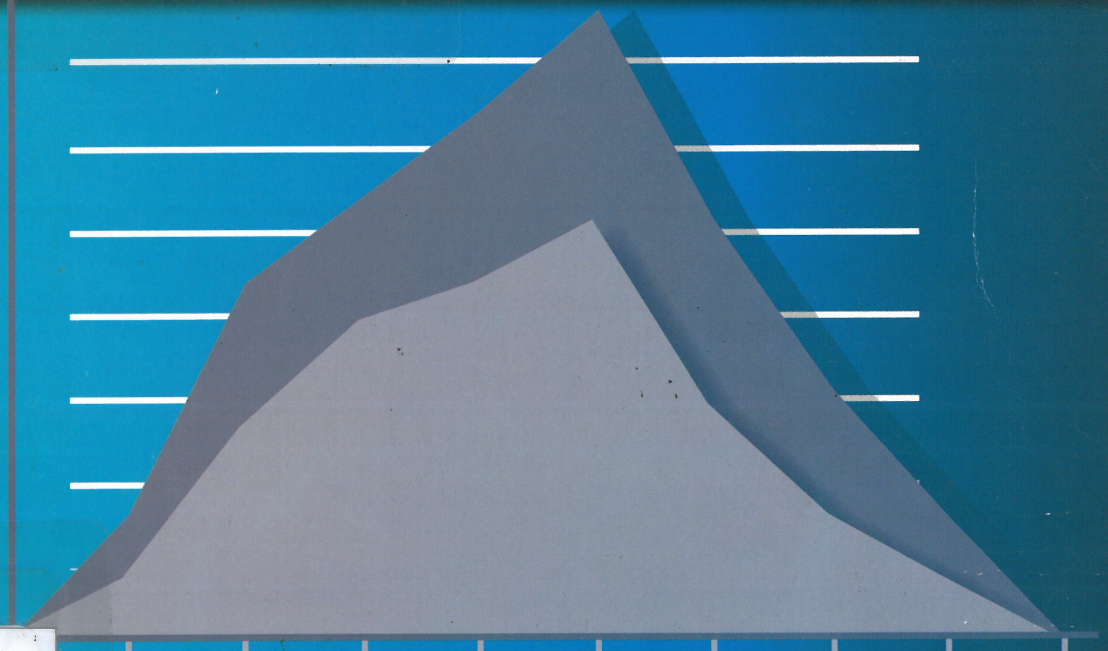
ELIO ARTURO BURGOS GÓMEZ

ESTADÍSTICA DESCRIPTIVA

APLICADA A LAS CIENCIAS DE LA SALUD

BIB002927



EUNED

Elio Arturo Burgos Gómez

ESTADÍSTICA DESCRIPTIVA
APLICADA A LAS CIENCIAS DE LA SALUD

CENTRO DE INFORMACIÓN TECNOLÓGICO

EUNED
EDITORIAL UNIVERSIDAD ESTATAL A DISTANCIA

PROMADE

Producción académica y asesoría metodológica:
Mario Marín Romero

CENIT - RECINTO ARANJUEZ
INSCRIPCIÓN
Nº 003951
CENTRO DE INFORMACIÓN TECNOLÓGICO

© Elio Arturo Burgos Gómez
© Sobre la presente edición:
Editorial Universidad Estatal a Distancia

Primera edición
Editorial Universidad Estatal a Distancia
San José, Costa Rica, 2016

UNIVERSIDAD HISPANOAMERICANA
CENIT
RECINTO ARANJUEZ

compra
21/04/2018
EDISA
$10.500

UNIVERSIDAD HISPANOAMERICANA
CENIT
RECINTO ARANJUEZ

SEDUCA
SISTEMA EDITORIAL UNIVERSITARIO CENTROAMERICANO

La Editorial EUNED es parte
del Sistema Editorial Universitario Centroamericano

Esta unidad didáctica fue confeccionada en la UNED, en el año 2016, para ser utilizada en la asignatura "Estadística en Salud I", código 3164, que se imparte en el programa de Registros y Estadísticas de Salud.

Equipo de producción

Coordinación académica:
Evelin Umaña Ramírez

Encargado de cátedra:
Luis Salazar Alvarado

Encargada de programa:
Hannia Castro Artavia

Especialista en contenido:
Jorge Espinoza Gutiérrez

Corrección de estilo:
Daniela Sánchez Sánchez

Elaboración de fórmulas:
Elio Arturo Burgos Gómez

Corrección de pruebas:
Mario Marín Romero

Ilustraciones:
Carmen Teresa Mora

Productor gráfico:
Jorge Delgado Castillo

Portada:
Jorge Delgado Castillo

Coordinador de Producción Editorial:
Daniel Villalobos Gamboa

Imposición digital:
Giovanna Calderón Zúñiga

978-9968-48-195-3

519.53
B957e Burgos Gómez, Elio Arturo, 1967- -

Estadística descriptiva aplicada a las ciencias de la salud / Elio Arturo Burgos Gómez. – San José, C.R. : EUNED, 2016

452 p

978-9968-48-195-3

1. ESTADISTICA DESCRIPTIVA 2. SERVICIOS DE SALUD. 3.CIENCIAS DE LA SALUD I. Título.

Impreso en Costa Rica.
Reservados todos los derechos.
Prohibida la reproducción total o parcial
bajo cualquier medio mecánico o electrónico
del contenido total o parcial de esta publicación
sin el permiso expreso del autor y de los editores.
Hecho el depósito que dicta la ley.

A mis hijos Elio, Alicia y Erick
Pues siempre me dan alegrías
e incentivos para seguir adelante

Prólogo

La atención de la salud en una colectividad demanda la participación de muchas personas capacitadas no solo en medicina, fisiología, química, microbiología, odontología, nutrición y otras ciencias de la salud, sino que también requiere de la participación de personal preparado en computación, electrónica, matemáticas y estadística. De esa cooperación, han surgido disciplinas mixtas como la demografía, la epidemiología, la bioestadística, la investigación experimental y los registros médicos, todas ellas con un denominador común: la estadística.

La estadística aplicada al estudio de los fenómenos poblacionales genera la demografía; orientada a la identificación de determinantes de la salud, de la enfermedad y su frecuencia da cuerpo a la epidemiología; encauzada a la pesquisa sobre seres vivos permite el desarrollo de la bioestadística; dirigida al muestreo, diseño de experimentos y prueba de hipótesis produce la investigación experimental; con sus técnicas de captura y procesamiento de datos posibilita mejores registros médicos.

Todas estas disciplinas potencian el conocimiento científico para prevenir la enfermedad, enfrentar sus causas, combatir las formas de contagio y diseminación, generar pronósticos y patrones de ocurrencia, además de brindar datos útiles para el establecimiento de políticas generales de salud, sistemas de vigilancia y campañas de divulgación que lleven a mejorar la salud de la comunidad.

Esos encuentros disciplinarios demandan fuertes bases teóricas en el campo de la estadística aplicada. Atender esas demandas, lleva a que en este primer documento se trabaje el instrumental general de la estadística descriptiva y se deje para un segundo tomo el desarrollo de la estadística inferencial. Ambos textos tienen una necesaria continuidad y mantienen relaciones de complementariedad; pero habrá que dar un paso primero, el otro después y regresar sobre la ruta pues "al andar se hace camino" (Machado).

Al poner en sus manos este trabajo, mi mayor aspiración es que disfrute su lectura y que las prácticas no le resulten aburridas; pero estaré conforme si, al menos, este material le ayuda a coronar sus metas académicas y lo sigue consultando durante su desempeño profesional.

EL AUTOR

Presentación

La estadística es la rama del saber que desarrolla y aplica técnicas de recolección, organización, presentación, análisis e interpretación de datos numéricos, obtenidos por observación o experimentación con el fin de obtener conclusiones válidas y apoyar la toma de decisiones.

Esta disciplina surgió por la necesidad de los Estados de contar con datos útiles para orientar el trabajo de sus dirigentes o estadistas. En su evolución como disciplina científica, se ha ramificado en dos vertientes: una teórica y otra aplicada. La primera se apoya en axiomas y razonamiento lógico-matemático para fundamentar sus métodos de análisis, en tanto que, la estadística aplicada se articula con otras disciplinas para proponer herramientas útiles en el manejo de los datos para usos prácticos.

La estadística aplicada se subdivide en descriptiva e inferencial. La primera presenta los datos efectivamente inspeccionados, revela sus características y sus relaciones exponiendo la información obtenida. La parte inferencial, revisa secciones o muestras que le permiten generalizar sus hallazgos a toda la población.

Metodológicamente, la parte descriptiva es la base para posteriores desarrollos teóricos, entre ellos las técnicas de pronóstico, la generalización y la inferencia. Didácticamente, la descripción de lo que se puede contar, pesar o medir es más accesible a los sentidos; luego se podrá ampliar el estudio para eventos futuros, probables o inciertos. Por ello, este texto comienza su exposición con la parte descriptiva del procesamiento de datos recolectados, analizados y vertidos en informes textuales, semitabulares, tabulares y gráficos.

La descripción también genera medidas de posición, dispersión y forma, a partir de ellas se calculan números relativos como las razones, proporciones, tasas e índices; con estos indicadores es posible estimar probabilidades, riesgos y oportunidades relativas. Al hacerlo, se estará en la frontera que separa la descripción de la inferencia. Hasta allí llega este documento, justo en el límite donde ya no solo se describe sino que se comienza a inferir.

Comencemos, entonces, este caminar por una disciplina que te será de gran aplicación en las investigaciones, tanto las de práctica durante los años de formación académica, como para aquellas a las que te enfrentarás durante toda tu vida profesional.

Contenido

PRÓLOGO IX

PRESENTACIÓN XI

CAPÍTULO 1

APLICACIÓN DE LA ESTADÍSTICA EN LAS CIENCIAS DE LA SALUD 1

1.1. **DEFINAMOS ESTADÍSTICA** 4

1.2. **CLASIFICACIÓN DE LA ESTADÍSTICA** 5

1.2.1. Estadística descriptiva 6

1.2.2. Estadística inferencial 6

1.3. **FASES DE LA INVESTIGACIÓN CIENTÍFICA** 8

1.4. **TIPOS DE INVESTIGACIÓN CIENTÍFICA** 9

1.4.1. Por la temporalidad del estudio 10

1.4.2. Investigaciones científicas según el tipo de resultado 11

1.4.3. Investigaciones científicas según la unidad de estudio 13

1.4.4. Investigaciones científicas según el grado de intervención del investigador 14

1.5. **ELEMENTOS POR CONSIDERAR EN UNA INVESTIGACIÓN** 16

1.5.1. La población 16

1.5.2. Unidad estadística, característica y observación 17

1.5.3. Atributos y variables 18

1.5.4. Censo y muestras 20

1.5.5. Selección de la muestra 21

1.5.6. Errores y sesgos 23

1.6. **OBTENCIÓN DE LA INFORMACIÓN** 24

1.6.1. Fuentes internas o externas, primarias o secundarias 25

1.6.2. Evaluación de los datos y las fuentes 26

1.6.3 . Recolección de datos no existentes 28

EJERCICIOS DE REPASO Y AUTOEVALUACIÓN 31

CAPÍTULO 2

ELABORACIÓN, PRESENTACIÓN Y ANÁLISIS ESTADÍSTICO DE LOS DATOS 35

2.1. **ETAPAS DE UNA INVESTIGACIÓN** 38

2.1.1. Etapa preparatoria 38

2.1.1.1. Identificación del problema 39

2.1.1.2. Estado de la cuestión 39

2.1.1.3. Establecimiento de los objetivos 40

2.1.1.4. Construcción de los instrumentos de recolección de datos 40

2.1.1.5. Tareas administrativas en la preparación 42

2.1.2. Etapa de ejecución 42

2.1.2.1. Prueba piloto 43

2.1.2.2. Selección y capacitación de los entrevistadores 43

2.1.2.3. Tareas administrativas en la ejecución 43

2.1.2.4. El trabajo de recolección de los datos 44

2.1.3. Etapa de procesamiento y análisis 44

2.1.3.1. El procesamiento de los datos 45

2.1.3.2. El análisis e interpretación 45

2.1.3.3. El informe de investigación 45

2.2. **DISEÑO DE CUESTIONARIOS** 46

2.2.1. Clases de preguntas 47

2.2.2. La redacción de los ítems 54

2.2.3. Obtención y procesamiento de los datos 59

2.2.4. Métodos de tabulación 63

2.2.5. Distribuciones simples y de doble entrada 65

2.3. **PRESENTACIÓN DE LOS DATOS ESTADÍSTICOS** 68

2.3.1. Forma textual 68

2.3.2. Forma semitabular 68

2.3.3. Cuadros estadísticos 69

2.3.3.1. Series estadísticas 71

2.3.3.2. Distribuciones de frecuencia de variables continuas 76

2.3.3.3. Detalles adicionales para la confección de los cuadros 84

EJERCICIOS DE REPASO Y AUTOEVALUACIÓN 87

CAPÍTULO 3

LOS INDICADORES ESTADÍSTICOS EN EL CAMPO DE LA SALUD 91

3.1 **LOS NÚMEROS RELATIVOS** 96

3.1.1. Las razones 96

3.1.2. Las proporciones 98

3.1.2.1. Aplicación de razones y proporciones: la prueba de tamizaje 100

3.1.2.2. Presunciones para la validación de pruebas de tamizaje. 103

3.1.3. Las tasas 105

3.2. **MORBILIDAD** 110

3.3. **INDICADORES PROPORCIONALES** 112

3.3.1. Indicadores proporcionales de natalidad 112

3.3.2. Indicadores proporcionales de mortalidad 114

3.3.3. Indicadores proporcionales de morbilidad 117

3.4. **CÁLCULOS DE RIESGOS Y OPORTUNIDADES** 119

3.4.1. La Oportunidad Relativa (OR = Odds Ratio) 120

3.4.2. El Riesgo Relativo (RR) 121

3.4.3. El Riesgo Atribuible (RA) 122

3.4.4. La Fracción Atribuible (FA) 122

3.5. **TASAS AJUSTADAS** 124

3.5.1. Método directo de ajuste de tasas 125

3.5.2. Método indirecto de ajuste de tasas 130

3.5.3. La razón de estandarización 132

3.5.4. La razón de estandarización como indicador de salud 136

EJERCICIOS DE REPASO Y AUTOEVALUACIÓN 137

CAPÍTULO 4

MEDIDAS DE POSICIÓN Y DE VARIABILIDAD 141

4.1. **MEDIDAS DE POSICIÓN** 144

4.1.1. La moda (Mo) 144

4.1.2. El promedio ($\bar{x}$) 145

4.1.3. La mediana (Me) 146

4.2. **CÁLCULO DE LAS MEDIDAS DE POSICIÓN PARA DATOS AGRUPADOS** 148

4.2.1. La moda (Mo) 148

4.2.2. El promedio ($\bar{x}$) 150

4.2.3. La mediana (Me) 151

4.3. **USO DE LAS MEDIDAS DE POSICIÓN** 153

4.4. **LOS CUANTILES** 158

4.4.1. Los percentiles 158

4.4.2. Los cuartiles 161

4.4.3. Los percentiles como medidas de posición 164

4.4.4. Rango del percentil 165

4.5. **LA VARIABILIDAD** 168

4.5.1. Medición de la variabilidad 168

4.5.2. La desviación media 169

4.5.3. La desviación estándar 171

4.5.4. La variancia 173

4.5.5. Cálculo de la variancia para datos agrupados 174

4.5.6. Coeficiente de variación 177

4.5.7. Percentiles como medidas de variabilidad 178

4.6. **FIJACIÓN DE LÍMITES DE TOLERANCIA** 179

4.7. **LA VARIABILIDAD EN EL CASO DE PROPORCIONES** 181

4.8. **APLICACIONES AL ÁREA DE SALUD: EL ÍNDICE ENDÉMICO** 183

EJERCICIOS DE REPASO Y AUTOEVALUACIÓN 185

CAPÍTULO 5

ESTADÍSTICAS DE LOS SERVICIOS DE SALUD Y EL ANÁLISIS GRÁFICO 189

5.1. **ELEMENTOS DE LOS GRÁFICOS** 192

5.2. **LIMITACIONES DE LOS GRÁFICOS** 195

5.3. **TIPOS DE GRÁFICOS** 195

5.3.1. Gráfico lineal 197

5.3.2. Gráficos de barras horizontales 204

5.3.3. Gráfico de bastones 205

5.3.4. Los gráficos de columnas 211

5.3.5. Representación de la participación relativa 215

5.3.6. El histograma 220

5.3.7. El polígono de frecuencias 222

5.3.8. La pirámide de población 224

5.3.9. Diagrama de Pareto 227

5.3.10. Diagrama de dispersión 229

5.3.11. Ojivas de frecuencias acumuladas 230

5.3.12. Pictogramas 236

5.3.13. Mapas estadísticos 237

5.3.14. El gráfico de índice endémico 238

EJERCICIOS DE REPASO Y AUTOEVALUACIÓN 241

CAPÍTULO 6

INTRODUCCIÓN AL ESTUDIO DE LAS PROBABILIDADES 245

6.1. **ESPACIO MUESTRAL Y EVENTOS DE UN EXPERIMENTO** 249

6.2. **TRES ENFOQUES SOBRE PROBABILIDADES** 253

6.2.1. Enfoque clásico 253

6.2.2. Enfoque empírico 254

6.2.3. Enfoque subjetivo 255

6.3. **REGLAS DE LAS PROBABILIDADES** 257

6.3.1. Eventos mutuamente excluyentes 257

6.3.2. Eventos que no son mutuamente excluyentes 259

6.3.3. Probabilidad complementaria 262

6.3.4. Probabilidad conjunta 262

6.4. **TABLAS DE CONTINGENCIA** 264

6.5. **INDEPENDENCIA Y DEPENDENCIA** 271

6.6. **LA PROBABILIDAD CONDICIONAL** 273

EJERCICIOS DE REPASO Y AUTOEVALUACIÓN 277

ANEXOS 281

RESPUESTAS A LOS EJERCICIOS DE AUTOEVALUACIÓN 307

HIPERVÍNCULOS 407

GLOSARIO 409

REFERENCIAS BIBLIOGRÁFICAS 423

Descripción de contenidos

Esta unidad didáctica se ha subdividido en seis capítulos que, en lo fundamental, tratan lo siguiente:

- Capítulo 1. **Aplicación de la estadística en las ciencias de la salud**

  Justifica la necesidad de trabajar con datos numéricos con el fin de describir situaciones o inferir impactos en la población, y para generar herramientas de análisis útiles en las ciencias de la salud.

- Capítulo 2. **Elaboración, presentación y análisis estadístico de los datos**

  Las investigaciones científicas pasan por etapas preparatorias, de ejecución, procesamiento y análisis. Para la obtención de los datos y su posterior análisis, es necesaria la construcción de instrumentos eficientes de recolección, de manera que posibiliten comunicar los hallazgos más importantes en forma textual, semitabular, tabular y gráfica.

- Capítulo 3. **Los indicadores estadísticos en el campo de la salud**

  Las razones, proporciones y tasas de hechos vitales se utilizan como indicadores del estado de salud, de la atención, saneamiento ambiental o situación socioeconómica de una población y sirven como guías de medición de los avances, estancamientos o retrocesos en el campo de la salud.

- Capítulo 4. **Medidas de posición y de variabilidad**

  Se estudian los métodos de cálculo e interpretación de las principales medidas de posición, forma y variabilidad, así como la construcción de intervalos de control, límites de tolerancia en distribuciones simétricas o asimétricas y la interpretación del índice endémico en una comunidad.

- Capítulo 5. **Estadísticas de los servicios de salud y el análisis gráfico**

  Este capítulo presenta las técnicas de representación gráfica de los datos estadísticos, proporcionando herramientas útiles en la confección, la interpretación y la selección del tipo de gráfico a utilizar en los informes de investigación.

- Capítulo 6. **Introducción al estudio de las probabilidades**

  Aunque no se puede tener plena certeza del rumbo de las actividades vitales como el nacimiento, la enfermedad o la muerte, sí se puede calcular el riesgo asociado a sus ocurrencias. Por eso se trabajan los enfoques de probabilidad y la resolución de problemas con eventos individuales, luego con pares de eventos mutuamente excluyentes o que pueden ocurrir conjuntamente, para finalmente trabajar la forma en que ocurrencias previas condicionan la probabilidad de eventos sucesivos.

Recomendaciones para su estudio

La estadística es práctica y requiere ser practicada. En esto se parece a los deportes o al arte.

Puede contar con recursos y habilidades, pero debera ejercitarte: disponer de piernas fuertes y bicicleta no es suficiente para convertirte en campeón(a) de ciclismo, para lograrlo tendrá que madrugar y hacer los recorridos indicados por su entrenador(a); poseer dedos largos y un piano en casa tampoco le convertirán en pianista, en algo ayuda, pero requerirá de horas practicando sobre el teclado.

Este texto equivale a su bicicleta o su piano, sus habilidades equivalen a las piernas o los dedos, pero ninguno le exime de hacer las prácticas.

El documento comienza exponiendo los conceptos y teorías desarrolladas por la disciplina, para luego apoyarse en uno o más ejemplos ilustrativos; seguidamente, viene un ejercicio práctico que es conveniente que complete. Algunos de estos ejercicios son de desarrollo, otros de cálculo, a veces tendrá que redactar preguntas, otras hacer conteos y también graficar.

Que no le preocupe equivocarse, ¡es parte del aprendizaje! Coteje su respuesta con el solucionarlo adjunto. Podría ser que tenga que repasar el concepto; igual es posible que las respuestas no sean únicas y permitan un buen coloquio con su tutor; también puede ocurrir que tenga razón y nos ayudará a mejorar el texto para beneficio de futuros estudiantes.

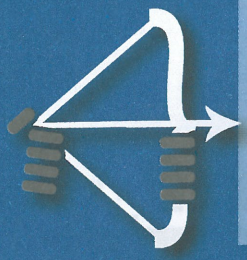
Recuerde que la UNED pone a su disposición varios recursos para estudiar, como tutorías, consulta telefónica, blog y preguntas por la red y aula virtual. ¡Mucho éxito!

ESTADÍSTICA DESCRIPTIVA
APLICADA A LAS CIENCIAS DE LA SALUD

CAPÍTULO 1

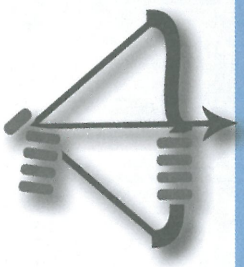
APLICACIÓN DE LA ESTADÍSTICA EN LAS CIENCIAS DE LA SALUD

OBJETIVO GENERAL



Adquirir conocimientos sobre los elementos básicos de la estadística y su aplicación para la comprensión de los fenómenos de salud para la toma de decisiones en este campo. Además, aplicar los conceptos estadísticos adquiridos para el planeamiento y la recolección de la información para efectos de investigación.

OBJETIVOS ESPECÍFICOS



Al finalizar el estudio de este capítulo usted estará en la capacidad de:

1. Redactar su propia definición de estadística.
2. Distinguir en los informes de investigación aquellos que son descriptivos o inferenciales.
3. Clasificar los estudios en transversales o longitudinales, descriptivos o analíticos, individuales o ecológicos, observacionales, experimentales o cuasi-experimentales.
4. Señalar las poblaciones finitas o infinitas, de entre un listado de poblaciones.
5. Ordenar las características de interés según sean atributos o variables discretas, continuas o dicotómicas.
6. Indicar las diferentes formas de obtención de una muestra.
7. Mencionar las ventajas y desventajas de los métodos de recolección de datos.

La estadística, al trabajar con datos numéricos para describir situaciones o para inferir impactos en la población, genera herramientas útiles en las ciencias de la salud. Las investigaciones demandan el uso del método científico, el cual pasa por las siguientes etapas: conteo, medición, inferencia, descripción cuantitativa, verificación de hipótesis y divulgación de los resultados para generar nuevo conocimiento y sugerir acciones correctivas.

Con números, pesos y medidas
Dios construyó todo.

ISAAC NEWTON

¿En qué tareas de las ciencias de la salud puede ayudar la estadística?

- Descripción cuantitativa del estado de salud de una comunidad
- Establecimiento de relaciones de causa y efecto en situaciones de morbilidad
- Señalización de tendencias de incidencia, contagio y mortalidad
- Generación de indicadores de cobertura, utilización y eficiencia en los servicios
- Indicación de herramientas para la evaluación de los diferentes programas de salud pública
- Establecimiento de tamaños de muestra óptimos para diferentes pruebas
- Prueba de tratamientos y lo significativo de su implementación
- Cálculo de las probabilidades y el riesgo de contagio
- Correlación de variables, su nivel de asociación y determinación

1.1. DEFINAMOS ESTADÍSTICA

Revise las siguientes definiciones de estadística:

"Disciplina científica dedicada al desarrollo y aplicación de la teoría y las técnicas apropiadas para la recolección, clasificación, presentación, análisis e interpretación de información cuantitativa obtenida por observación o experimentación" (Gómez, 1998, p. 5).

"Rama del saber que trata del desarrollo y aplicación de *métodos* eficientes de recolección, elaboración, presentación, análisis e interpretación de datos numéricos" (Moya, 2009, p. 5).

"La estadística se ocupa de los métodos científicos que se utilizan para recolectar, organizar, resumir, presentar y analizar datos, así como para obtener conclusiones válidas y tomar decisiones razonables con base en este análisis" (Spiegel y Stephens, 2009, p. 1).

"Ciencia que recoge, organiza, presenta, analiza e interpreta datos con el fin de propiciar la toma de decisiones más eficaz" (Lind, Marchal y Wathen, 2009, p. 4).

Con base en estas definiciones se confecciona el siguiente cuadro analítico:

Cuadro 1.1

Cuadro analítico de algunas definiciones de estadística

Referencia a la ciencia			
Gómez Disciplina científica	Moya Rama del saber	Spiegel y Stephens ---	Lind *et al.* Ciencia
Objeto de estudio			
Gómez Dedicada al desarrollo y aplicación de la teoría de y las técnicas apropiadas	Moya Trata del desarrollo y aplicación de métodos eficientes	Spiegel y Stephens Se ocupa de los métodos científicos que se utilizan	Lind *et al.* ---
Tareas			
Gómez Recolección, clasificación, presentación, análisis e interpretación de información cuantitativa	Moya Recolección, elaboración, presentación, análisis e interpretación de datos numéricos	Spiegel y Stephens Recolectar, organizar, resumir, presentar y analizar datos	Lind *et al.* Recoge, organiza, presenta, analiza e interpreta datos
Métodos			
Gómez Observación Experimentación	Moya ---	Spiegel y Stephens Con base en este análisis	Lind *et al.* ---
Fin (¿Para qué?)			
Gómez ---	Moya ---	Spiegel y Stephens para obtener conclusiones válidas y tomar decisiones razonables	Lind *et al.* con el fin de propiciar la toma de decisiones más eficaz

EJERCICIO 1.1

Utilice el cuadro analítico de las definiciones de estadística y redacte su propia definición de forma que contenga:

- La referencia a la ciencia
- El objeto de estudio de la estadística
- Las tareas que efectúa
- Los métodos que aplica
- El fin de la estadística

(Recomendación: muestre su definición al tutor y coméntela)

1.2. CLASIFICACIÓN DE LA ESTADÍSTICA

La evolución de la estadística la ha llevado a desarrollar un sustento teórico y una gran variedad de aplicaciones. La parte teórica se basa en axiomas y el razonamiento lógico-matemático para dar fundamento a sus métodos de análisis. En las aplicaciones, la estadística se hermana con otras disciplinas para generar instrumentos de trabajo, de esta forma la demografía analiza poblaciones humanas y la bioestadística estudia a los seres vivos, además, se generan múltiples herramientas para la economía, la ingeniería, la sociología y la salud.

La *estadística aplicada* se clasifica en:

- Estadística descriptiva
- Estadística inferencial

1.2.1. ESTADÍSTICA DESCRIPTIVA

La *estadística descriptiva* se especializa en la recolección, el resumen y la presentación de los datos. Destaca sus características y relaciones para hacer la información más comprensible.

Entre los instrumentos utilizados por la estadística descriptiva se encuentran:

- La confección de cuadros y tablas
- La elaboración de gráficos informativos
- El cálculo de porcentajes, tasas e indicadores
- La generación de medidas de posición, variabilidad y forma

1.2.2. ESTADÍSTICA INFERENCIAL

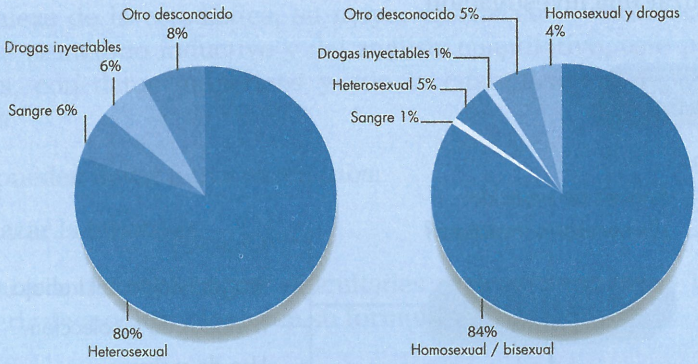
La *estadística inferencial* trabaja con muestras para hacer generalizaciones acerca de la población o, partiendo de experimentos, predice el comportamiento de las variables en situaciones similares.

Entre los instrumentos utilizados por la estadística inferencial están:

- El cálculo de probabilidades y riesgo
- Las distribuciones teóricas de frecuencias
- Las pruebas de hipótesis y límites de confianza
- El establecimiento de pronósticos y tendencias

EJERCICIO 1.2

Tras leer los siguientes informes de investigación, clasifíquelos en descriptivos o inferenciales marcando con una X en la casilla que corresponda.

		Descriptivo	Inferencial
a)	Méndez Fonseca y Araya Umaña (2003) señalan: "Las dos últimas décadas del siglo pasado, periodo en el cual el ritmo de mejoramiento de la esperanza de vida baja sensiblemente y se insinúa una tendencia a la estabilización en un valor cercano a los 78 años".		
b)	Mann et al. (1992), en el reporte sobre SIDA en el mundo, informa lo siguiente: **Proporción de casos de SIDA en Canadá por modo de transmisión y género al 1 de julio de 1991**  Figura 1.1. **Proporción de casos de SIDA en canadá por modo de transmisión y género al 1 de julio de 1991.**		
c)	Sawyer RG, Raymond DP, Pelletier SJ et al (2001), investigadores del Departamento de Cirugía de la Universidad de Virginia, estudiaron 2457 infecciones sufridas por el personal hospitalario, a partir de esos datos sugieren que "el riesgo de muerte es del 13 % (Tasa Global de Mortalidad), en un rango que va desde el 5 % en las infecciones extrahospitalarias hasta un 25 % en las intrahospitalarias".		
d)	Roselló-Araya Marlene et al. (2004) destacan de su investigación que "la prevalencia de diabetes fue 2,5 % (1,9 % en hombres y 3,1 % mujeres). El 89 % de la muestra eran mayores de 40 años cuya prevalencia fue de 9,4 % (7,4 % hombres, 11,4 % mujeres). La mayoría de diabéticos eran casados, tenían educación primaria y se clasificaron como "no pobres". El 38,2 % usaba insulina. En el último semestre, el 82,2 % asistió a consultas médicas y el 8,4 % requirió hospitalización. La enfermedad causó complicaciones en 36,1 %, especialmente en la agudeza visual".		
e)	Jiménez-Montero, José Guillermo (2004), basado en fuentes secundarias y en una muestra de pacientes, infiere que "el sobrepeso y la obesidad aumentan el riesgo de desarrollar diabetes mellitus tipo 2, hipertensión arterial, dislipidemia, enfermedad coronaria, insuficiencia cardiaca congestiva, accidente cerebrovascular y otras condiciones como litiasis vesicular, esteatosis hepática, apnea de sueño, cáncer de próstata, de endometrio, de mama y de colon".		
f)	Baltodano-Zúñiga, Víctor Julio et al. (2012), sabiendo que un 30 % de la población padece de obesidad, existe una prevalencia de diabetes del 2,5 % y la existencia de un 2 % de pacientes que es a la vez obeso y diabético, coligen que la probabilidad de atender de emergencia a un paciente obeso dado que es diabético es de 80 %.		

1.3. FASES DE LA INVESTIGACIÓN CIENTÍFICA

La investigación científica es el estudio sistemático, organizado y objetivo de los datos con el fin de comprender, comprobar, corregir o utilizar el conocimiento. Toda investigación científica conlleva al menos tres momentos:

- Observación
- Formulación de hipótesis
- Verificación o prueba

En la etapa de *observación* se detecta un problema o un fenómeno de interés, por ejemplo, la práctica clínica puede alertar de un aumento en la atención de casos o brotes de una enfermedad que llevan a plantearse un análisis de la situación y a preguntarse si es necesario realizar una investigación.

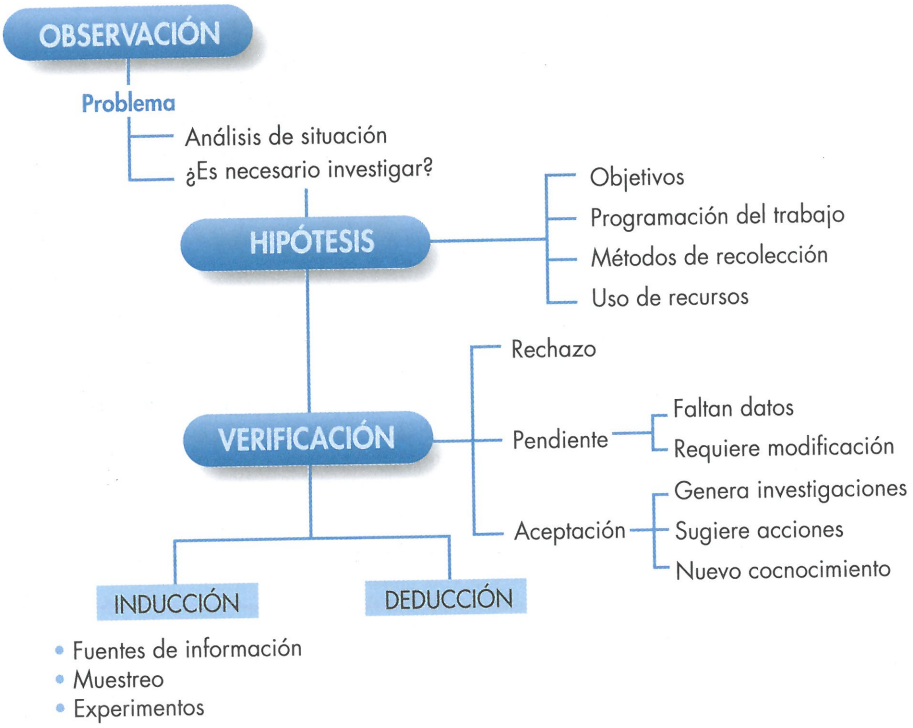


Figura 1.2. **Principales etapas de la investigación científica**

Fuente: Elaboración con fines didácticos basado en Gallardo (2005)

En la fase de *formulación de hipótesis* se realiza un enunciado basado en sospechas o en resultados de otros estudios. Este propone una relación entre los datos, una explicación preliminar del problema, una posible solución o un método de comprobación de las determinaciones entre variables.

Una adecuada formulación de la hipótesis permite precisar los objetivos de la investigación, programar las fases del trabajo, los métodos, los instrumentos y las técnicas de recolección de la evidencia, además de establecer una estrategia para el uso de los recursos disponibles.

La etapa de *verificación o comprobación* requiere del uso del método inductivo, del deductivo o de una combinación de ambos; además, esta puede hacerse tanto sobre bases cuantitativas como cualitativas.

El método inductivo comienza con la observación y el análisis de una muestra para inferir la totalidad, mientras que el método deductivo parte de leyes generales para explicar un fenómeno específico. Inducir es estudiar una parte y generalizar para el todo, en cambio deducir es formular una hipótesis general para luego advertir sobre fenómenos individuales.

Por la naturaleza de la estadística, su mayor aporte en la verificación de hipótesis se hace desde el método inductivo y del análisis cuantitativo, esto porque se trabaja con muestras, con datos numéricos y con la cuantificación de los resultados de experimentos.

Las pruebas pueden llevar a tres escenarios:

- Rechazar la hipótesis
- Declararla incompleta por dificultades para obtener datos relevantes o por requerir de modificaciones en su formulación
- Aceptar la hipótesis

Una hipótesis aceptada puede:

- Proponer nuevas investigaciones para profundizar en el abordaje del tema
- Sugerir acciones para enfrentar el problema
- Aportar información nueva para la comunidad

1.4. TIPOS DE INVESTIGACIÓN CIENTÍFICA

Las investigaciones científicas se clasifican de acuerdo con:

- La temporalidad del estudio
- El tipo de resultado obtenido
- La unidad de estudio
- El grado de intervención del investigador

1.4.1. POR LA TEMPORALIDAD DEL ESTUDIO

El espacio temporal en que se efectúa la investigación permite clasificar los estudios en:

- Transversales
- Longitudinales

Los estudios *transversales* son como una fotografía de una situación en un tiempo y un espacio determinados. En el campo de la salud, posibilitan el análisis de la prevalencia de una enfermedad y en su gran mayoría son descriptivos.

Las ventajas de las investigaciones transversales son que se efectúan con mayor rapidez, permiten actualizar los registros médicos, aportan insumos para la implementación de políticas de salud y proporcionan datos para futuros estudios analíticos.

Las investigaciones *longitudinales* son como una película, ya que necesitan del transcurso del tiempo para identificar posibles causas de un fenómeno, el comportamiento de una enfermedad y apreciar su tendencia.

Los estudios longitudinales se subdividen en:

- Retrospectivos
- Prospectivos

Un análisis es *retrospectivo* cuando se realiza con datos del pasado y está basado en registros de salud, expedientes, archivos, observaciones clínicas o reportes de los laboratorios. Generalmente, las fuentes de los estudios retrospectivos son secundarias y parten de una hipótesis en la cual se sospecha de una causa en el pasado que provoca una enfermedad en el presente.

Alberto Polanco, citado por Alba Lobato, dice de las investigaciones retrospectivas buscan "las causas a partir de un efecto que ya se presentó. Los estudios retrospectivos parten de un efecto y regresan a buscar una causa. Es como si fuésemos hacia atrás, por esto es retrospectivo" (Alba, 2010, p. 21).

Los estudios *prospectivos* comienzan en el presente con datos que se recolectarán en el futuro. Estos análisis se diseñan hoy y parten de una causa sospechosa que estará bajo observación en el futuro, los datos se obtendrán a medida que aparezcan y, por ello, tienden a originarse en fuentes primarias, recogidas por el equipo investigador evitando los sesgos de medición, recolección y memoria.

EJERCICIO 1.3

Clasifique los siguientes estudios en transversales o longitudinales.

		Transversal	Longitudinal
a)	Se quiere conocer el porcentaje de mujeres embarazadas con HIV, atendidas en la Maternidad Carit en agosto de 2011. Asimismo, se busca información sobre el número de compañeros sexuales, utilización del condón, uso de drogas inyectadas, otras enfermedades venéreas sufridas, padecimientos de hemofilia y práctica de transfusiones sanguíneas.		
b)	2500 estudiantes de secundaria serán observados durante 10 años para determinar la evolución del consumo de alcohol. Se pretende evaluar la relación de causa-efecto entre eventuales conductas antisociales y el consumo, además de la influencia de la desinhibición, susceptibilidad y reciprocidad, el comportamiento por género, clase social y el contexto en que se da la ingesta.		
c)	Una propuesta de investigación plantea la necesidad de generar conocimiento sobre el proceso de envejecimiento de la población nicoyana mayor de 60 años. Para ello solicitan asignación presupuestaria por 20 años para financiar las observaciones, detectar cambios y transiciones, además del comportamiento colectivo e individual conforme se da el proceso de envejecimiento.		
d)	Los informes internacionales reportan que el porcentaje de pacientes curados de cáncer después de 5 años de tratamiento es del 50 %. Para evaluar si esa proporción es válida en el país, toma una muestra de 1200 pacientes con situación similar para efectuar la prueba de significación.		
e)	En una población de tamaño N, hay r enfermos y N-r individuos sanos. Se toma una muestra de n personas y se identifican x enfermos y n-x sanos. Con esos datos se desea estimar la prevalencia de la enfermedad en la población.		
f)	El PANI está interesado en conocer los patrones de crecimiento en talla y peso de los niños albergados durante sus primeros 5 años de vida, mientras no hayan sido dados en adopción.		

1.4.2. INVESTIGACIONES CIENTÍFICAS SEGÚN EL TIPO DE RESULTADO

Por el resultado obtenido, las investigaciones se clasifican en:

- Descriptivas
- Analíticas

Las investigaciones *descriptivas* buscan detallar las manifestaciones de un fenómeno a partir de su observación. Pretenden explicar los casos clínicos, identificar su ocurrencia, la frecuencia con que se presentan, las zonas en que suceden y determinar la tendencia a partir de series de casos y de los indicadores.

Las investigaciones *analíticas* pretenden establecer conexiones entre variables, medir su correlación o nivel de asociación, comprobar su causalidad, plantear escenarios futuros, medidas preventivas y las consecuencias sobre la salud. Con frecuencia, los estudios analíticos se utilizan para probar hipótesis planteadas por investigaciones descriptivas.

EJERCICIO 1.4

Clasifique los siguientes estudios en descriptivos o analíticos.

		Descriptivo	Analítico
a)	El Programa "Niño Sano" quiere medir el nivel de conocimiento que tienen las madres de niños en edad preescolar sobre la prevención de la diarrea. Entre otras variables, interesa conocer la edad de las madres, la escolaridad alcanzada, su nivel de ingresos y las condiciones habitacionales.		
b)	A partir de informes de casos y series de casos que describen hombres jóvenes con neumonía o sarcoma de Kaposi, se plantea una investigación para correlacionar estos padecimientos con la preferencia sexual, el número de compañeros sexuales y la exposición a drogas recreativas.		
c)	Se busca conocer las habilidades comunicativas del personal de enfermería que imparte los cursos de saneamiento ambiental. Además de la habilidad de expresión, interesa medir la empatía y las técnicas empleadas.		
d)	Se busca correlacionar las características epidemiológicas de las personas involucradas en accidentes de tránsito en la carretera San José-Guápiles, eventos ocurridos durante el mes de diciembre de 2010. Para ello, se registró el tipo de accidente, el porcentaje de mortalidad, el tipo de lesión, el diagnóstico hospitalario, el sexo de los conductores, la edad, la presión arterial, la hora del accidente y el tipo de vehículo involucrado.		
e)	Tras recoger muestras de contaminantes particulados del medio ambiente, y relacionarlas con la incidencia de asma e infecciones de las vías respiratorias, se busca establecer la influencia de la contaminación en el ADN y en el proceso de apoptosis de las células bronquio-epiteliales.		
f)	Dado que la hemorragia digestiva alta (HDA) en personas de la tercera edad tiene una elevada mortalidad, se pretende establecer un perfil del paciente de bajo riesgo, clasificándolo por rango de edad, sexo, ingesta de medicamentos, alcohol, tabaquismo y hábitos de alimentación.		

1.4.3. INVESTIGACIONES CIENTÍFICAS SEGÚN LA UNIDAD DE ESTUDIO [1]

De acuerdo con la unidad de estudio, las investigaciones se clasifican en:

- Individuales
- Ecológicas

Los estudios *individuales* o de caso analizan a un solo paciente. De hecho, el individuo es la unidad de estudio y el sujeto de las investigaciones clínicas. El individuo es la fuente de los datos y provoca la necesidad de investigación. El estudio de caso es útil para estudiar enfermedades raras, de baja incidencia o prevalencia, además requiere de pocos recursos financieros y absorbe menos tiempo. Sin embargo, pueden presentar un sesgo de selección por mayor riesgo de tener casos no representativos del total, también pueden llevar a clasificaciones erróneas y dificultar establecer la causa de la enfermedad.

En los estudios *ecológicos* las unidades de estudio son las poblaciones, las comunidades o los grupos específicos. Son útiles para establecer efectos que no se identifican en un paciente aislado y, también, son de gran ayuda para evaluar el impacto de programas, los resultados de intervenciones, la efectividad de las políticas de salud en una comunidad y permiten plantear nuevas hipótesis para futuros trabajos.

1 La unidad de estudio es el individuo o conjunto de individuos de donde se obtendrán los datos.

EJERCICIO 1.5

Clasifique los siguientes estudios en individuales o ecológicos.

		Individuales	Ecológicos
a)	Se busca estudiar los niveles de concentración de sales de calcio en el agua utilizada para beber o cocinar y su relación con la nefrolitiasis (piedra en el riñón).		
b)	Una niña autista es observada durante dos años por medio de filmaciones libres y semi estructuradas tanto en la Escuela Centeno Güell como en la casa de sus padres. Se busca analizar la evolución de su lenguaje, sus relaciones sociales y el uso que hace de los objetos con los que tiene contacto.		
c)	Se busca establecer las relaciones existentes entre los niveles de consumo de sal, los patrones alimenticios de una comunidad y la incidencia del cáncer gástrico.		
d)	Existe la sospecha de que la prevalencia de labio leporino y paladar hendido está afectada por factores genéticos, por lo que se propone llevar un registro del color de la piel, ojos y cabello, además de la estatura, rasgos faciales y apariencia física de los padres de niños con este padecimiento.		
e)	Un equipo de terapeutas del lenguaje estudia el proceso de adquisición de habilidades comunicativas de un hijo único de madre soltera sordomuda. Se sabe que este niño de cuatro años no socializaba con otras personas y que su contacto con el lenguaje articulado está basado en muchas horas frente al televisor.		
f)	Se quiere establecer el tipo de relación entre el porcentaje de colecistectomías (extracción de la vesícula biliar) y la tasa de mortalidad por cáncer de vesícula.		

1.4.4. INVESTIGACIONES CIENTÍFICAS SEGÚN EL GRADO DE INTERVENCIÓN DEL INVESTIGADOR

Por la intervención del investigador, los estudios se clasifican en:

- Observacionales
- Experimentales
- Cuasi-experimentales

Todas las investigaciones científicas necesitan de la observación; sin embargo, se llaman estudios *observacionales* a aquellos en los que no existe interferencia del investigador, sino que este se dedica a observar las variables, a cuantificarla o medirlas pero sin manipularlas ni controlarlas. La observación favorece la recolección de los datos y manifestaciones del fenómeno, lo cual permite un análisis estadístico para generar razones, proporciones, tasas e indicadores. También facilita los insumos para correlacionar variables, establecer tendencias, medidas de riesgo y medidas de impacto.

Los estudios *experimentales* pueden ser clínicos, ecológicos o de laboratorio. En estos se da la intervención del investigador pues introduce una alteración intencional para estudiar los efectos y llegar hasta el control de las variables. En muchas investigaciones se definen varios grupos que se someten a diversos grados de exposición a una variable causal y un grupo de control que no está expuesto a la variable de análisis, luego se analizan los efectos y se establecen conclusiones.

Las investigaciones son *cuasi-experimentales* cuando la variable en estudio no es manipulada por el investigador sino que se produce exógenamente, por ejemplo: los niveles de exposición a agentes contaminantes del ambiente o los efectos de los patrones alimenticios de diferentes comunidades.

EJERCICIO 1.6

Clasifique los siguientes estudios en observacionales (O), experimentales (E) o cuasi-experimentales (C).

		O	E	C
a)	Una comunidad guanacasteca está expuesta al polvo que produce un quebrador de piedra caliza. El personal médico quiere cuantificar la incidencia de enfermedades bronquiales en los niños de edad escolar, relacionando la distancia entre las casas y el quebrador más la dirección de los vientos. Para ello identifica diez barrios de los cuales seleccionará una muestra de niños en edad escolar para someterlos a estudio.			
b)	Se afirma que desayunar con avena, por lo menos tres veces a la semana, reduce los niveles de colesterol en la sangre. Se da seguimiento a un grupo de voluntarios que afirma que tiene este hábito alimenticio y cada tres meses se toman muestras de sangre para, al cabo de un año, evaluar la afirmación.			
c)	Un fabricante de analgésicos afirma que su medicamento es mucho más efectivo para aliviar los dolores reumáticos. Para probar la efectividad del analgésico, se establece un grupo experimental al cual se le aplicará el nuevo tratamiento y un grupo de control que seguirá tomando la medicación tradicional. A ambos grupos se les dará seguimiento y luego se hará un análisis estadístico de los datos.			
d)	Un grupo de nutricionistas quiere evaluar la relación de la dieta y el nivel de hierro en la sangre. Con ese propósito, seleccionan tres comunidades donde extraerán muestras para efectuar los análisis respectivos. Una muestra se recolectará en la isla de Chira, donde se sabe que la dieta típica cuenta con muchos productos del mar; otra muestra se obtendrá en un caserío de Upala cuya dieta se basa en granos y tubérculos, y la tercera muestra será tomada en Zarcero en razón de basar su alimentación en productos lácteos y hortalizas.			
e)	100 enfermos de tuberculosis de la provincia de San José, de cuyo padecimiento se tenía noticia al mes de octubre de 2012, son sometidos a un chequeo médico para determinar su situación de salud, perfil demográfico y situación socioeconómica.			

1.5. ELEMENTOS POR CONSIDERAR EN UNA INVESTIGACIÓN

El investigador debe identificar la población objeto de estudio, así como la unidad estadística que se observará, las características e información que pueda clasificarse bien como atributo o como variable. Hecho esto, se pasa a decidir si se estudia a toda la población, esto es hacer un censo, o a una muestra; en caso de inclinarse por una muestra, se debe tener en cuenta las técnicas de selección de la muestra y los posibles errores y sesgos que puedan presentarse en el curso de la investigación.

1.5.1. LA POBLACIÓN

La *población* se conforma por el total de elementos con una característica bajo estudio. Así, se tiene que los elementos son las *unidades de estudio*, la presencia o ausencia de la *característica* produce valores numéricos asociados y la población es el conjunto de elementos delimitado en un espacio y un tiempo.

En este sentido, tanto la población como sus componentes dependen de los objetivos de la investigación y de su delimitación espacial y temporal. Si, por ejemplo, la unidad de estudio es la mujer embarazada atendida en la maternidad Carit durante el año 2012, se pueden generar diferentes poblaciones según los objetivos de la investigación: habrá una población multípara, una población con riesgo de aborto, una población por estados civiles, una población por incidencia de enfermedades venéreas y una población expuesta o no a la violencia doméstica.

La población puede ser:

- Finita
- Infinita

Una población *finita* tiene una cantidad de elementos que es posible enumerar, puede ser pequeña o muy grande pero tiene un número limitado de unidades. En tanto que una población *infinita* cuenta con un número ilimitado de elementos o estos son parte de procesos de producción, reproducción o generación de nuevas unidades.

EJERCICIO 1.7

Señale con una X cuáles de las siguientes poblaciones son finitas o infinitas:

		Finita	Infinita
a)	Niños a los que se debe aplicar la vacuna de la poliomielitis.		
b)	Quemados atendidos en el Hospital de Niños en diciembre de 2012.		
c)	Producción de ácido-acetil-salicílico en los laboratorios de la CCSS.		
d)	Pacientes con dengue hemorrágico atendidos en el Hospital de la Anexión durante el invierno de 2010.		
e)	Especialistas en Imágenes Médicas graduados por el convenio CCSS-UCR durante la primera década del siglo XXI.		
f)	Personal de obstetricia graduado en Costa Rica en los últimos 5 años.		
g)	Funcionarios participantes en la última huelga médica, declarada ilegal por los tribunales de trabajo.		
h)	Estudiantes que se matricularán en Microbiología en los próximos años.		
i)	Socorristas voluntarios que actualmente se encuentran acreditados por la Cruz Roja.		

1.5.2. UNIDAD ESTADÍSTICA, CARACTERÍSTICA Y OBSERVACIÓN

Las investigaciones examinan cada *unidad estadística* y registran las *observaciones* que se ajustan a una determinada *característica de interés*.

Por ejemplo, dado que tanto la ley de pensiones como la seguridad social establecen una edad de retiro similar para todos los trabajadores del sector público e igual atención médica, el Director del Cuerpo de Bomberos necesita justificar un régimen especial para el personal de operaciones basado en el tipo de riesgo que corren los bomberos, las edades y su condición de salud.

En este caso, la *unidad estadística* es el bombero, las *características de interés* son la edad, el sexo, el peso, la estatura, los años de servicio, las enfermedades profesionales y los accidentes sufridos durante los incendios o rescates.

Cuadro 1.2

Observación de las características de interés a la Teniente Kembly Alfaro

Unidad estadística	Característica de interés	Observación
El bombero	• Edad	• 24 años
	• Sexo	• Femenino
	• Peso	• 57 kilogramos
	• Estatura	• 168 centímetros
	• Años de servicio	• 3 años
	• Estado civil	• Casada
	• Número de hijos	• 1
	• Tipo (voluntario, asalariado, emérito)	• Asalariado
	• Rango	• Teniente
	• Enfermedades padecidas	
	– Hipertensión	• No
	– Diabetes	• No
	– Respiratorias	• No
	• Número de accidentes reportados	
	– Fracturas	• 1
	– Inhalación de gases	• 3
	– Caídas de más de 2 metros	• 1
	– Quemaduras	• 2

1.5.3. ATRIBUTOS Y VARIABLES

Los datos de una investigación se pueden contar o se pueden medir. Los *atributos* no pueden ser medidos, pero sí pueden ser contados. Por ejemplo, se puede contar el número de bomberos voluntarios, asalariados y eméritos; se puede contar los bomberos rasos, los operadores, los oficiales y los jefes, y se puede saber el número de casados, solteros, etc.

Las *variables* pueden ser medidas en unidades de longitud, peso, temperatura, precio o tiempo. Las variables al ser observadas dan directamente un valor numérico. Las variables son:

- Discretas
- Continuas

Las *variables discretas* generalmente toman valores enteros como el número de hijos o la cantidad de fracturas sufridas en el cumplimiento del deber, pero puede darse el caso de variables discretas que tomen valores fraccionarios, por ejemplo, las notas de aprobación en un curso que pueden ser 7.0, 7.5, 8.0, 8.5, 9.0, 9.5 y 10.0. Para que las notas fraccionarias del curso sean variables discretas, deben haber sido

estipuladas finitamente de esta manera y no será posible encontrar una nota entre 7.0 y 7.5. Precisándolo más, las variables discretas únicamente toman ciertos valores predeterminados en un intervalo.

Las *variables continuas* pueden asumir infinito número de valores en un intervalo; por ejemplo, el tiempo de atención de una emergencia, la temperatura soportada por los equipos, la altura de la ventana por la cual se realizará el rescate o la distancia entre el hidrante y el incendio. Convencionalmente, algunas variables continuas pueden trabajarse como discretas, por ejemplo, cuando la edad del funcionario solamente se registra en años cumplidos, a pesar de que la variable tiempo es de naturaleza continua.

Existen también las variables *dicotómicas* que solamente asumen dos valores, sí o no, verdadero o falso, defectuoso o aceptable, embarazada o no. Estas variables no se pueden medir pero se pueden contar, por lo que reciben un tratamiento similar al de los atributos.

EJERCICIO 1.8

Organice las características de interés y las observaciones del Cuadro 1.2 según sean atributos, variables discretas, continuas o dicotómicas:

ATRIBUTOS

Característica de interés	Unidad de medida	Ejemplo de observación

VARIABLES DISCRETAS

Característica de interés	Unidad de medida	Ejemplo de observación

VARIABLES CONTINUAS

Característica de interés	Unidad de medida	Ejemplo de observación

VARIABLES DICOTÓMICAS

Característica de interés	Unidad de medida	Ejemplo de observación

1.5.4. CENSO Y MUESTRAS

Un *censo* es una investigación sobre las características de *toda* la población, en tanto que una *muestra* estudia a un subconjunto de aquella para luego hacer inferencias o generalizaciones sin tener que examinar uno a uno todos los elementos.

Los términos censo y muestra no son rígidos. Por ejemplo, puesto que todas las embarazadas atendidas en la maternidad Carit durante el año 2012 podrían considerarse una población y su análisis un censo además, de entre ellas se puede escoger una muestra. En otro estudio, estas señoras podrían considerarse una muestra de todas las mujeres embarazadas de Costa Rica en el 2012.

Tampoco es rígida la clasificación de poblaciones finitas e infinitas; por ejemplo, todas las mujeres embarazadas atendidas en la maternidad Carit en 2012 es una población finita pues se puede enumerar, pero se vuelve una población infinita cuando eliminamos la delimitación temporal: "Todas las mujeres embarazas atendidas en la maternidad Carit". Si se elimina tanto la delimitación temporal como espacial, también tendríamos una población infinita: "Todas las mujeres embarazadas".

Se recomienda la realización de un censo cuando:

- La población es pequeña
- El tamaño óptimo de muestra es casi tan grande como la población
- Se presenta mucha variabilidad entre los elementos
- Se busca la exactitud en los resultados
- Es asunto de vida o muerte el examinar a todos los miembros de una población
- Se dispone de los recursos para obtener información completa
- Existen razones de imagen institucional o presiones políticas

Aunque los resultados censales parezcan más confiables, la realidad obliga a examinar muestras puesto que:

- Hay poblaciones infinitas o procesos que no terminan
- En muchos ensayos las unidades se destruyen o transforman
- Los elementos son homogéneos o de poca variabilidad
- El costo de realizar un censo es muy alto
- No hay tiempo para capacitar a muchos tomadores de muestras
- Existe urgencia por conocer los datos o se podría perder la oportunidad de obtener información valiosa
- Se necesita un equilibrio entre tiempo, costo y tipo de información requerida

Con la información que genera una muestra se producen medidas e indicadores muestrales que técnicamente se llaman *estadísticas o estimadores*. En tanto que, con la información censal, se obtienen medidas e indicadores globales llamados *parámetros*.

Una información muestral o estadística se convierte en el *estimador* de un parámetro. Por ejemplo, si se considera a las mujeres embarazadas atendidas en la maternidad Carit en el año 2012 como una buena muestra, el promedio de edad de ellas (estadística) puede ser utilizado para *estimar* el promedio de edad de todas las mujeres embarazadas de Costa Rica en 2012 (parámetro).

1.5.5. SELECCIÓN DE LA MUESTRA

Una muestra puede seleccionarse de forma:

- Intencional
- Convencional
- Aleatoria

El muestreo *intencional* parte del buen juicio, del conocimiento o la experiencia del investigador para identificar los elementos de la población que le pueden aportar mejores datos.

El muestreo *por conveniencia* se basa en la comodidad para estudiar los elementos accesibles, son más fáciles de encontrar, son voluntarios o se muestran más anuentes a colaborar con la investigación.

Una muestra *aleatoria* se basa en el azar y todos los elementos de la población tienen una probabilidad medible (calculable) de ser tomados en cuenta y esta probabilidad es diferente de cero. En el muestreo aleatorio simple, todas las unidades de estudio tienen igual probabilidad de ser incluidos en la muestra [2].

2 En el segundo curso de Estadística para la Salud, se estudian las técnicas de muestreo. De momento se informa que, además del muestreo aleatorio simple, existen muestreos por conglomerados, por estratos y sistemáticos.

EJERCICIO 1.9

Indique la forma de obtención de la muestra en las siguientes investigaciones (intencional, por conveniencia o aleatoria):

		I	C	A
a)	La Presidencia Ejecutiva de la CCSS quiere conocer las necesidades de los servicios de atención de emergencias en el sistema hospitalario nacional y para ello se convoca a reunión al médico-jefe del turno de noche de 8 hospitales: Tony Facio de Limón, San Vicente de Paúl de Heredia, San Rafael de Alajuela, Monseñor Sanabria de Puntarenas, Enrique Baltodano de Liberia, Escalante Pradilla de Pérez Zeledón y San Juan de Dios de la capital.			
b)	El médico de la selección mayor de fútbol masculino solicita un análisis microscópico de sedimento en la orina que informe los eritrocitos y leucocitos por campo de 400x, los filamentos mucosos, células epiteliales, y flora bacteriana de 10 de los 50 preseleccionados. A cada deportista se le asigna un número que se coloca en una caja de la cual un niño extraerá la muestra de 10 jugadores por estudiar.			
c)	El Ministerio de Salud solicita un estudio acerca del cumplimiento a las regulaciones para manipulación de alimentos en las carnicerías del país, este estudio se realizará a una muestra de 1200 carnicerías escogidas por sorteo dentro de una lista de concesionarios y patentados de todo el país.			
d)	La Gerencia Administrativa de la CCSS busca inventariar las necesidades de las unidades de lavandería en el sistema hospitalario nacional y para ello se convoca a reunión a los Jefes de Servicios Generales de los hospitales San Juan de Dios, México y Calderón Guardia.			
e)	El Ministro de Educación quiere saber el nivel de apoyo popular a las restricciones a la venta de productos con alto contenido de azúcar, sal y grasas saturadas en los comedores escolares; para ello se programa una conexión de Internet en la página oficial del Ministerio para que los visitantes puedan manifestarse, si así lo desean, a favor, en contra o indiferentes con las regulaciones.			
f)	La Gerencia Médica de la CCSS necesita evaluar la exposición a las infecciones intrahospitalarias de los trabajadores del Hospital Dr. Max Peralta de Cartago y para ello se estudiarán los registros clínicos del personal, escogiendo uno de cada treinta expedientes de funcionarios administrativos y uno de cada quince expedientes del personal médico y de enfermería.			

1.5.6. ERRORES Y SESGOS

En toda investigación es posible llegar a conclusiones inexactas por la presencia de errores y sesgos. Estos pueden comenzar desde la incorrecta identificación de la población hasta la digitación del informe final y su impresión, pasando por una desatinada definición del problema de investigación, una vaga redacción de los objetivos, un subregistro o sobre-enumeraciones y múltiples problemas técnicos (con programas de cómputo, cálculos matemáticos, métodos estadísticos).

Partiendo de un riguroso diseño de investigación, estos errores y sesgos pueden reducirse, corregirse, evitarse y hasta eliminarse. Lo primero es conocer de su existencia.

Al analizar muestras, siempre estará rondando un *margen de error*, que es el riesgo de que la muestra con que se trabaja no sea representativa. Según Miguel Gómez, este es el "precio que debe pagarse por los beneficios que en costo y rapidez brinda el muestreo" (1998, p. 9).

Cuando el muestreo se realiza intencionalmente o por conveniencia, el error muestral está presente pero no se puede calcular su magnitud. En cambio, cuando la muestra es aleatoria, el error de muestreo se puede calcular, evaluar y hasta reducir por aumento del tamaño de la muestra o por muestras sucesivas en las cuales los errores se alternarán alrededor del parámetro.

Los *sesgos* son inclinaciones sistemáticas en un solo sentido, es decir, reflejan una tendencia por sobreestimación o subestimación insistente y constante. Los sesgos pueden ser:

- De selección
- De medición

Los *sesgos de selección* solo estarán presentes cuando se trabaja con muestras, su mayor incidencia está en la incorrecta delimitación de la población de la cual se toma la muestra y se da en forma de: subregistro, omisiones, no respuesta o grupos no representativos de la población.

Los *sesgos de medición* pueden ocurrir tanto en muestras como en censos, por lo que aumentar el tamaño de muestra no los corrige. La clave de este sesgo es la forma de medir y clasificar a los informantes, así, si se pregunta sobre el ingreso devengado, existe la tendencia a declarar menos ingresos para reducir la cuota de seguridad social y pagar menos; si el médico pregunta sobre la cantidad de cigarrillos diarios que un paciente fuma este miente hacia la baja, o un tensiómetro mal calibrado informará persistentemente de mayores (o menores) niveles de presión arterial.

Si se logra identificar la tendencia y la magnitud de un sesgo de medición, este podría corregirse, por ejemplo: si se identifica que al metro le faltaba un centímetro, pueden ajustarse las estaturas de los pacientes. Pero lo frecuente es que los sesgos pasen inadvertidos y no puedan corregirse satisfactoriamente.

EJERCICIO 1.10

A continuación se presentan varias investigaciones sesgadas, se le pide que redacte una explicación sobre el posible origen del sesgo.

Para analizar la efectividad de una dieta se analizan los resultados de todas las personas que completaron el programa.	Sesgo de selección	
Se quiere comprobar la eficacia de un tratamiento para dejar de fumar y para ello se llama a voluntarios fumadores.	Sesgo de selección	
En un experimento participan dos grupos, uno expuesto a la variable estudiada y otro de control. El investigador principal hace personalmente las mediciones en el grupo experimental y manda a sus asistentes a hacer las mediciones del grupo de control.	Sesgo de medición	
Para estudiar la práctica de medidas de salud reproductiva se envía un cuestionario por la vía del correo electrónico a 2000 informantes.	Sesgo de selección	
En un estudio retrospectivo se pregunta a los participantes sobre los niveles de consumo de un fármaco en diferentes etapas de su vida.	Sesgo de medición	
Con una encuesta se quiere medir la frecuencia de la violencia doméstica y, para ello, se pregunta sobre la cantidad de incidentes y ataques intrafamiliares, además de su periodicidad.	Sesgo de medición	

1.6. OBTENCIÓN DE LA INFORMACIÓN

Una vez que se tiene establecido el propósito de una investigación, surge la pregunta, ¿qué tipo de datos se requiere? Y, seguidamente, ¿esos datos ya existen o se deben recolectar?

Los *datos existentes* pueden provenir de:

- Publicaciones periódicas
- Investigaciones especiales

Las *publicaciones periódicas* existen, en muchos casos, por disposición legal que obliga a recoger y difundir esta información, este es el caso de los Censos Nacionales de Población, las Encuestas de Hogares y los registros administrativos de oficinas gubernamentales como el Registro Civil, el Ministerio de Salud, la Caja Costarricense del Seguro Social y el Instituto Nacional de Seguros. Adicionalmente,

existen organismos internacionales como La Organización Mundial de la Salud y la Organización Panamericana de la Salud, que publican periódicamente estadísticas de salud de cobertura regional y mundial.

Las *investigaciones especiales* se producen en universidades, institutos de investigación, informes de equipos ad-hoc [3], clínicas, hospitales y consultorías contratadas para un tema concreto.

1.6.1. FUENTES INTERNAS O EXTERNAS, PRIMARIAS O SECUNDARIAS

Los datos existentes pueden haberse originado en fuentes internas o externas. Si un equipo de investigadores de la Caja Costarricense del Seguro Social realiza un estudio basándose en datos recopilados por el Departamento de Estadísticas de Salud de su misma institución, se dice que trabajan con fuentes *internas*. En cambio, si tienen que basarse en datos del Instituto Nacional de Seguros o de la Oficina Sanitaria Panamericana se dice que se basan en fuentes *externas*.

Cuando una institución informa solamente datos recogidos por sus propias dependencias, sus publicaciones o informes se consideran *fuentes primarias*. Mientras que se considera *fuente secundaria* a la información proporcionada por agencias que hacen acopio de estadísticas procesadas por otros, la hacen comparable y luego la publican, por ejemplo: el Anuario Estadístico del Ministerio de Salud o el World Health Statistics Report.

Las fuentes secundarias no se limitan a los anuarios y resúmenes estadísticos, puesto que también lo son los datos que se citan de revistas, tesis, informes de labores, consultorías y otras publicaciones.

Gómez Barrantes (1998) indica que es preferible trabajar con fuentes primarias que con secundarias, porque:

a) La fuente secundaria puede contener errores de transcripción cometidos al tomarse la información de la fuente primaria o aún de otra fuentes secundaria.

b) La fuente primaria incluye a menudo definiciones de los términos y unidades usadas.

c) La fuente primaria incluye generalmente un ejemplar del cuestionario y una descripción del procedimiento que se usó al elegir la muestra y al recoger los datos.

d) Una fuente primaria suele proporcionar más detalle. Una fuente secundaria omite a menudo información o combina categorías. (p. 33)

3 Ad-hoc: Equipos organizados para un fin específico, para una tarea especial y que luego de cumplirla se disuelven. Una investigación ad-hoc es específica, solicitada con un propósito particular.

1.6.2. EVALUACIÓN DE LOS DATOS Y LAS FUENTES

Tanto los datos como las fuentes deben evaluarse en al menos cuatro aspectos:

- Pertinencia
- Calidad
- Confiabilidad
- Comparabilidad

La *pertinencia* lleva a cuestionarse si los datos proporcionan información útil y si son aplicables al problema.

La valoración de la *calidad de los datos* implica preguntar si estos son de origen censal o muestral, si la muestra fue aleatoria o por conveniencia y si se trata de estimadores o de parámetros.

La *confiabilidad* pasa por evaluar la imparcialidad de la fuente, la posibilidad de sesgos de medición o de selección, la capacitación o profesionalismo de los captadores del dato bruto y la adecuada formulación del cuestionario, las pruebas o el experimento.

Finalmente, la *comparabilidad* de los datos requiere el conocer la definición usada de población, si las clasificaciones son uniformes y estandarizadas, si se trabajó con métodos similares de recolección y la forma en que se procesó la información.



Figura 1.3. **Evaluación de los datos y las fuentes**

Fuente: **Elaboración con fines didácticos basado en Gallardo (2005)**

1.6.3. RECOLECCIÓN DE DATOS NO EXISTENTES

Si los datos no existen, el investigador los tiene que obtener por medio de los siguientes métodos:

- Observación
- Entrevista
- Registro
- Correo

La *observación* se basa en conteos o mediciones que realiza el mismo investigador en el laboratorio, en el trabajo de campo o por el examen físico directo del paciente.

La observación se vuelve obligatoria cuando la unidad estadística no puede comunicarse adecuadamente, como es el caso de niños y pacientes en estado terminal, o si se trabaja con plantas, animales y cosas (reactivos, medicinas, tejidos). Tiene la ventaja de ser de bajo costo, pero se señala como desventaja el hecho de depender mayoritariamente de la experiencia y el conocimiento del observador.

La *entrevista* es viable cuando la unidad de análisis se puede comunicar verbalmente con el investigador, puesto que tiene que contestar un listado de preguntas plasmadas en un formulario o boleta. El cuestionario puede aplicarse directamente y toca al entrevistador llenarlo a medida que va realizando la entrevista. También puede hacerse en forma autoadministrada, entregando la boleta para ser llenada por el informante en algún lugar que le parezca cómodo. Adicionalmente, esta entrevista podría hacerse en forma telefónica.

El método de entrevista es más caro, requiere de entrevistadores capacitados, criterios de clasificación homogéneos y exige ganarse la confianza del informante para reducir la no respuesta, la hostilidad o la deshonestidad en lo declarado.

El *registro* es muy utilizado en oficinas gubernamentales y en los movimientos hospitalarios. Se basa en el conteo de los hechos conforme se van presentando (ingreso/egreso hospitalario), por obligación de ley (registro de nacimientos, defunciones, naturalizaciones), como requisitos para tramitar indemnizaciones (accidentes de tránsito) o por necesidades evidentes (vigilancia epidemiológica).

El costo adicional de un registro de estadística continua es relativamente bajo para la institución que lo realiza, pero debe garantizarse la confidencialidad, facilidades administrativas (oficinas de primer piso, gratuidad de las boletas, franquicia postal) y campañas de concientización acerca de su utilidad, las leyes que lo obligan y la forma correcta de informar. Todo esto para disminuir el subregistro, las omisiones o duplicaciones de datos, así como las incongruencias. La desventaja de los registros es que permiten obtener solo información específica, en el registro puedo encontrar la cantidad de adolescentes embarazadas, pero no la proporción de estas que son agredidas por sus parejas, por ejemplo.

El *correo* consiste en el envío de un cuestionario por medio de la oficina postal o por Internet. Este cuestionario debe ser llenado por el informante y devuelto a la dirección del equipo investigador. Si bien es uno de los métodos más baratos, su aplicación requiere del conocimiento de las direcciones (físicas o electrónicas) de los informantes, de facilidades para la devolución, interés por participar, que las instrucciones sean claras y detalladas, además se sugiere que el formulario no sea muy extenso.

La mayor debilidad de este método es su alta probabilidad de no respuesta, respuestas incompletas y omisiones; además, no existe posibilidad de retroalimentación en el caso de que no se entienda bien una pregunta. En todo caso, este método es muy útil para estudios exploratorios o para conocer la opinión de una asociación profesional en donde todos sus miembros manejan el tipo de lenguaje utilizado y tienen interés en el tema.

Téngase presente que estos métodos de recolección no son excluyentes y que, más bien, deben combinarse para mejorar la calidad de la información. Además, en todos se pueden utilizar los adelantos técnicos como grabadoras, equipo de video, teléfono, redes sociales, contadores electrónicos, equipos y programas informáticos.

EJERCICIO 1.11

Señale, al menos, dos ventajas y dos desventajas de cada uno de los siguientes métodos de recolección de datos.

OBSERVACIÓN

Ventajas	Desventajas

ENTREVISTA

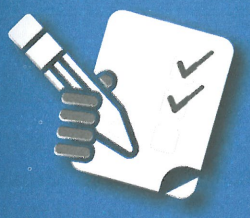
Ventajas	Desventajas

REGISTRO

Ventajas	Desventajas

CORREO

Ventajas	Desventajas



EJERCICIOS DE REPASO Y AUTOEVALUACIÓN

Esta sección pretende que haga un repaso general de los contenidos del Capítulo.

Ejercicio 1.12

Encierre en un círculo la opción que se ajusta al enunciado, Falsa (F) o Verdadera (V).

F V 1. La estadística aplicada se hermana con otras disciplinas para producir instrumentos de trabajo, lo cual genera múltiples herramientas para la economía, la ingeniería, la sociología, la demografía, la biología y la salud.

F V 2. La estadística inferencial se especializa en la recolección, el resumen y la presentación de los datos, destaca sus características y relaciones para hacer la información más comprensible.

F V 3. La estadística inferencial trabaja con muestras para hacer generalizaciones acerca de la población o, partiendo de experimentos, predice el comportamiento de las variables en situaciones parecidas.

F V 4. Una hipótesis es un enunciado que propone una relación entre los datos, una explicación preliminar a un problema, una posible solución o un método de comprobación de las determinaciones entre variables.

F V 5. El mayor aporte de la estadística para la verificación de una hipótesis se hace por medio del análisis cualitativo y utilizando el método deductivo.

F V 6. Los estudios transversales necesitan del transcurso de un tiempo considerable para identificar posibles causas de un fenómeno, el comportamiento de una enfermedad en el tiempo y describir su tendencia.

F V 7. Un estudio prospectivo se realiza con datos del pasado y se basa en registros, expedientes y reportes de laboratorio almacenados en los archivos.

F V 8. En los estudios ecológicos las unidades de estudio son las poblaciones, las comunidades o los grupos específicos.

F V 9. Las investigaciones son cuasi-experimentales cuando la variable en estudio es manipulada por el investigador.

F V 10. Una población se conforma por el total de elementos con una característica bajo estudio.

F V 11. Los atributos pueden ser medidos en unidades de longitud, peso, temperatura, precio o tiempo.

F V 12. Las variables discretas pueden asumir infinito número de valores en un intervalo.

F V 13. Las variables dicotómicas solamente asumen dos valores: sí o no, verdadero o falso, defectuoso o aceptable.

F V 14. Un censo es una investigación sobre las características de un subconjunto de la población.

F V 15. Los sesgos son inclinaciones sistemáticas en un solo sentido, es decir, reflejan una tendencia por sobreestimación o subestimación insistente y constante.

F V 16. Los sesgos de medición pueden corregirse con un aumento del tamaño de la muestra.

F V 17. Es preferible trabajar con fuentes secundarias puesto que han sido sometidas a un doble proceso de evaluación.

F V 18. La valoración de la calidad de los datos implica preguntar si estos son de origen censal o muestral, si la muestra fue aleatoria o por conveniencia y si se trata de estimadores o de parámetros.

F V 19. La comparabilidad de los datos requiere conocer la definición usada de población, si las clasificaciones son uniformes y estandarizadas, si se trabajó con métodos similares de recolección y la forma en que se procesó la información.

F V 20. Las encuestas por correo siempre tienen un alto porcentaje de respuestas.

Ejercicio 1.13

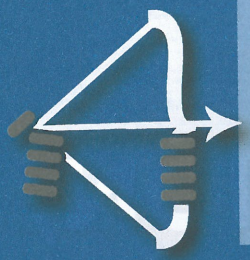
Pareo. Llene los paréntesis con el número que acompaña a cada definición de la izquierda, de forma que cada definición coincida con su concepto. Quedan dos paréntesis en blanco.

1.	Es la ciencia que trata del desarrollo y aplicación de las técnicas de recolección, organización, presentación, análisis e interpretación de datos numéricos obtenidos por observación o experimentación, para obtener conclusiones válidas y fundamentar la toma de decisiones.	( )	Entrevista
2.	Estudios que pretenden establecer conexiones entre variables, medir su correlación o nivel de asociación, comprobar su causalidad, plantear escenarios futuros, medidas preventivas y las consecuencias sobre la salud.	( )	Hipótesis
3.	Es una investigación sobre las características de toda la población.	( )	Muestra
4.	Método de recolección de datos en el cual el informante debe contestar preguntas o llenar un cuestionario.	( )	Censo
5.	Es el riesgo de que la muestra con que se trabaja no sea representativa y surge por la variabilidad del muestreo.	( )	Fuentes primarias
6.	Características de las unidades de estudio que pueden ser medidas en su longitud, peso, temperatura, precio o duración en el tiempo.	( )	Fuentes secundarias
7.	Datos, documentos e informes generados por la misma institución o sus dependencias.	( )	Variables
8.	Enunciado que propone una relación entre los datos, una explicación preliminar a un problema, una posible solución o un método de comprobación de las determinaciones entre variables.	( )	Estadística
9.	Subconjunto de la población que se examina para observar las características de interés.	( )	Error de muestreo
10.	Conteo de los hechos conforme se van presentando, por obligación de ley, como requisitos para tramitar indemnizaciones o por necesidades justificadas.	( )	Descriptivos
		( )	Analíticos
		( )	Registro

CAPÍTULO 2

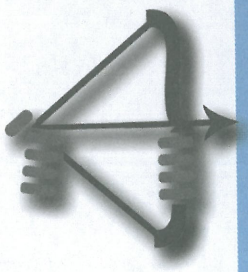
ELABORACIÓN, PRESENTACIÓN Y ANÁLISIS ESTADÍSTICO DE LOS DATOS

OBJETIVO GENERAL



Adquirir conocimientos sobre los elementos básicos relacionados con la elaboración de datos estadísticos para su presentación, análisis e interpretación en el campo de la salud.

OBJETIVOS ESPECÍFICOS



Al finalizar el estudio de este capítulo usted estará en la capacidad de:

1. Elaborar un instrumento de recolección de los datos para su posterior análisis e interpretación.
2. Señalar cuáles preguntas versan sobre hechos, comportamiento, intención u opinión.
3. Cambiar preguntas abiertas a formato de preguntas cerradas o semiabiertas.
4. Corregir enunciados negativos, redactándolos afirmativamente.
5. Transformar preguntas con eufemismos y tecnicismos a un lenguaje estándar.
6. Distinguir categorías mutuamente excluyentes y colectivamente exhaustivas.
7. Confeccionar cuadros y distribuciones de frecuencias para la presentación de los informes de investigación.

Las investigaciones científicas pasan por etapas preparatorias, de ejecución, procesamiento y análisis. Para la obtención de los datos y su posterior interpretación, es necesaria la construcción de instrumentos de recolección eficientes, de manera que posibiliten comunicar los hallazgos más importantes en distribuciones de frecuencias simples, acumuladas y relativas.

Las preguntas adecuadas
constituyen la mitad de la ciencia.
ROGER BACON

La investigación científica ha ido desarrollando una serie de técnicas e instrumentos para la recolección de los datos, la aplicación de instrumentos de captura, validación, interpretación y análisis. Es necesario insistir en que los hallazgos seguirán en la oscuridad si no se presentan adecuadamente y se comunican en forma de textos, tablas o gráficos que posibiliten la socialización con la comunidad y sirvan de base para posteriores avances. Por ello, en este capítulo, se presenta el camino general que sigue una investigación desde la identificación del problema hasta la presentación del informe final.

2.1. ETAPAS DE UNA INVESTIGACIÓN

En toda investigación se distinguen claramente tres etapas:

- Preparatoria
- De ejecución
- De procesamiento y análisis

Cada etapa incluye un conjunto de tareas científicas y tareas administrativas que avanzan en forma concomitante, se retroalimentan y se apoyan mutuamente.

2.1.1. ETAPA PREPARATORIA



Figura 2.1. **Tareas de la etapa preparatoria de una investigación**

Fuente: Elaboración propia con base en Gallardo (2005)

En la *etapa preparatoria*, las tareas *sustanciales o científicas* son:

- Identificación del problema
- Revisión del estado de la cuestión
- Establecimiento de los objetivos
- Construcción de los instrumentos de recolección de datos
- Selección de la muestra, (si no se trabaja con la población)

2.1.1.1. Identificación del problema

Todo proyecto de investigación comienza con la identificación de un problema, esto es, con la búsqueda de una respuesta a la pregunta ¿qué ignoramos? o ¿qué se necesita conocer a profundidad?

Esta interrogante puede ser subdividida en varias preguntas: ¿existe ese problema?, ¿vale la pena investigarlo?, ¿cómo se puede plantear y justificar?, ¿cuál es su importancia?, ¿cuál es la población y cuál la unidad de estudio?, ¿dónde se obtiene la información relevante?, ¿está disponible o hay que recolectarla?, ¿cómo se delimita este problema en el tiempo y en el espacio?, ¿la investigación será descriptiva o analítica?, ¿transversal o longitudinal?, ¿individual o ecológica?, ¿observacional o experimental?, ¿es factible de investigar?

Una buena identificación del problema permite establecer el rumbo de la investigación y especificar una pregunta que demanda respuesta, a la vez que economiza recursos financieros, materiales y humanos, y reduce esfuerzos y tiempo.

2.1.1.2. Estado de la cuestión

Una vez identificado el problema de investigación, se deben revisar los escritos acerca del tema o estudios relacionados, la experiencia de otros especialistas sobre asuntos parecidos y los resultados obtenidos en estudios similares publicados en:

- Revistas especializadas
- Libros
- Tesis de grado
- Informes de investigación
- Casos clínicos
- Ponencias en seminarios y congresos

Esta revisión ayuda a solucionar problemas teóricos y metodológicos, a conocer lo que existe sobre el tema y determinar qué se puede agregar, además de dar luz sobre dificultades prácticas que puedan presentarse más adelante. A este respecto, Gallardo (2005) señala que toda investigación:

> Parte de ciertos conocimientos o informaciones y de determinados procedimientos de trabajo (ya sea para apoyarse críticamente en ellos, ya sea para refutarlos) que han sido propuestos por otros investigadores. En cuanto esos otros investigadores se han ocupado también y previamente de lo que consideramos nuestro objeto de estudio resulta pertinente conocer y reseñar, por escrito, lo que esos otros investigadores académicos han encontrado y también qué criterios han empleado para encontrar lo que encontraron.
>
> A la búsqueda de lo que otros han producido como conocimiento sobre mi objeto de trabajo o tema-problema, y al informe escrito que consuma esta indagación, los consideramos como los elementos que configuran la realización del **estado de la cuestión** (p. 125).

2.1.1.3. Establecimiento de los objetivos

Teniendo delimitado el problema y revisados los estudios previos, se puede entonces tener una idea clara de la tarea por realizar y se procede a establecer los objetivos. Moya (2009) indica que:

> Los objetivos describen el resultado que se espera de la investigación; deben expresarse en forma afirmativa, utilizando verbos en infinitivo y estableciendo claramente las restricciones que el estudio tenga respecto a tiempo, área geográfica, población y características por investigar (p. 20).

El establecimiento de los objetivos permite señalar si se va a investigar globalmente el problema o solamente una parte, definir el grado de profundidad que tendrá, decidir si el estudio será evaluativo o si se buscarán relaciones de causalidad, plantear si se dedicará a la búsqueda de los parámetros o si se probará alguna hipótesis.

2.1.1.4. Construcción de los instrumentos de recolección de datos

Los datos son atributos o características que describen hechos y pueden representarse en forma numérica, alfabética u otras formas simbólicas. Dato no es sinónimo de información, pero puede llegar a serlo, toda vez que:

- Esté claramente definido o se pueda saber qué es
- Pueda ser observado en sus manifestaciones físicas
- Sea medible cuantitativamente
- Se pueda categorizar cualitativamente
- Tenga atributos o sea variable

Los datos permiten responder a la pregunta del problema de investigación, pero no son equivalentes a los resultados o las conclusiones de esta, les hace falta un proceso de análisis, interpretación y sustento teórico que les dará el investigador para convertirlos en información y, eventualmente, en conocimiento científico.

Para la recolección de los datos se cuenta con instrumentos directos e indirectos. Los instrumentos *directos* obtienen el dato en la fuente, entre ellos están los experimentos, el registro y la observación, tanto cuantitativa como cualitativa. Los instrumentos *indirectos* permiten recolectar los datos utilizando intermediarios, por ejemplo, a través de entrevistas, encuestas, grupos focales, análisis de contenido o en fuentes secundarias.

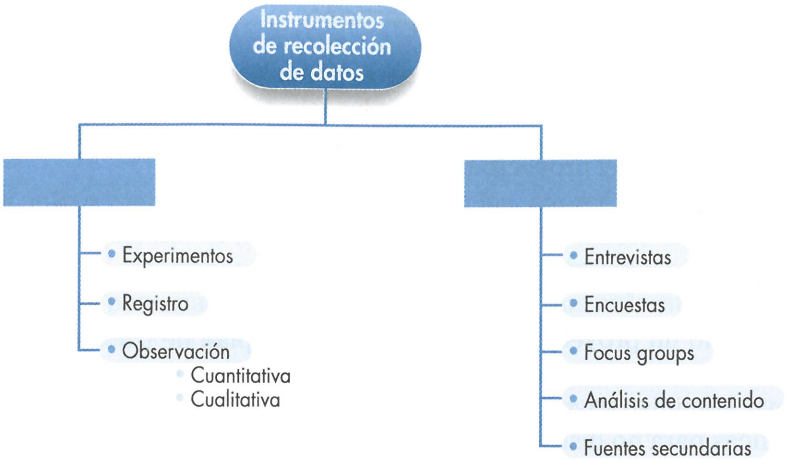


Figura 2.2. **Tipos de instrumentos de recolección de datos para una investigación**

Fuente: Elaboración propia con base en Gallardo (2005)

Cualquiera que sea el instrumento de recolección de datos utilizado, deberá garantizarse que cumple con tres requisitos:

- Confiabilidad
- Validez
- Objetividad

La *confiabilidad* del instrumento exige que si se aplica en forma repetida al mismo informante en condiciones similares, siempre produzca datos iguales o equivalentes. Así, si se mide la presión arterial a un paciente con un tensiómetro que genera 110/70 en un brazo y al minuto siguiente 140/95 en el otro, este instrumento no será confiable.

La *validez* es la convicción de que un instrumento de recolección de datos mide lo que busca medir, por ejemplo, si se quiere saber la presión arterial y el tensiómetro indica 60 pulsaciones por minuto en cada uno de los brazos, ese instrumento podrá ser confiable, pero no es válido puesto que no mide lo que se pretende medir.

La *objetividad* busca eliminar la presencia de sesgos o la tendencia a influenciar los resultados por parte del equipo de recolección, de los que procesan los datos o del propio investigador.

2.1.1.5. Tareas administrativas en la preparación

Paralelamente, en la etapa preparatoria se requiere poner atención a algunas *tareas administrativas* que dan soporte y fluidez a las labores sustantivas. Entre ellas están:

- La elaboración del proyecto o plan de trabajo
- La aprobación o consecución de recursos
- La organización del equipo

Al tener elaborado un proyecto y su correspondiente plan de trabajo, se procede a la consecución de los recursos presupuestarios que darán soporte a la investigación junto con un apoyo administrativo que permitirá a los investigadores concentrarse en las tareas científicas para no distraer tiempo y esfuerzos en trámites burocráticos.

2.1.2. ETAPA DE EJECUCIÓN

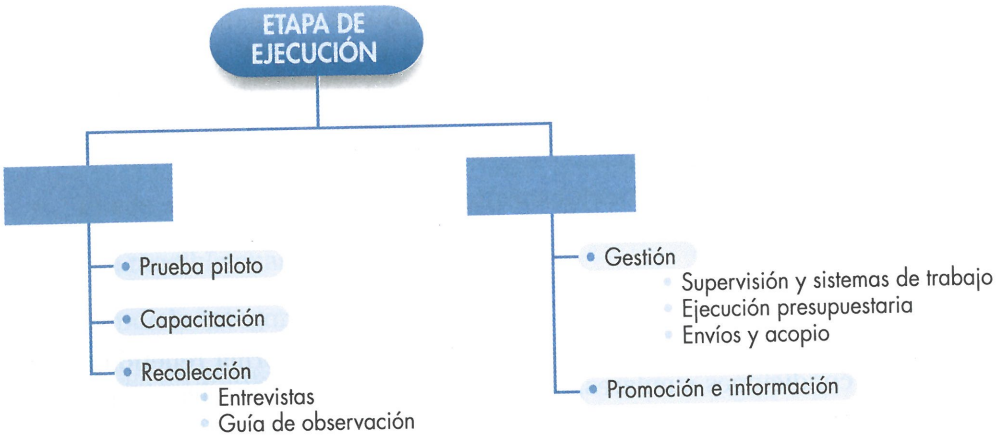


Figura 2.3. **Tareas de la etapa de ejecución de una investigación**

Fuente: Elaboración propia con base en Gallardo (2005)

En la *etapa de ejecución*, las tareas *sustanciales o científicas* son:

- La aplicación de una prueba piloto
- La selección y capacitación del personal que hará el trabajo de campo
- El trabajo de recolección de los datos

2.1.2.1. Prueba piloto

La implementación de una prueba piloto es muy útil para evaluar el instrumento de recolección de los datos, especialmente cuando se trata de un diseño experimental, de una guía de observación o cuando se busca comprobar la claridad de un cuestionario.

Una prueba piloto siempre ayudará a corregir las imprecisiones del instrumento, además sirve para mejorar el nivel de comunicación por medio de la adecuación del lenguaje al nivel del informante. La prueba piloto también evidencia las necesidades de capacitación del personal en los diferentes aspectos que enfrentará en el trabajo de campo y en su relación con los encuestados.

2.1.2.2. Selección y capacitación de los entrevistadores

Tiene que establecerse un perfil profesional del funcionario que realizará el trabajo de campo, en la mayor parte de los casos hasta se demanda formación académica: laboratoristas, enfermeros, médicos, microbiólogos, maestros o estudiantes universitarios, esto porque la labor de recopilación de datos es delicada, seria y acuciosa. El personal, además de estar motivado, debe conocer la importancia de la investigación que se está realizando y tener capacidad para ganarse la confianza de los informantes, máxime cuando se trata de temas de salud en donde la información es íntima.

Además del perfil y otros filtros de selección, a los encuestadores debe dárseles cursos de capacitación específica para aclarar múltiples aspectos de la investigación y del instrumento de recolección a fin de eliminar o, al menos, minimizar los errores y sesgos en la recopilación de los datos.

2.1.2.3. Tareas administrativas en la ejecución

Concomitantemente, se necesita poner atención a múltiples *tareas administrativas* que hacen más fluida la investigación, entre ellas están:

- El diseño de sistemas de trabajo
- El control de la ejecución presupuestaria
- Los métodos de acopio, traslado y protección de los datos recolectados
- Las relaciones públicas para promocionar e informar sobre los avances

Las tareas administrativas incluyen la conformación de equipos de trabajo, los mecanismos de supervisión, el planeamiento de giras de campo, el pago de viáticos, salarios, combustible y otros gastos, y el diseño de procedimientos para hacer llegar los formularios desde el campo hasta las oficinas en las cuales se procesarán.

Gómez (1998) aconseja:

> Esta labor implica numerosas decisiones administrativas y una serie de cuidados que muchos investigadores tienden a subestimar; es conveniente advertir, sin embargo, que aun en encuestas pequeñas una mala organización del trabajo de campo puede llevar a datos de mala calidad y a errores difíciles de corregir posteriormente, así como a una elevación significativa de los costos (p. 39).

Algunos trabajos de investigación en el área de la salud son silenciosos, pero con frecuencia se necesita sensibilizar a la población o informar sobre su existencia, buscando reducir la desconfianza del informante y ganarse el apoyo de los líderes religiosos, políticos o comunales. Alguna información puede comunicarse por medios gratuitos (programas radiales, noticias, púlpitos) pero otras veces se requerirá de gastos en afiches, volantes, pautas publicitarias y otros medios que deben ser presupuestados.

2.1.2.4. El trabajo de recolección de los datos

Cuando se trabaja con cuestionarios, la recolección de datos puede hacerse de dos maneras: autoadministrado o por entrevista. En el primer caso, la boleta se envía (por correo o Internet) o se lleva (a la casa, oficina o a un evento) para que el informante la llene, luego se recoge y se verifica que se completó. En el segundo caso, la boleta es llenada por un encuestador que pregunta y anota directamente, tanto si la entrevista se hace cara a cara, por teléfono u otro medio de comunicación en línea (chat).

El método autoadministrado de recolección es más barato y de aplicación sencilla, además el informante puede llenar la boleta en el momento que considere adecuado, en un lugar tranquilo, y tiene tiempo para meditar su respuesta; sin embargo, el porcentaje de respuestas es menor, puede venir incompleta y no se tiene certeza de quién lo llenó. La entrevista es más cara pero reduce las omisiones y errores de interpretación por parte del entrevistado, se tiene certeza de que quien brindó la información es la persona que realmente se quiere entrevistar, se cuenta con un grupo de encuestadores entrenados y capacitados que además es objeto de supervisión.

2.1.3. ETAPA DE PROCESAMIENTO Y ANÁLISIS



Figura 2.4. **Tareas de la etapa de procesamiento y análisis en una investigación**

Fuente: Elaboración propia con base en Gallardo (2005)

En la *etapa de procesamiento y análisis*, las tareas *centrales* son:

- El procesamiento de los datos
- El análisis e interpretación
- La preparación del informe de investigación

2.1.3.1. El procesamiento de los datos

Recopilada la información, se procede a revisar las boletas y, si no estaban previamente codificados, a codificar los diferentes ítems para facilitar el trabajo informático de los datos. El procesamiento implica también la evaluación crítica de la cobertura de la muestra o del censo.

2.1.3.2. El análisis e interpretación

Con los datos digitados, se construyen bases de datos, se elaboran tablas, distribuciones de frecuencia y otros informes (Capítulo II), se determinan razones, proporciones, tasas e indicadores (Capítulo III), se calculan las medidas de posición y de variabilidad (Capítulo IV), se evalúan los servicios y se hace análisis gráfico (Capítulo V), además de proyectar a la población, al futuro y generar probabilidades de ocurrencias (Capítulo VI).

La interpretación sobrepasa al simple análisis estadístico puesto que se requiere de la formación específica del investigador (médico, microbiólogo, epidemiólogo), así como de la experiencia, cuidado, habilidad interpretativa dentro de un marco teórico y metodológico validado por la disciplina o que rompe esquemas para proponer nuevos paradigmas.

2.1.3.3. El informe de investigación

El cuerpo central del informe lo constituyen los hallazgos de la investigación: los resultados. Al redactar el informe, siempre debe tenerse en cuenta al público al que va dirigido: colegas, académicos, patrocinadores, lector promedio, etc. En términos generales, el informe tiene una introducción, señala el problema y los objetivos planteados, la metodología aplicada, los hallazgos, conclusiones y recomendaciones.

Entre las *tareas administrativas* que coadyuvan con esta parte final de la investigación se encuentran:

- La liquidación presupuestaria
- La publicación del informe
- El cierre del proyecto

Esto es importante porque, en un alto porcentaje, el respaldo a las futuras investigaciones depende de las *publicaciones* efectivas que puedan presentar los miembros del equipo investigador y de que no queden dudas sobre la ejecución presupuestaria y el destino de los dineros que facilitó la institución patrocinadora, sea universitaria, instituto de investigación, fundación, empresa, órgano del Estado u organismo internacional.

2.2. DISEÑO DE CUESTIONARIOS

Al confeccionar el cuestionario, tienen que tomarse en cuenta los siguientes aspectos:

- El nivel de dificultad
- La confidencialidad de la información
- El grado de intimidad o presencia de terceros
- La edad y sexo del informante
- Los valores del entrevistado
- No provocar desconfianza o miedos
- La apariencia y acreditación del entrevistador
- El medio por el que se aplicará el cuestionario
- La extensión del cuestionario
- El orden de las preguntas
- La redacción
- El tipo de lenguaje utilizado

Entre las funciones del cuestionario se encuentran: el obtener datos para cumplir los objetivos planteados por la investigación y el guiar al entrevistador para generar una buena relación con el informante.

Al confeccionar el cuestionario, es prudente que no se confunda la pregunta de investigación con las preguntas del cuestionario. La pregunta de investigación es general y muchas veces abstracta, por lo que no puede plantearse directamente y debe desglosarse en varias preguntas accesibles al nivel de comprensión del entrevistado.

Si la pregunta de investigación es, por ejemplo, ¿es aceptable el nivel de atención en esta clínica?, esa interrogante de investigación deberá traducirse en varias preguntas que informen del nivel de atención en esa institución: la conveniencia de los horarios para sacar citas, los tiempos de espera, la comodidad en las salas de atención, el amueblamiento, la iluminación, la temperatura del local, la existencia de rampas y pasamanos, las salidas de emergencia, los pisos resbalosos, las zonas de seguridad, la limpieza, tipo y cantidad de servicios sanitarios, el trato con el personal médico, el tiempo para despacho de las recetas, la eventual sobrepoblación de pacientes, la rotulación de oficinas y rutas de acceso.

2.2.1. CLASES DE PREGUNTAS

Según la naturaleza de las respuestas que se solicitan, las preguntas fundamentales del cuestionario pueden ser:

- De hechos
- De comportamiento
- De intención
- De opinión

Las preguntas accesorias cumplen una función indirecta y son:

- De calentamiento
- De filtro
- Tipo test

De acuerdo con el grado de libertad de la respuesta, las preguntas son:

- Cerradas
- Abiertas
- Semiabiertas

En las preguntas de *hechos*, se quiere saber algo concreto, preciso, tangible, fácil de responder, por ejemplo:

¿Cuántas veces ha estado hospitalizado en el último año?

¿Tiene problemas dentales que le hacen difícil comer?

¿Con qué frecuencia lo visitan sus familiares en el asilo de ancianos?

¿Ha utilizado dinero de la empresa para jugar en el casino?

¿Cuánto tiempo pasa entre el levantarse y fumar el primer cigarrillo?

En las preguntas de hechos, generalmente, el porcentaje de respuestas obtenido es grande y el informante tiende a contestar con sinceridad, salvo si sospecha que su respuesta puede acarrearle una consecuencia negativa: reducción del tiempo de incapacidad, aumento de la cuota obrero-patronal, cobro de más impuestos, sanciones.

Las preguntas de *comportamiento* buscan conocer si se ha realizado tal o cual acción, o qué se hizo frente a tal situación, por ejemplo:

Cuando está en el casino, ¿se empeña en seguir jugando hasta ganar?

¿Se ha asustado o ha tenido pánico sin motivo aparente?

¿Cuántas pastillas auto-recetadas toma diariamente?

¿Qué hace cuando no puede conciliar el sueño?

¿Le remuerde la conciencia después de beber en exceso?

Las respuestas obtenidas en estas preguntas son concretas y precisas, salvo que existan problemas de memoria o el informante se avergüence por confesar su comportamiento.

Las preguntas de *intención* indagan cómo se actuaría si se presentara tal situación, es decir, preguntan sobre algo hipotético o imaginado; el entrevistado no dice cómo se comportó sino cómo se comportaría frente a tal situación, así:

¿Le gustaría dejar de fumar?

¿En qué piensa emplear su tiempo tras la aprobación de su pensión?

¿Desearía delegar el cuidado de su abuela en otra persona?

¿Le parece que es mejor "quitarse de en medio"?

¿Usaría técnicas de relajación en lugar de pastillas para dormir?

Para el entrevistado, este tipo de preguntas implica un esfuerzo adicional, dado que es más fácil recordar un hecho que imaginar una situación probable. Además, la respuesta no asegura que la acción se realizará, apenas revela una intención o deseo cuya ejecución puede cambiar frente a la situación concreta.

Las preguntas de *opinión* piden al informante que diga lo que piensa, opina o cree, no necesariamente una actuación específica, tampoco indaga qué conoce ni cómo se comporta ni qué haría en tal situación, por ejemplo:

¿Cree que está gorda aunque otros digan que está demasiado delgada?

¿Cómo valora la atención que se brinda al adulto mayor en esta clínica?

¿Piensa que su abuelo pide más ayuda de la que realmente necesita?

¿Considera que su vida social ha desmejorado por tener que cuidar a su madre?

¿Cuál hospital cree que está mejor equipado?

La respuesta es menos precisa pues obedece a subjetividades; además, está más afectada por sesgos, el acomodo a conveniencia de la declaración, falta de objetividad y poca sinceridad; sin embargo, revela la percepción mayoritaria de los informantes.

EJERCICIO 2.1

Señale si las siguientes preguntas son de hechos, comportamiento, intención u opinión, marcando con una X según corresponda.

		Hechos	Comp.	Intención	Opinión
a)	¿Comería más frutas y verduras?				
b)	¿A qué edad comenzó a fumar?				
c)	¿Fuma aunque esté enfermo?				
d)	¿Sus preocupaciones le han hecho perder el sueño?				
e)	¿Cuál candidato presidencial cree que mejorará los servicios de salud pública?				
f)	¿Intentaría dejar de apostar?				
g)	¿Recientemente ha perdido más de 6 kilogramos en un periodo de seis meses?				
h)	¿Se siente mejor atendido en el hospital o en el EBAIS?				
i)	¿Qué especialidad médica escogerá cuando se gradúe?				
j)	¿Por quién votó en las pasadas elecciones?				
k)	¿Cuántos miembros conforman su hogar?				
l)	¿Cuántos de ellos trabajan?				

Las preguntas *de calentamiento* son poco relevantes para el problema de investigación, pero cumplen un rol introductorio y ayudan a "romper el hielo" con el entrevistado facilitando la comunicación. Con frecuencia la respuesta es obvia, pero introducen y motivan a los participantes; sin embargo, un buen cuestionario no debe incluir muchas preguntas de calentamiento puesto que consumen tiempo, cansan y distraen del tema investigado.

Las preguntas *filtro* ayudan a clasificar a los informantes y reducen el tiempo de aplicación de un cuestionario al eliminar un grupo de ítems que no son aplicables en el caso concreto, por ejemplo:

Pregunta 3. ¿Intentará dejar de fumar en las próximas dos semanas?

1. ❑ No (Agradezca y finalice la entrevista)
2. ❑ Dudoso (Pase a la pregunta 8)
3. ❑ Probable (Pase a la pregunta 12)
4. ❑ Sí (Continúe con la pregunta siguiente)

EJERCICIO 2.2

En el cuestionario del Censo 2011, se realizaron las siguientes preguntas sobre la etnia o grupo cultural al que considera pertenecer el entrevistado.

7. ¿[Nombre] se considera indígena?

SI ❑ 1 NO ❑ 2 → Pasea 10

8. ¿A qué pueblo indígena pertenece?

Bribri	❑ 1	maleku o buatuso	❑ 6
Brunca o Boruca	❑ 2	Ngöbe o Guaymí	❑ 7
Cabécar	❑ 3	Teribe o Térraba	❑ 8
Chorotega	❑ 4	De otro país	❑ 9
Huetar	❑ 5	Ningún pueblo	❑ 10

9. ¿Habla [Nombre] alguna lengua indígena indígena?

SI ❑ 1 → Pasea 10

NO ❑ 2

10. ¿[Nombre] se considera:?

- Negro(a) o Afrodescendiente ❑ 1
- Mulato(a) ❑ 2
- Chino ❑ 3
- Blanco(a) o Mestizo(a) ❑ 4
- Otro ❑ 5
- Ningún ❑ 6

Figura 2.5. **Preguntas del Censo 2011 sobre etnicidad**

Fuente: INEC (2011)

De ese apartado, señale las preguntas filtro y explique cómo reducen el número de ítems y disminuyen el tiempo de aplicación de la entrevista.

Las preguntas *tipo test* permiten establecer un puntaje o índice que revela lo que para el informante no es evidente o no se atreve a confesar directamente. Al factor evaluado se llega por un rodeo de preguntas que en forma individual parecen de poca trascendencia pero que juntas configuran un hecho, una actitud o un padecimiento del entrevistado, por ejemplo el "Índice de Esfuerzo del Cuidador" de Moral Serrano y otros (2003):

- *Población diana*: Población cuidadora de personas dependientes en general. Se trata de una *entrevista semiestructurada* que consta de 13 ítems con respuesta dicotómica Verdadero – Falso. Cada respuesta afirmativa puntúa 1. Una puntuación total de 7 o más sugiere un nivel elevado de esfuerzo.

- *Instrucciones para el profesional*: Voy a leer una lista de cosas que han sido problemáticas para otras personas al atender a pacientes que han regresado a casa tras una estancia en el Hospital ¿Puede decirme si alguna de ellas se puede aplicar a su caso? (aporte ejemplos).

1	Tiene trastornos de sueño (Ej.: porque el paciente se acuesta y se levanta o pasea por la casa de noche)	Sí	No
2	Es un inconveniente (Ej.: porque la ayuda consume mucho tiempo o se tarda mucho en proporcionar)	Sí	No
3	Representa un esfuerzo físico (Ej.: hay que sentarlo, levantarlo de una silla)	Sí	No
4	Supone una restricción (Ej.: porque ayudar limita el tiempo libre o no puede hacer visitas)	Sí	No
5	Ha habido modificaciones en la familia (Ej.: porque la ayuda ha roto la rutina o no hay intimidad)	Sí	No
6	Ha habido cambios en los planes personales (Ej.: se tuvo que rechazar un trabajo o no se pudo ir de vacaciones)	Sí	No
7	Ha habido otras exigencias de mi tiempo (Ej.: por parte de otros miembros de la familia)	Sí	No
8	Ha habido cambios emocionales (Ej.: causa de fuertes discusiones)	Sí	No
9	Algunos comportamientos son molestos (Ej.: la incontinencia, al paciente le cuesta recordar las cosas, el paciente acusa a los demás de quitarle las cosas)	Sí	No
10	Es molesto darse cuenta de que el paciente ha cambiado tanto comparado con antes (Ej.: es un persona diferente de antes)	Sí	No
11	Ha habido modificaciones en el trabajo (Ej.: a causa de la necesidad de reservarse tiempo para la ayuda)	Sí	No
12	Es una carga económica	Sí	No
13	Nos ha desbordado totalmente (Ej.: por la preocupación acerca de persona cuidada o preocupaciones sobre cómo continuar el tratamiento)	Sí	No
	PUNTUACIÓN TOTAL		

En las *preguntas cerradas*, las alternativas de respuesta se han limitado previamente; estas pueden ser dicotómicas: falso o verdadero, positivo o negativo, de acuerdo o en desacuerdo, sí o no (como en el test anterior), o pueden presentar varias alternativas, por ejemplo:

PREGUNTA 2. ¿Cuál es su estado civil?

1. ❑ Soltero
2. ❑ Casado
3. ❑ Viudo
4. ❑ Divorciado
5. ❑ Unión libre

PREGUNTA 3. ¿Cuál es su nivel educativo máximo alcanzado?

1. ❑ Sin estudios formales
2. ❑ Primaria incompleta
3. ❑ Primaria completa
4. ❑ Secundaria incompleta
5. ❑ Secundaria completa
6. ❑ Universitaria sin graduar
7. ❑ Grado universitario
8. ❑ Posgrado

Las preguntas cerradas se formulan con rapidez, son fáciles de registrar y están codificadas de antemano, pero el entrevistado puede contestarlas a la carrera sin que tenga una idea clara del tema que se está investigando ni se sienta obligado a reflexionar sobre una respuesta precisa o, incluso, puede tener una posición diferente de las alternativas proporcionadas y contestar para salir del paso.

Las *preguntas abiertas* permiten contestar libremente, se utilizan cuando la respuesta no es previsible o cuando hay muchas categorías que incluirlas a todas implicaría un cuestionario con muchas hojas, por ejemplo:

PREGUNTA 1. ¿Cuál es el diagnóstico de ingreso del (la) paciente?

_____

Las preguntas abiertas también permiten una respuesta espontánea y posibilitan que el entrevistado se exprese en sus propios términos, desde sus vivencias o según su perspectiva de análisis, así:

PREGUNTA 7. ¿Con qué tareas podría colaborar esta comunidad para reducir la contaminación?

_____

Frecuentemente, se recurre a preguntas abiertas cuando el tema es novedoso, complejo o la investigación está en una etapa exploratoria donde las categorías no están bien definidas. Entre las desventajas de estas preguntas está la dificultad para la codificación y clasificación dada la variedad de respuestas, también demandan referentes teóricos que permitan ordenarlas en categorías útiles para un adecuado procesamiento y análisis estadístico.

Una solución intermedia la brindan las preguntas semiabiertas que establecen un listado de opciones cerradas, pero dejan una alternativa abierta que muchas veces propone:

7. Otra

   Especifique: _____

O plantean una pregunta adicional, por ejemplo:

PREGUNTA 11. En el conflicto de poder entre el Ministerio de Salud y la Caja Costarricense del Seguro Social, ¿cómo valora las decisiones de la presidenta Laura Chinchilla?

1. ❑ Excelentes
2. ❑ Buenas
3. ❑ Regulares
4. ❑ Malas
5. ❑ Deficientes

¿Por qué las valora así?

_____

EJERCICIO 2.3

A continuación encontrará algunas preguntas abiertas y se pide que las transforme en preguntas cerradas o semiabiertas.

a) ¿Edad en años cumplidos?

_____

b) ¿Cuál es su ocupación principal?

_____

c) Si es estudiante, ¿dónde estudia?

_____

d) Si desempeña alguna actividad remunerada, ¿cuál es?

_____

e) ¿Cuál es su ingreso mensual aproximado?

_____

No puede predefinirse el *número de preguntas*, pues si son pocas se corre el riesgo de desaprovechar las entrevistas y no alcanzar los objetivos de la investigación; pero si se plantean muchas, el informante se cansa y podría desistir de seguir colaborando o comenzar a falsear las respuestas para terminar rápido. La regla general es que si las respuestas son cortas, predeterminadas y fáciles de responder, se pueden plantear muchas; pero si son de respuesta larga, abiertas y de difícil respuesta, el cuestionario deberá ser breve.

El *orden en que se plantean* afecta el nivel de confianza/desconfianza del informante y evita la contaminación de la respuesta. Por ello, las preguntas delicadas, comprometedoras o con probabilidad de ser vistas como amenazantes no pueden ser planteadas al inicio de la entrevista, por ejemplo, si la primera o segunda pregunta fuera: "¿Está al día con las cuotas obrero-patronales de la CCSS?", la investigación podría abortarse. En este caso, conviene comenzar con preguntas de calentamiento, redactar preguntas tipo test y dispersar los ítems delicados ubicándolos después de la mitad del cuestionario o hacia el final.

La *contaminación* ocurre cuando las preguntas iniciales van llevando, en un "efecto embudo", a una respuesta bastante previsible que altera en un sentido los resultados de la investigación. Una regla útil, para evitar la contaminación de la respuesta, es plantear al comienzo las preguntas generales y después ubicar las específicas.

2.2.2. LA REDACCIÓN DE LOS ÍTEMS

Cuando se redactan las preguntas, debe evitarse:

- La utilización de frases con negación
- Las preguntas bidimensionales
- La inclusión de juicios de valor
- El abuso de eufemismos
- El uso de un registro técnico muy especializado

Las preguntas con *negación* confunden al informante, además tienen mayor riesgo de que lo anotado sea contrario a la opinión que se quiso manifestar. Si, por ejemplo, previamente se hubiera redactado el ítem:

9. No está de acuerdo en que las sodas escolares no tienen que vender alimentos con exceso de azúcar y grasas saturadas.

1. ❑ De acuerdo 2. ❑ En desacuerdo 3. ❑ NS/NR

Este debe corregirse redactándolo afirmativamente:

9. Las sodas escolares solamente tienen que vender alimentos bajos en azúcar y libres de grasas saturadas.

1. ❑ De acuerdo 2. ❑ En desacuerdo 3. ❑ NS/NR

EJERCICIO 2.4

Corrija el siguiente cuestionario:

Marque una X en la casilla correspondiente si usted está de acuerdo o en desacuerdo con las siguientes medidas:

		Acuerdo	Desacuerdo
a)	No debería venderse cigarrillos a menores de edad		
b)	Nunca debe venderse cigarrillos sueltos		
c)	Jamás hay que abandonar la campaña antitabaco		
d)	El Estado no debe contratar empleados que padecen de tabaquismo		
e)	No tiene que reducirse la información masiva sobre las enfermedades causadas por el fumado		

Las redacciones *bidimensionales* solicitan una sola respuesta en un ítem que, en realidad, contiene dos preguntas, por ejemplo:

3. ¿Está a favor o en contra de que el Ministerio de Salud Pública intervenga en la crisis financiera de la Caja Costarricense del Seguro Social y se aumente el porcentaje de la cuota obrero-patronal?

   1. ❑ A favor 2. ❑ En contra 3. ❑ NS/NR

La pregunta genera dudas por demandar dos respuestas diferentes en una sola formulación. Algunas personas podrían estar a favor de las dos, otras en contra de las dos, un grupo podría favorecer la intervención pero no el aumento de las cuotas y otro grupo manifestarse en contra de la intervención a la vez que apoya el aumento en el porcentaje de la cuota obrero-patronal. En todo caso, los informantes no podrían dar una respuesta clara y tendrían que explicarla o condicionarla, incluso algunos estarían tentados a declinar la respuesta.

Para evitar ambigüedades, basta con transformarla en dos ítems cuya redacción sea simple y donde cada uno incluya un solo cuestionamiento.

3. ¿Está a favor o en contra de que el Ministerio de Salud Pública intervenga en la crisis financiera de la Caja Costarricense del Seguro Social?

1. ❑ A favor 2. ❑ En contra 3. ❑ NS/NR

4. ¿Está a favor o en contra de que se aumente el porcentaje de la cuota obrero-patronal?

1. ❑ A favor 2. ❑ En contra 3. ❑ NS/NR

La presencia de *juicios de valor* en una pregunta puede darse por referencias emocionales (éticas, religiosas, políticas), por la orientación de la respuesta en un sentido predeterminado o por la inclusión de figuras de autoridad. Algunos ejemplos:

7. ¿Apoyaría la legalización de la eutanasia en pacientes terminales, práctica que sería aplicada por médicos que se han educado para salvar vidas?

1. ❑ Sí 2. ❑ No 3. ❑ NS/NR

La frase "práctica que sería aplicada por médicos que se han educado para salvar vidas" introduce un referente emocional y con ello sesga la respuesta. Este juicio debe ser eliminado dejando solamente: "¿Apoyaría la legalización de la eutanasia en pacientes terminales?".

Otro ejemplo:

5. ¿Cómo valora el trabajo del ministro de Salud, Dr. Pedro Moreno Gómez, sabiendo que es nieto del Dr. Ricardo Moreno Cañas?

1. ❑ Excelente
2. ❑ Bueno
3. ❑ Regular
4. ❑ Malo
5. ❑ Deficiente

La información adicional del ascendente familiar del Ministro orienta la respuesta hacia una mejor valoración y este sesgo debe ser evitado. Basta con recortar el enunciado a: "¿Cómo valora el trabajo del ministro de Salud, Dr. Pedro Moreno Gómez?".

Las figuras de autoridad pueden ser, además de personas prestigiosas, grupos con características especiales, veamos:

3. ¿Al igual que la comunidad científica, apoya la dotación de más presupuesto para el Ministerio de Ciencia y Tecnología?

1. ❑ Sí 2. ❑ No 3. ❑ NS/NR

Referirse a "la comunidad científica" lleva al entrevistado a responder en forma diferente a si solo se preguntara: "¿Apoya la dotación de más presupuesto para el Ministerio de Ciencia y Tecnología?".

También debe reducirse el uso de *eufemismos*. Estas son palabras "menos ofensivas" que se usan en lugar de otras consideradas de mal gusto o tabú. Por ejemplo, un paramédico que pregunta ¿el niño tiene picazón atrás?, lo hace porque considera de mal gusto mencionar el ano; otro le pregunta al paciente si "¿lo ha hecho con alguien diferente de su cónyuge?" elevando a nivel de tabú el preguntar: "¿ha tenido sexo con alguien diferente de su cónyuge?".

El problema del abuso con los eufemismos es que restan precisión tanto a la pregunta como a la respuesta, pues si bien muchos informantes pueden entender, por contexto, el significado de un término como "atrás", otros podrían pensar que se trata de la nuca o de la espalda.

Situación parecida puede darse con el uso de un *registro técnico* en el lenguaje empleado, ya que pueden establecer una barrera de comunicación entre el enumerador y el entrevistado o llevar a que este responda sin saber exactamente qué se le está preguntando.

Un ejemplo de registro muy técnico:

5. ¿Cómo valora la práctica de la salpingectomía a pacientes que ya tienen dos hijos o más?
   1. Excelente
   2. Buena
   3. Regular
   4. Inconveniente

La mayoría de los entrevistados preguntará ¿qué es salpingectomía?, y podría ocurrir que el entrevistador tampoco lo sepa. Algunos responderán sin saber qué contestan y otros declinarán responder. Una solución práctica es "traducir" a un registro de lenguaje más accesible para el gran público y preguntar: "¿Cómo valora la práctica de 'cortar los tubos' a pacientes mujeres que ya tienen dos hijos o más?".

Tanto en el caso de los eufemismos como en el de los tecnicismos, no puede establecerse una regla absoluta sobre incluirlos o no. Su empleo depende del público meta, esto es, de las personas a las que se aplicará el cuestionario. En ejemplos previos se utilizó el eufemismo "quitarse de en medio" en lugar de suicidarse, y un término técnico: "eutanasia" en lugar de muerte asistida. En ambos casos su utilización dependerá del contexto social de la entrevista y del nivel de información o educación del encuestado.

EJERCICIO 2.5

Con la ayuda de un diccionario, cambie el excesivo registro técnico del siguiente cuestionario por un lenguaje más accesible para el informante promedio:

Un otorrinolaringólogo entrevista a una paciente:

		Sí	No
a)	¿Tiene disnea?		
b)	¿Ha padecido broncoespasmos por la noche?		
c)	¿Sufre de odinofagia por la mañana?		
d)	¿Usted esputa con frecuencia?		
e)	¿Le han salido aftas bucales?		

EJERCICIO 2.6

Cambie los eufemismos utilizados en el siguiente cuestionario por términos más precisos:

		Sí	No
a)	¿Cree que la mayoría de sus alumnos están "rellenitos"?		
b)	Es decir, ¿piensa que están "muy alentaditos"?		
c)	¿Les ha afectado "el incremento negativo" en el presupuesto del comedor escolar?		
d)	¿Ya les aplicaron "la tasa temporal de solidaridad" presupuestaria?		
e)	¿El Director "se hace de la vista gorda" frente a las limitaciones y carencias?		

2.2.3. OBTENCIÓN Y PROCESAMIENTO DE LOS DATOS

El material recolectado pasará a ser convertido en cifras numéricas facilitadoras del análisis estadístico, por eso las boletas entran en un proceso de revisión y crítica para luego ser codificadas y tabuladas.

Se recomienda que la *revisión* sea realizada por el supervisor del trabajo de campo en el momento en que se recolectan las boletas aplicadas. Este procederá en forma inmediata a:

a) Verificar el resultado de la entrevista

b) Corroborar el control de visitas

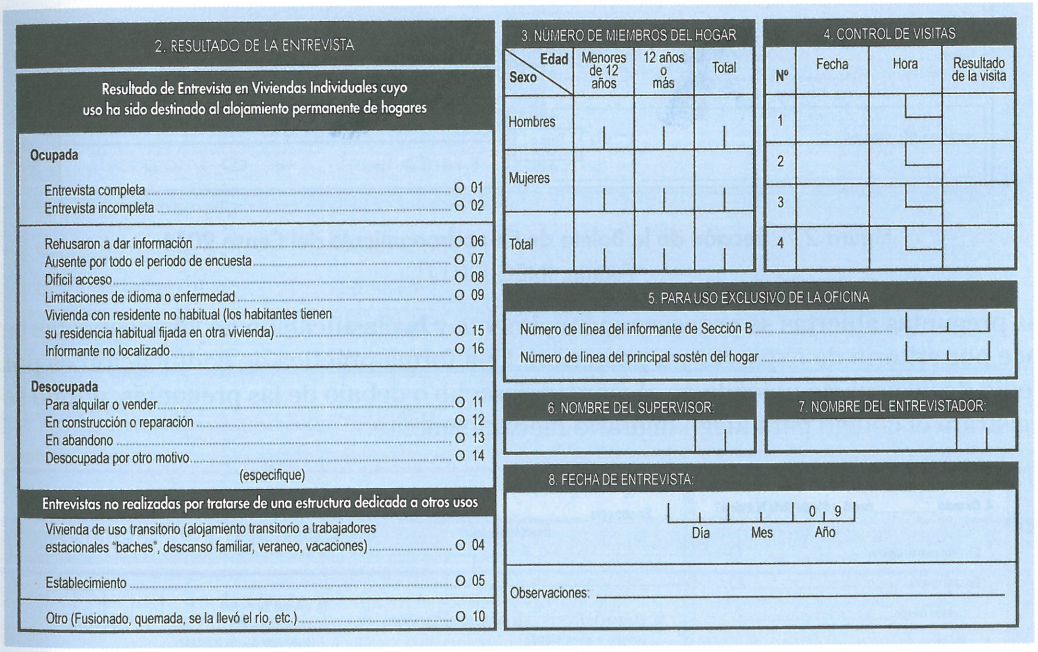
c) Cotejar que todas las preguntas hayan sido contestadas

d) Subsanar respuestas ilegibles o manchadas

e) Inspeccionar posibles incongruencias y casos dudosos

f) Reenviar al encuestador para que rectifique los faltantes

g) Instruir al personal de campo para evitar reincidir en los errores



2. RESULTADO DE LA ENTREVISTA

Resultado de Entrevista en Viviendas Individuales cuyo uso ha sido destinado al alojamiento permanente de hogares

Ocupada

- Entrevista completa ..... O 01
- Entrevista incompleta ..... O 02
- Rehusaron a dar información ..... O 06
- Ausente por todo el periodo de encuesta ..... O 07
- Difícil acceso ..... O 08
- Limitaciones de idioma o enfermedad ..... O 09
- Vivienda con residente no habitual (los habitantes tienen su residencia habitual fijada en otra vivienda) ..... O 15
- Informante no localizado ..... O 16

Desocupada

- Para alquilar o vender ..... O 11
- En construcción o reparación ..... O 12
- En abandono ..... O 13
- Desocupada por otro motivo ..... O 14

(especifique)

Entrevistas no realizadas por tratarse de una estructura dedicada a otros usos

- Vivienda de uso transitorio (alojamiento transitorio a trabajadores estacionales "baches", descanso familiar, veraneo, vacaciones) ..... O 04
- Establecimiento ..... O 05
- Otro (Fusionado, quemada, se la llevó el río, etc.) ..... O 10

3. NÚMERO DE MIEMBROS DEL HOGAR

Sexo \ Edad	Menores de 12 años	12 años o más	Total
Hombres			
Mujeres			
Total			

4. CONTROL DE VISITAS

Nº	Fecha	Hora	Resultado de la visita
1			
2			
3			
4			

5. PARA USO EXCLUSIVO DE LA OFICINA

Número de línea del informante de Sección B .....

Número de línea del principal sostén del hogar .....

6. NOMBRE DEL SUPERVISOR:

7. NOMBRE DEL ENTREVISTADOR:

8. FECHA DE ENTREVISTA:

Día Mes 0 9 Año

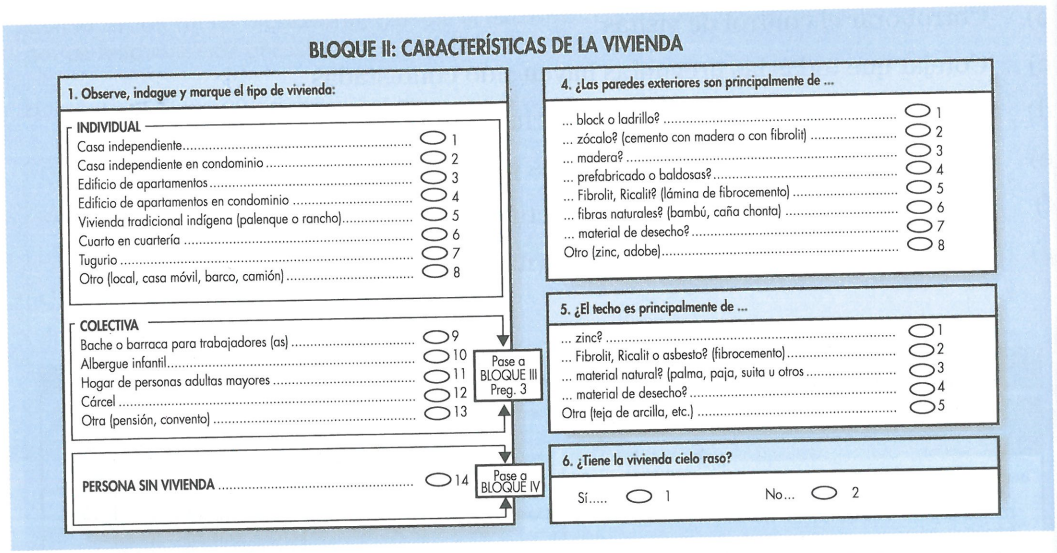
Observaciones: _____

Figura 2.6. **Boleta de recolección "Encuesta de Hogares para Propósitos Múltiples"**

Fuente: INEC (2009)

La *crítica* es una nueva revisión hecha por los especialistas que evalúan a los supervisores, analizan la coherencia de lo recolectado, los aspectos a corregir y, eventualmente, deciden si hay que volver al campo. Es de utilidad, tabular una muestra del trabajo de cada encuestador buscando detectar datos anormales, posibles fraudes o fallas en la supervisión.

La *codificación* es más sencilla en las preguntas cerradas o en las semiabiertas porque previamente se conocen las categorías. Los datos codificados facilitan el conteo de ocurrencias; cuando las encuestas son pocas, pueden usarse sistemas de conteo manual (palotes) aunque ya son muy accesibles programas computacionales que extraen la información con lectores ópticos.


BLOQUE II: CARACTERÍSTICAS DE LA VIVIENDA

1. Observe, indague y marque el tipo de vivienda:

INDIVIDUAL
Casa independiente ... 1
Casa independiente en condominio ... 2
Edificio de apartamentos ... 3
Edificio de apartamentos en condominio ... 4
Vivienda tradicional indígena (palenque o rancho) ... 5
Cuarto en cuartería ... 6
Tugurio ... 7
Otro (local, casa móvil, barco, camión) ... 8

COLECTIVA
Bache o barraca para trabajadores (as) ... 9
Albergue infantil ... 10
Hogar de personas adultas mayores ... 11
Cárcel ... 12
Otra (pensión, convento) ... 13
Pase a BLOQUE III Preg. 3

PERSONA SIN VIVIENDA ... 14
Pase a BLOQUE IV

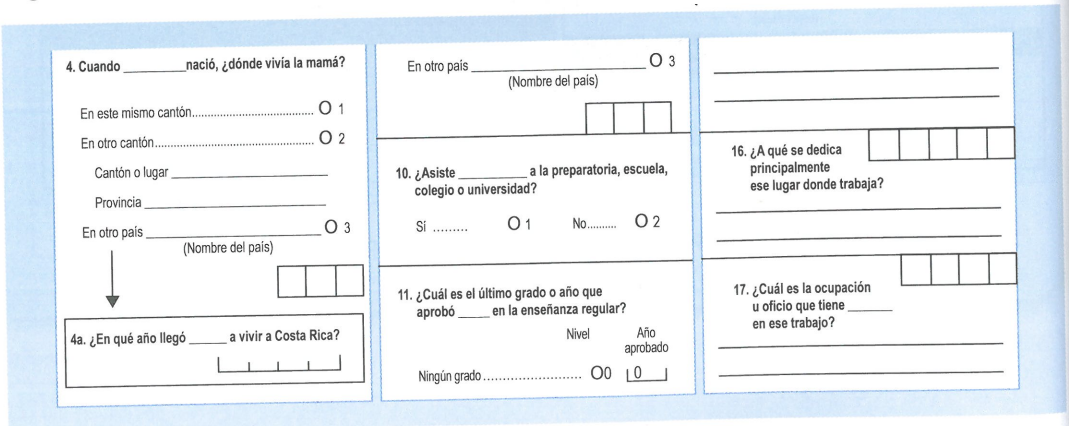
4. ¿Las paredes exteriores son principalmente de ...
... block o ladrillo? ... 1
... zócalo? (cemento con madera o con fibrolit) ... 2
... madera? ... 3
... prefabricado o baldosas? ... 4
... Fibrolit, Ricalit? (lámina de fibrocemento) ... 5
... fibras naturales? (bambú, caña chonta) ... 6
... material de desecho? ... 7
Otro (zinc, adobe) ... 8

5. ¿El techo es principalmente de ...
... zinc? ... 1
... Fibrolit, Ricalit o asbesto? (fibrocemento) ... 2
... material natural? (palma, paja, suita u otros ... 3
... material de desecho? ... 4
Otra (teja de arcilla, etc.) ... 5

6. ¿Tiene la vivienda cielo raso?
Sí..... 1 No... 2

Figura 2.7. **Sección de la Boleta de Empadronamiento del Censo 2011**
Fuente: INEC (2011)

Las preguntas abiertas se procesan manualmente y la clasificación de las respuestas se hace por criterio de expserto y a posteriori. Una forma práctica de codificación es por medio de un espacio ubicado en el margen derecho o debajo de las preguntas, donde se registrará el código para luego digitarlo mecánicamente.


4. Cuando _____ nació, ¿dónde vivía la mamá?
En este mismo cantón ... O 1
En otro cantón ... O 2
Cantón o lugar _____
Provincia _____
En otro país _____ O 3
(Nombre del país)

4a. ¿En qué año llegó _____ a vivir a Costa Rica?

En otro país _____ O 3
(Nombre del país)

10. ¿Asiste _____ a la preparatoria, escuela, colegio o universidad?
Sí ..... O 1 No..... O 2

11. ¿Cuál es el último grado o año que aprobó _____ en la enseñanza regular?
Nivel Año aprobado
Ningún grado ... O0 0

16. ¿A qué se dedica principalmente ese lugar donde trabaja?

17. ¿Cuál es la ocupación u oficio que tiene _____ en ese trabajo?

Figura 2.8. **Sección de la Boleta de Empadronamiento del Censo 2000**
Fuente: INEC (2000)

Existen algunos sistemas de codificación de uso generalizado como los códigos geográficos de Costa Rica (provincias del 1 al 7, cantones del 01 en adelante y distritos del 01 en adelante) que puede consultarse en <http://www.inec.go.cr/A/MS/Encuestas/>. También, los códigos de enfermedades están estandarizados en la Clasificación Internacional de Enfermedades (CIE) de la Organización Mundial de la Salud, que puede revisar en la dirección electrónica:

<http://www.coinsasalud.com.ar/~coinsasa/media/CIE_10_COMPLETO.xls>.

El trabajo de *tabulación* implica ordenar los datos en clases (categorías con X rasgo) que, idealmente, sean mutuamente excluyentes y colectivamente exhaustivas.

Son clases *mutuamente excluyentes* si un dato que pertenece a una de ellas, no puede pertenecer a otra al mismo tiempo, esto es, que un resultado solamente puede ser ubicado bajo una categoría, sin dudas ni duplicaciones.

Las clases son *colectivamente exhaustivas* cuando cualquier resultado posible, siempre encontrará una categoría en la cual registrarse, es decir, que todas las clases juntas representan el 100 % y no hay omisiones.

EJERCICIO 2.7

Señale si los siguientes pares de categorías son mutuamente excluyentes.

a)	Soltero	Casado	
b)	Trabajar	Tener caries	
c)	Contaminado	No contaminado	
d)	Tener tos	Tener diarrea	
e)	Enfermo	Sano	
f)	Enfermo	Hospitalizado	
g)	Fuma	No fuma	
h)	Nacional	Extranjero	

EJERCICIO 2.8

Anote si las siguientes clases son colectivamente exhaustivas.

a)	De Heredia o Cartago	
b)	Soltero, casado, unión libre, separado, divorciado o viudo	
c)	Contaminado o no contaminado	
d)	Dengue o paludismo	
e)	Estar incapacitado o no	
f)	Ganar, perder o empatar un partido de fútbol	
g)	Obtener 7.0, 8.0 o 9.0 en el final de Estadística	

La tabulación permite obtener *distribuciones de frecuencias*, estas informan del número de ocurrencias obtenidas para las diferentes clases y pueden ser frecuencias absolutas o relativas.

Las *frecuencias absolutas* contienen el número simple de elementos encontrados de cada clase. Siempre que las clases sean mutuamente excluyentes y colectivamente exhaustivas, la sumatoria de todas las frecuencias absolutas es igual al número de observaciones realizadas.

Cuadro 2.1

Datos de egresos hospitalarios por parto y aborto en Costa Rica durante el año 2003, según estado civil de la parturienta

Estado civil	Frecuencia absoluta
Casada	33 192
Soltera	16 953
Unión libre	26 324
Divorciada	821
Viuda	156
Desconocido	782
TOTAL	78 228

Fuente: CCSS (2004)
Resumido con propósitos didácticos

Las *frecuencias relativas* se obtienen dividiendo cada frecuencia absoluta entre el total de las observaciones e informan el porcentaje que cada clase representa del total, por ejemplo:

$$Frecuencia\ relativa = \frac{frecuencia\ absoluta}{TOTAL} = 0,2167$$

$$Frecuencia\ relativa_{Soltera} = \frac{16\ 953}{70\ 228} = 0,2167$$

En el año 2003 el 21,67 % de los egresos hospitalarios por parto y aborto correspondió a mujeres solteras.

Cuadro 2.2

Frecuencias de egresos hospitalarios por parto y aborto en Costa Rica durante el año 2003, según estado civil de la parturienta

Estado civil	Frecuencia absoluta	Frecuencia Relativa
Casada	33192	0,4243
Soltera	16 953	0,2167
Unión libre	26 324	0,3365
Divorciada	821	0,0105
Viuda	156	0,0020
Desconocido	782	0,0100
	78 228	1,0000

Fuente: CCSS (2004)
Resumido para propósitos didácticos

2.2.4. MÉTODOS DE TABULACIÓN

Cuando los datos son pocos, la tabulación puede hacerse en forma manual utilizando palotes. Pero cuando la cantidad de datos es grande, es necesario echar mano de programas de cómputo que permiten ordenar y tabular los datos en forma práctica.

Los *palotes* son instrumentos sencillos de conteo que solo requieren de lápiz o lapicero y una hoja de papel en la cual se hace una matriz para anotar las clases. A la par de la respectiva clase, se traza un palote por cada observación registrada. Con el fin de facilitar el conteo, se recomienda hacer bloques cada cinco observaciones (///// /////).

Basados en el Cuadro 2.3, tabule las frecuencias del tipo de seguro de las parturientas.

Cuadro 2.3

Informe de atención de partos, Clínica de Guabito
Mayo de 2015

ESTADO CIVIL	CLASE DE PARTO	TIPO DE SEGURO	SEXO DEL RECIÉN NACIDO
1 Casada	Vaginal	Indirecto	Femenino
2 Soltera	Cesaria	Directo	Masculino
3 Unión libre	Episiotomía	Directo	Femenino
4 Divorciada	Vaginal	Directo	Masculino
5 Viuda	Vaginal	Directo	Femenino
1 Casada	Cesaria	Indirecto	Femenino
3 Unión libre	Episiotomía	Directo	Masculino
2 Soltera	Cesaria	Directo	Masculino
3 Unión libre	Vaginal	Directo	Femenino
3 Unión libre	Cesaria	Directo	Masculino
1 Casada	Vaginal	Indirecto	Femenino
4 Divorciada	Cesaria	Sin seguro	Masculino
2 Soltera	Vaginal	Indirecto	Femenino
1 Casada	Vaginal	Directo	Masculino
1 Casada	Vaginal	Indirecto	Masculino

Fuente: Datos hipotéticos

Para ello, se listan las clases en la primera columna, luego se traza un palote a la par del tipo de seguro al que está adscrita cada parturienta. Finalmente se cuentan los palotes y se presentan los resultados en forma de cuadro, así:

Cuadro 2.4

Tipo de seguro de las parturientas atendidas en la Clínica de Guabito durante el mes de mayo de 2015

TIPO DE SEGURO	PALOTES	FRECUENCIA ABSOLUTA
Directo	///// ////	9
Indirecto	/////	5
Sin seguro	/	1
TOTAL		15

EJERCICIO 2.9

Utilizando la técnica de palotes y los datos del Cuadro 2.3, elabore:

a) Un informe de la distribución de frecuencias para la variable "Estado civil".

ESTADO CIVIL	PALOTES	FRECUENCIA ABSOLUTA
Casada		
Soltera		
Unión libre		
Divorciada		
Viuda		
TOTAL		

b) Un informe de la distribución de frecuencias para la variable "Clase de Parto".

CLASE DE PARTO	PALOTES	FRECUENCIA ABSOLUTA
Cesaria		
Episiotomía		
Vaginal		
TOTAL		

c) Un informe de la distribución de frecuencias para la variable "Sexo del Recién Nacido".

SEXO DEL RECIÉN NACIDO	PALOTES	FRECUENCIA ABSOLUTA
Femenino		
Masculino		
TOTAL		

2.2.5. DISTRIBUCIONES SIMPLES Y DE DOBLE ENTRADA

Hasta el momento, se han tabulado *distribuciones simples* de frecuencia o univariadas, pues solamente se ha considerado una variable. Seguidamente se pasará a tabular *distribuciones de doble entrada*, que se llaman así porque toman en cuenta simultáneamente dos o más características.

Las distribuciones de doble entrada, cruzadas o multivariadas, clasifican una característica en las filas y otra (u otras) en las columnas. Cada cruce presenta la relación de las variables involucradas, de forma que se puedan observar posibles asociaciones, dependencias o interdependencias.

Ejemplo: Utilizando los datos del Cuadro 2.3, que presenta el informe de atención a los partos en la Clínica de Guabito durante mayo de 2015, construya un cuadro de doble entrada que en las filas presente el "Estado Civil" de las parturientas y en las columnas señale el "Tipo de seguro" que las protege.

Lo primero que se hace es dibujar la matriz de la siguiente forma:

	TIPO DE SEGURO						
ESTADO CIVIL	DIRECTO		INDIRECTO		SIN SEGURO		TOTAL
	Palotes	Frecuencia	Palotes	Frecuencia	Palotes	Frecuencia	
Casada							
Soltera							
Unión libre							
Divorciada							
Viuda							
TOTAL							

Luego se hace el conteo, colocando un palote por cada parturienta en la intersección de su Estado Civil y Tipo de Seguro.

Nótese que al final de cada fila y cada columna se ha colocado una celda para registrar el total de cada variable. Esto tiene, además de un propósito informativo, la función de control puesto que el gran total tanto por filas como por columnas debe ser igual al número de registros (15 en este ejemplo).

Al finalizar el conteo se informa:

Cuadro 2.5

Condición de las parturientas atendidas en la Clínica de Guabito ordenadas por su estado civil y tipo de seguro, mayo de 2015

	TIPO DE SEGURO						
ESTADO CIVIL	DIRECTO		INDIRECTO		SIN SEGURO		TOTAL
	Palotes	Frecuencia	Palotes	Frecuencia	Palotes	Frecuencia	
Casada	/	1	////	4			5
Soltera	//	2	/	1			3
Unión libre	////	4					4
Divorciada	/	1			/	1	2
Viuda	/	1					1
TOTAL		9		5		1	**15**

Si bien este cuadro está basado en cifras hipotéticas, puede ensayarse una *interpretación* de los datos con frecuencias mayores o los casos de riesgo social, como es el de la Divorciada que no cuenta con seguro. Las frecuencias más altas indican la tendencia de las mujeres en Unión Libre a tener Seguro Directo y de las Casadas a asegurarse en forma Indirecta.

EJERCICIO 2.10

Utilizando la técnica de palotes y los datos del Cuadro 2.3, elabore:

a) Un cuadro de doble entrada que relacione el estado civil de la parturienta con la clase de parto.

	CLASE DE PARTO						
ESTADO CIVIL	CESARIA		EPISIOTOMÍA		VAGINAL		TOTAL
	Palotes	F	Palotes	F	Palotes	F	
Casada							
Soltera							
Unión libre							
Divorciada							
Viuda							
TOTAL							

b) Un cuadro de doble entrada que relacione tipo de parto y sexo del recién nacido.

	SEXO DEL RECIÉN NACIDO				
TIPO DE PARTO	FEMENINO		MASCULINO		TOTAL
	Palotes	F	Palotes	F	
Cesaria					
Episiotomía					
Vaginal					
TOTAL					

También pueden hacerse cuadros con triple clasificación como el siguiente:

Cuadro 2.6

Informe de situación de las parturientas atendidas en la Clínica de Guabito, mayo de 2015 ordenados por estado civil de la madre, sexo del recién nacido y clase de parto.

ESTADO CIVIL	TOTAL GENERAL	FEMENINO			TOTAL FEMENINO	MASCULINO			TOTAL MASCULINO
		Cesaria	Episiotomía	Vaginal		Cesaria	Episiotomía	Vaginal	
Casada	5	1		2	3			2	2
Soltera	3			1	1	2			2
Unión libre	4		1	1	2	1	1		2
Divorciada	2				0	1		1	2
Viuda	1			1	1				0
TOTAL	**15**	1	1	5	7	4	1	3	8

Además, se pueden realizar cuadros generales mucho más complejos, pero su confección sobrepasa los objetivos planteados en este curso.

Una vez que usted domina el conteo manual con palotes, lea el ANEXO 1 que explica la forma de tabulación utilizando Excel de Microsoft.

2.3. PRESENTACIÓN DE LOS DATOS ESTADÍSTICOS

Las formas usuales de presentación de la información estadística son:

- Textual
- Semitabular
- Cuadros
- Gráficos

Se debe asegurar que todas ellas expliquen cuáles fueron las unidades observadas, cómo se clasificaron, en qué periodo de tiempo y en cuál lugar se recolectaron los datos.

2.3.1. FORMA TEXTUAL

Es la forma más simple de presentar y de entender los hallazgos de una investigación o un registro, especialmente cuando no está muy cargado de datos numéricos. En la presentación del informe, las cifras importantes se van mencionando y comentando en una forma fluida, subrayando y explicando los aspectos relevantes, por ejemplo:

> A pesar de que cada año hay un aumento de cerca del 10 % en la cantidad de vehículos que circulan en nuestras calles, la tasa de accidentes de tránsito se ha reducido en el transcurso de estos 10 años, pasando de 97,5 por cada mil vehículos en el año 2001 a 41,2 en el 2010.
>
> También las muertes por accidentes de tránsito se han reducido, pasando de 10,05 muertes por cada 10 000 vehículos en el 2001 a 3,93 muertes en el 2010 (RITEVE, 2012).

Debe resaltarse, como desventajas, que se hace necesario leer todo el texto para captar el mensaje, que no pueden hacerse comparaciones y que, si el volumen de datos es grande, el informe se vuelve confuso y cansado de leer.

2.3.2. FORMA SEMITABULAR

Se utiliza cuando hay un volumen de datos numéricos un poco mayor que vuelven poco práctico mencionarlos en forma textual. Se requiere que los datos estén organizados de manera simple, con pocas clases y pocas columnas.

La forma general de presentarlos es mediante un texto introductorio, que señala el lugar y el tiempo en que se recolectaron los datos, luego se mencionan los hallazgos principales seguidos de un cuadro sencillo. Posteriormente se hace énfasis en algunos elementos. Por ejemplo:

A la semana epidemiológica 20 (del 13 al 19 de mayo), se han registrado en el país un total de 2 237 casos de dengue, lo que representa un incremento del 46,2 % respecto al mismo período del año anterior. La Región Huetar Atlántica continúa concentrando la mayor carga del problema (60,0 %)

REGIÓN	2012	%
Pacífico Central	251	11,2
Chorotega	259	11,6
Central Norte	165	7,4
Huertar Atlántica	1 343	60,0
Central Sur	88	3,9
Brunca	63	2,8
Central Este	33	1,5
Occidente	2	0,1
Huetar Norte	33	1,5
TOTAL	2 237	100,0

A la semana 20 terminada el 19 de mayo 2012

Fuente: Dirección de Vigilancia de la Salud

La Región Huetar Atlántica se mantiene en zona de alerta del canal endémico observándose una tendencia en las semanas anteriores, situación que se refleja, principalmente en el área de mayor incidencia de Pococí. (Ministerio de Salud, 2012, p. 23).

2.3.3. CUADROS ESTADÍSTICOS

Presentan los datos cuantitativos ordenados en filas y columnas, de manera que estos puedan ser comprendidos sin que sea necesaria mucha explicación textual o gráfica. Todo cuadro contiene necesariamente:

- Título
- Columna matriz
- Encabezados
- Cuerpo
- Fuente

Adicionalmente puede incluir:

- Número de cuadro
- Nota introductoria
- Notas al pie

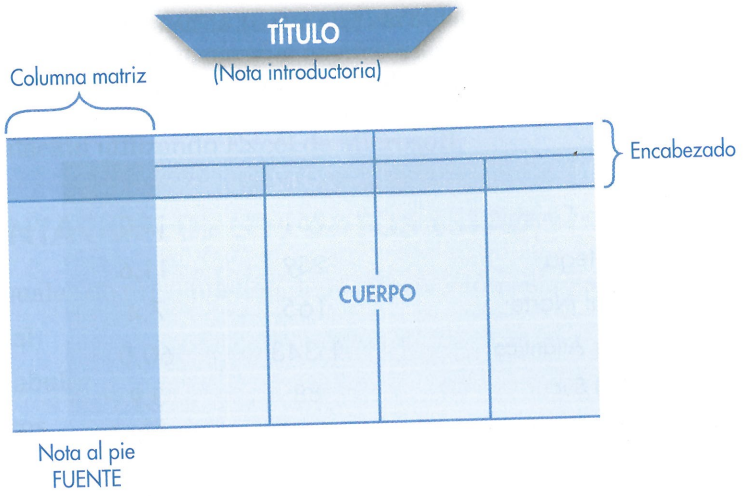


Figura 2.9. **Componentes de un cuadro estadístico**

Fuente: Reproducido de Gómez (1998, p. 148)

El *título* presenta brevemente, pero con claridad, el tipo de datos que contiene el cuadro, respondiendo "a las preguntas: *qué* son los datos, *dónde* se recogieron, *cuándo y cómo* y bajo *qué* criterio se clasificaron" (Gómez, p. 149). Si el título fuera de dos o tres líneas, se recomienda presentarlo en forma de "trapecio" con el primer renglón más largo y el resto con extensión cada vez menor.

La *columna matriz* es, generalmente, la primera de la izquierda y contiene la clasificación principal de las clases.

Los *encabezados* van el la primera fila e indican una o varias categorías en que se van a presentar los datos.

El *cuerpo* es el contenido fundamental de los hallazgos e informa los datos según la fila y la columna respectiva.

La *fuente* es obligatoria cuando los datos han sido obtenidos por otros investigadores. Es parte de la honestidad intelectual, además permite guiar a quienes quieren corroborar la información o ampliar en el conocimiento de las cifras.

El *número de cuadro* facilita las referencias y la presentación textual de análisis o comentarios, esto solamente es necesario cuando un informe contiene varios cuadros.

Las *notas introductorias* se ubican después del título y se estilan en una letra ligeramente más pequeña que la de este. Ellas anuncian las unidades en que va el informe, sirven para aclarar elementos del título, advertir sobre alguna definición o establecer una fecha a partir de la cual se hacen las comparaciones.

Las *notas al pie* explican o advierten sobre un dato específico o una categoría en particular. Se ubican en la parte inferior del cuadro, después del cuerpo y antes de la fuente.

2.3.3.1. Series estadísticas

Ya se ha visto que para transformar los datos en *información* estadística, deben ordenarse siguiendo algún criterio o conforme alguna característica que los presente como series:

- Cualitativas
- Geográficas
- Cronológicas
- Cuantitativas

En las *series cualitativas* la información va ordenada de acuerdo con un atributo, cualidad o característica. Estas series pueden presentarse en forma usual (como el orden del sexo o del estado civil) o alfabéticamente, con lo que se facilita la búsqueda de la información, por ejemplo:

Cuadro 2.7

Apertura de historias clínicas según servicios en el año 2011
Instituto Regional de Enfermedades Neoplásicas

APERTURA DE HISTORIAS CLÍNICAS POR SERVICIOS MÉDICOS				
Servicios	Masculino	Femenino	TOTAL	%
TOTAL	490	1 005	1 495	100
Abdomen	167	211	378	25,3
Cabeza y cuello	114	194	308	20,6
Ginecología	0	339	339	22,7
Medicina oncológica	26	20	46	3,1
Senos y tumores mixtos	45	217	262	17,5
Urología	138	24	162	10,8

Fuente: IREN (2011)

En las *series geográficas* las categorías se pueden ordenar alfabéticamente, por orden de magnitud, de menor a mayor o viceversa, o en forma usual (como el orden de las provincias de Costa Rica), por ejemplo:

Cuadro 2.8

Indicadores de recursos, acceso y cobertura por provincia Costa Rica, 2005

	Cobertura de inmunizaciones DPT3 niños menores de un año %	Cobertura de inmunizaciones VOP 3 niños menores de un año %	Cobertura de inmunizaciones BCG niños menores de un año %	Cobertura de inmunizaciones Sarampión niños menores de un año %	EBAIS funcionando
Costa Rica	89,30	89,18	81,50	87,87	903
San José	84,59	84,15	67,19	82,66	254
Alajuela	90,38	90,45	93,32	88,08	173
Cartago	87,86	87,19	82,88	89,53	102
Heredia	86,88	87,02	85,69	86,00	77
Guanacaste	93,88	93,88	98,70	98,87	84
Puntarenas	91,44	91,63	82,49	87,54	108
Limón	99,70	100,00	87,45	98,96	105

Fuente: Organización Panamericana de la Salud (2006)

En las *series cronológicas* el criterio de clasificación es el tiempo. Las clases se ordenan respetando la ocurrencia de los eventos: de lunes a domingo o de domingo a sábado, de enero a diciembre, conforme las etapas de un proceso o por volúmenes anuales. Por ejemplo:

Cuadro 2.9

Población de Costa Rica estimada al 30 de junio de cada año, del 2001 al 2010

Año	Población al 30 de junio		
	Total	Hombres	Mujeres
2001	4 005 538	2 035 118	1 970 420
2002	4 071 879	2 068427	2 003 452
2003	4 136 250	2 100 839	2 035 411
2004	4 200 278	2 133 102	2 067 176
2005	4 263 479	2 164 807	2 098 672
2006	4 326 071	2 196 093	2 129 978
2007	4 389 139	2 227 538	2 161 601
2008	4 451 205	2 258 500	2 192 705
2009	4 509 290	2 287 405	2 221 885
2010	4 563 539	2 314 293	2 249 246

Fuente: *INEC (2011)*

Las *series cuantitativas* tienen su origen en variables que pueden ser discretas o continuas. Las variables pueden ser ordenadas por magnitud o por importancia, y siempre es prudente indicar la escala con que están hechas las mediciones: peso, distancia, temperatura, tiempo, unidades, cientos, miles, etc.

Un ejemplo de serie cuantitativa de variable discreta lo construiremos a partir de los siguientes datos hipotéticos sobre el número de hijos de 25 estudiantes de posgrado:

ESTUDIANTE	NÚMERO DE HIJOS
AB	4
AF	3
AG	2
AJ	2
BG	1
CB	1
EA	3
EC	2
EF	3
ER	4
FR	2
GG	3
JG	2

ESTUDIANTE	NÚMERO DE HIJOS
JJ	0
JR	2
JV	3
KV	5
PP	4
RG	3
RL	5
RS	6
SG	3
SL	3
TQ	2
VG	4

Con esos datos se construye la siguiente distribución de frecuencias:

Cuadro 2.10

Número de hijos de los estudiantes del Posgrado en Registros Médicos UNED, Segundo Ciclo 2015

CLASES N.° de hijos	Palotes	FRECUENCIAS ABSOLUTAS	FRECUENCIAS RELATIVAS
0	/	1	0,04
1	//	2	0,08
2	///// //	7	0,28
3	///// ///	8	0,32
4	////	4	0,16
5	//	2	0,08
6	/	1	0,04
		25	1

Fuente: Datos hipotéticos

Tanto las frecuencias absolutas como las relativas pueden ser acumuladas en sentido "X o menos" sumándolas de arriba hacia abajo, o en sentido "X o más" sumándolas de abajo hacia arriba.

Cuadro 2.11

Frecuencias absolutas y relativas del número de hijos de los estudiantes del Posgrado en Registros Médicos

CLASES	FRECUENCIA ABSOLUTA			FRECUENCIA RELATIVA		
N.° de hijos	Simple	Acumulada "X o menos"	Acumulada "X o más"	Simple	Acumulada "X o menos"	Acumulada "X o más"
0	1	1	25	0,04	0,04	1,00
1	2	3	24	0,08	0,12	0,96
2	7	**10**	22	0,28	**0,40**	0,88
3	8	18	**15**	0,32	0,72	**0,60**
4	4	22	7	0,16	0,88	0,28
5	2	24	3	0,08	0,96	0,12
6	1	25	1	0,04	1,00	0,04
	25			1,00		

Fuente: Datos hipotéticos

La frecuencia acumulada "X o menos" se llama así porque surge de la formulación de la pregunta: ¿cuántos estudiantes tienen "2 hijos o menos"? La respuesta es 10 estudiantes puesto que en la frecuencia se han sumado de arriba hacia abajo los que no tienen hijos más los que tienen 1 y los que tienen 2 hijos (1+2+7=10).

La frecuencia acumulada "X o más" recibe esta denominación porque se formula la pregunta: ¿cuántos estudiantes tienen "3 hijos o más"?, cuya respuesta se busca directamente en el cuadro y es 15 puesto que ha sumado de abajo hacia arriba 1+2+4+8=15

En el caso de las frecuencias relativas acumuladas, la pregunta que se plantea es: ¿qué porcentaje de estudiantes tiene "2 hijos o menos"? y la respuesta es 0,40, que amplificado es igual a 40 %. Desde otro ángulo se pregunta: ¿qué porcentaje de estudiantes tiene "3 hijos o más"?, cuya respuesta es 60 % (la suma de 0,04+0,08+0,16+0,32=0,6).

EJERCICIO 2.11

Complete el cuadro siguiente que informa del número de terapias trimestrales a las que acudió un grupo de 80 pacientes diagnosticados con artrosis.

CLASES	FRECUENCIA ABSOLUTA			FRECUENCIA RELATIVA		
N.° de terapias en el trimestre	Simple	Acumulada "X o menos"	Acumulada "X o más"	Simple	Acumulada "X o menos"	Acumulada "X o más"
0	8					
1	2					
2	8					
3	16					
4	24					
5	8					
6	8					
7	4					
8 o más	2					
	80			1 000		

Una vez confeccionado el cuadro, conteste:

a) ¿Cuántos pacientes participaron en 4 sesiones de terapia?

b) ¿Cuántos pacientes participaron en "4 terapias o menos"?

c) ¿Cuántos pacientes participaron en "4 terapias o más"?

d) ¿Qué porcentaje de pacientes participó en 3 sesiones de terapia?

e) ¿Qué porcentaje de pacientes participó en "3 terapias o menos"?

f) ¿Qué porcentaje de pacientes participó en "3 terapias o más"?

2.3.3.2. Distribuciones de frecuencia de variables continuas

Cuando se tabula una serie cuantitativa de variable continua, primero debe definirse el número de clases en que se dividirá la distribución; luego, debe establecerse el ancho de las clases, los criterios para determinar los límites de clase, el punto medio y la forma de tratar las clases abiertas.

Para ilustrar la técnica, se usarán los siguientes datos hipotéticos del peso en gramos de 40 recién nacidos, los cuales ya han sido ordenados de menor a mayor:

900	2 200	3 400	3 800	4 200
1 100	2 600	3 400	3 800	4 300
1 300	2 600	3 400	3 800	4 400
1 600	2 800	3 400	3 900	4 400
1 900	2 800	3 400	3 900	4 400
2 100	2 900	3 700	3 900	4 600
2 100	3 200	3 700	3 900	4 600
2 200	3 300	3 800	4 100	4 700

Para establecer el número de clases y el ancho de ellas, debe identificarse el peso más alto y el más bajo de la distribución, esto con el fin de definir dónde podría comenzar la primera clase. En este ejemplo, el peso mayor es 4 700 gramos y el peso menor es 900 gramos por lo que el *rango* o *recorrido* es 3 800 gramos.

La definición del ancho de clase está íntimamente ligada con la determinación del número de clases. Esta decisión dependerá de la cantidad de observaciones disponibles y del rango o recorrido de la distribución.

Siempre se debe tener presente que el objetivo principal que lleva a construir una distribución de frecuencias es el de convertir los datos en *información*. Si se hacen pocas categorías con amplitud de clase muy ancha, se perderán detalles relevantes; si por el contrario se hacen muchas clases, la distribución presentará formas irregulares, clases vacías o con pocas observaciones y un deficiente resumen de la información.

Por ejemplo:

Número muy pequeño de Clases (Amplitud de clase muy ancha)		Número excesivo de Clases (Poca amplitud de clase)	
CLASE	FRECUENCIA	CLASE	FRECUENCIA
Más de 500 a 2 000	5	Más de 750 a 1 000	1
Más de 2 000 a 3 500	16	Más de 1 000 a 1 250	1
Más de 3 500 a 5 000	19	Más de 1 250 a 1 500	1
	40	Más de 1 500 a 1 750	1
		Más de 1 750 a 2 000	1
		Más de 2 000 a 2 250	4
		Más de 2 250 a 2 500	0
		Más de 2 500 a 2 750	2
		Más de 2 750 a 3 000	3
		Más de 3 000 a 3 250	1
		Más de 3 250 a 3 500	6
		Más de 3 500 a 3 750	2
		Más de 3 750 a 4 000	8
		Más de 4 000 a 4 250	2
		Más de 4 250 a 4 500	4
		Más de 4 500 a 4 750	3
			40

Sobre el número de clases, recomienda Miguel Gómez (1998):

> Lo ideal es buscar un término medio, que dé un balance óptimo entre la necesidad de resumir la información por un lado y la de mantener el detalle necesario para poder apreciar las características principales de los datos, por el otro. En general, se acepta como un elemento de orientación la regla de que el número de clases no debe ser menor de 6 ni mayor de 15, aunque es obvio que, en circunstancias especiales, puede hacerse un número de clases superior o inferior a estos límites citados.
>
> En la práctica, el procedimiento usual es partir de la amplitud general y luego probar con diferente número de clases o diferentes amplitudes hasta que se llega a un número de clases y una amplitud adecuada y conveniente. Se trata de que las clases sean todas de igual amplitud y se prefiere, por comodidad, que el intervalo sea 5, 10, o un múltiplo de estos números (p. 247).

En el ejemplo del peso de los recién nacidos, el rango o recorrido es 3 800 gramos, que dividido entre un número tentativo de 7, 8 o 9 clases daría:

$$\frac{3\,800}{7} = 542{,}8 \text{ de ancho de clase}$$

$$\frac{3\,800}{8} = 475 \text{ de ancho de clase}$$

$$\frac{3\,800}{9} = 422{,}2 \text{ de ancho de clase}$$

Donde se insinúa un ancho de clase de 500 gramos. Luego se tienen que definir los límites de la primera clase. Puesto que en el caso del peso al nacer, un registro menor a 1 000 gramos se considera un peso extremadamente bajo, parece conveniente proponer las siguientes clases:

500 a 1 000 gramos
1 000 a 1 500 gramos
1 500 a 2 000 gramos
Etc.

Observe que estas clases no son *mutuamente excluyentes*, pues un bebé que pesa 1 500 gramos al nacer, generaría la duda para clasificarlo en la segunda o en la tercera clase.

Tampoco es recomendable establecer los límites de clase como sigue:

501 a 999 gramos
1 001 a 1 499 gramos
1 501 a 1 999 gramos
Etc.

Porque no son clases *colectivamente exhaustivas*, ya que el bebé que pesó 1 500 gramos al nacer, ahora no pertenece a ninguna clase: no es parte de la segunda pues pesa un gramo más y tampoco de la tercera ya que pesa un gramo menos que su límite inferior.

Existen varias soluciones para este problema, una de ellas es el establecimiento de límites indicados y límites reales. Si el instrumento de medición informa los pesos en gramos enteros, podría adicionarse 0,5 de gramo para establecer los límites reales, así:

Límites indicados	Límites reales
500 – 999	499,5 – 999,5
1 000 – 1 499	999,5 – 1 499,5
1 500 – 1 999	1 499,5 – 1 999,5
2 000 – 2 499	1 999,5 – 2 499,5

Esto se ilustra de la siguiente manera:

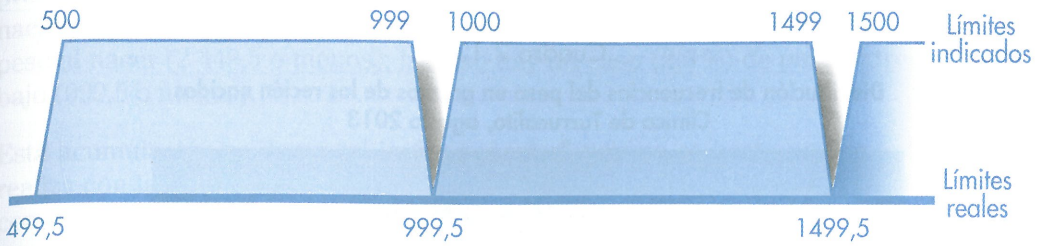


Figura 2.10. **Límites indicados y límites reales**

Fuente: Elaboración propia con fines didácticos

También pueden establecerse límites reales con los criterios:

- "A menos de" o hacia abajo
- "Más de" o hacia arriba

Hacia abajo	.....	Hacia arriba
"A menos de"	.	"Más de"
500 a menos de 1 000	.	Más de 500 a 1 000
1 000 a menos de 1 500	.	Más de 1 000 a 1 500
1 500 a menos de 2 000	.	Más de 1 500 a 2 000
2 000 a menos de 2 500	.	Más de 2 000 a 2 500

Que gráficamente se representa:



Figura 2.11. **Representación gráfica del establecimiento de límites reales Hacia abajo o "a menos de" y hacia arriba o "Más de"**

Fuente: Elaboración propia con fines didácticos

Una vez establecido un criterio para establecer los límites reales de las clases, se procede al conteo, resultando:

Cuadro 2.12

Distribución de frecuencias del peso en gramos de los recién nacidos Clínica de Turrucalito, agosto 2013

CLASES (peso en gramos)	Conteo	FRECUENCIA
499,5 – 999,5	/	1
999,5 – 1 499,5	//	2
1 499,5 – 1 999,5	//	2
1 999,5 – 2 499,5	////	4
2 499,5 – 2 999,5	/////	5
2 999,5 – 3 499,5	///// //	7
3 499,5 – 3 999,5	///// /////	10
3 999,5 – 4 499,5	///// /	6
4 499,5 – 4 999,5	///	3
		40

La distribución revela que una buena cantidad de los pesos de los recién nacidos se concentra entre 2 999,5 y 4 499,5. Debe ponerse atención a los casos con peso bajo al nacer, esto es los pesos menores de 2 500 gramos, para ello es muy útil acumular tanto las frecuencias absolutas como las relativas "al límite superior".

Cuadro 2.13

Frecuencias absolutas, relativas y acumuladas del peso en gramos de los recién nacidos, Clínica de Turrucalito, agosto 2013

CLASES	FRECUENCIAS ABSOLUTAS		FRECUENCIAS RELATIVAS	
(Peso en gramos)	Simples	Acumuladas	Simples	Acumuladas
499,5 – 999,5	1	1	0,025	**0,025**
999,5 – 1 499,5	2	3	0,050	0,075
1 499,5 – 1 999,5	2	5	0,050	0,125
1 999,5 – 2 499,5	4	9	0,100	**0,225**
2 499,5 – 2 999,5	5	14	0,125	**0,350**
2 999,5 – 3 499,5	7	21	0,175	0,525
3 499,5 – 3 999,5	10	31	0,250	0,775
3 999,5 – 4 499,5	6	37	0,150	0,925
4 499,5 – 4 999,5	3	40	0,075	1,000
	40		1,000	

Aunque estos datos son hipotéticos, una distribución parecida a esta tendría que preocupar grandemente a las autoridades de salud, puesto que el 35 % de los recién nacidos presenta peso insuficiente al nacer (2 999,5 o menos) y el 22,5 % revela bajo peso al nacer (2 449,5 o menos); incluso hay un caso (2,5 %) de peso extremadamente bajo (999,5 o menos).

Esta acumulación se llama "al límite superior" porque es hasta un límite tope que se realiza con la suma de las frecuencias de clases menores. Para este ejemplo, los niños con riesgo son los de pesos bajos, por lo que se establece un peso crítico (o límite) y se acumulan las observaciones pertenecientes a clases inferiores.

En otros casos, la acumulación recomendada es "al límite inferior" puesto que los casos problemáticos son los que sobrepasan un cierto valor crítico, por ejemplo, presiones arteriales por encima de 140/90.

Hay también casos de "doble cola", como el de la edad de la madre al momento del alumbramiento, pues preocupan tanto las parturientas de poca como de mucha edad. En estudios de este tipo, se deberán acumular las frecuencias tanto "al límite inferior" como "al límite superior".

Elementos adicionales a tomar en cuenta cuando se construyen cuadros de variables continuas son: el ancho de clase, el punto medio y el tratamiento a las clases abiertas.

El *ancho de clase* es la diferencia del límite real superior menos el límite real inferior, por ejemplo, 999,5 – 499,5 = 500. En la medida de lo posible, se recomienda que todas las clases de una distribución tengan el mismo ancho de clase.

El *punto medio* es el valor central de cada clase y se calcula como un promedio de los límites reales: $\frac{499,5+999,5}{2}=749,5$. Es el valor representativo de toda la clase y, más adelante, se empleará para calcular medidas de posición y variabilidad en datos agrupados.

Las *clases abiertas* generalmente se ubican al comienzo o al final de una distribución y solamente definen uno de los límites, por ejemplo, "999,5 gramos o menos", "75 años o más". El problema con las clases abiertas es que obligan a utilizar procedimientos arbitrarios para establecer el punto medio o el ancho de la clase; la recomendación general es evitarlas, pero no siempre es posible dada la existencia de datos extremos.

EJERCICIO 2.12

Se cuenta con la siguiente información sobre los años cumplidos de 50 mujeres al momento de dar a luz en el Hospital de la Anexión, durante diciembre de 2014.

12	24	28	31	36
13	25	29	32	36
17	25	29	32	37
17	25	30	32	37
19	25	30	32	38
19	25	30	34	39
20	27	31	34	41
23	27	31	35	42
23	28	31	35	44
24	28	31	36	47

a) Construya una distribución de frecuencias absolutas y relativas, utilizando los intervalos de clase por edad de la parturienta: 10 a menos de 15 años, 15 a menos de 20 años, etc.

CLASES	Palotes	FRECUENCIA ABSOLUTA	FRECUENCIA RELATIVA
10 a menos de15			
15 a menos de 20			
20 a menos de 25			
25 a menos de 30			
30 a menos de 35			
35 a menos de 40			
40 a menos de 45			
45 a menos de 50			
		50	1,00

b) Calcule las frecuencias acumuladas "al límite superior" y "al límite inferior", tanto las absolutas como las relativas.

	Frecuencias acumuladas al límite superior		Frecuencias acumuladas al límite inferior	
CLASES	Absoluta	Relativa	Absoluta	Relativa
10 a menos de15				
15 a menos de 20				
20 a menos de 25				
25 a menos de 30				
30 a menos de 35				
35 a menos de 40				
40 a menos de 45				
45 a menos de 50				

c) Conteste las siguientes preguntas:

i. ¿Cuántas mujeres tenían de 25 a menos de 30 años al momento de dar a luz?

ii. ¿Qué porcentaje de mujeres tenía menos de 35 años para la fecha del alumbramiento?

iii. ¿Qué porcentaje de mujeres tenía 40 años o más?

iv. ¿Cuántas mujeres tenían menos de 15 años en la fecha del parto?

v. ¿Qué porcentaje de mujeres tenía menos de 20 años?

2.3.3.3. Detalles adicionales para la confección de los cuadros

Si se desea *destacar algunas cifras*, pueden subrayarse, usar negritas, itálicas o colocarlas en un lugar privilegiado; por ejemplo, si se quiere destacar un total, este podría colocarse en la esquina superior izquierda del cuerpo del cuadro. Algunos totales pueden resaltarse con solo separarlos un renglón del cuerpo del informe, usando mayúsculas o letra de otro tipo o colocándolos inmediatamente después de los encabezados.

Si se quiere *acentuar comparaciones*, es mejor colocarlas en columnas contiguas que en filas. Cuando las cifras a comparar son relativas (porcentajes, tasas, indicadores), debe aclararse cuál fue la base para su cálculo.

Sobre la *forma* del cuadro, lo recomendado es que este no sea excesivamente alargado (muchas columnas y pocas filas o pocas columnas y muchas filas).

Respecto del *tamaño* del cuadro, se aconseja que se ubique en una sola página, si tiene que presentarse en dos o más páginas, se deberán repetir los encabezados al comienzo de cada página. Esta situación solamente se justifica en los cuadros generales y nunca en los cuadros resumen.

Cuando el *número de clases* es muy grande, es conveniente incluir un espacio adicional cada cinco clases, esto con el fin de darle mejor presencia visual y ayudar a la interpretación.

Cuando se han *redondeado* las cifras (al entero, décimo, centésimo, milésimo), debe indicarse en la nota introductoria o en nota al pie el criterio con el cual han sido redondeadas (hacia arriba, hacia abajo o a la unidad más próxima).

El *rayado* no debe ser excesivo ni "cerrar" el cuadro, esto es, que deben dejarse abiertos los extremos del cuadro. Generalmente, se separan con rayas la columna matriz, los encabezados y los totales. Hasta donde sea posible debe prescindirse del rayado en el cuerpo del cuadro, es mejor trabajar con espaciados para separar tanto las filas como las columnas.

Las *cifras faltantes* no se dejan en blanco, sino que se señalan con: tres puntos (...), ND (no disponible), NS/NR (no sabe o no responde), cero (0 o 0,0), cuando la cantidad es muy pequeña o nula; un guión (–), cuando el dato no es aplicable, o doble guión (– –) cuando las cifras no son comparables.

Los cuadros pueden ser:

- Generales
- De resumen
- De una entrada
- De doble entrada

Los *cuadros generales*, llamados también de referencia o primarios, presentan los datos con el mayor detalle posible, pero no pretenden destacar informaciones específicas ni demostrar aspectos concretos. También se llaman de referencia porque su función es servir como insumo para la confección de los cuadros resumen y ser fuente para otros análisis.

Los *cuadros de resumen* tienen como objetivo destacar aspectos específicos, facilitar el análisis y responder a preguntas de investigación concretas. Se derivan de los cuadros generales. En ellos se buscará que la columna matriz y los encabezados sean simples para que se capte con facilidad lo que se quiere demostrar y se noten las relaciones de interés.

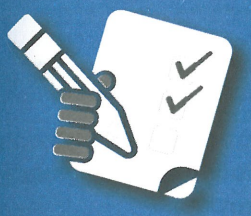
Los cuadros *de una entrada* son sencillos pues incluyen una sola clasificación, ubicada generalmente en la columna matriz. En cambio, los cuadros de *doble entrada* contienen dos o más categorías, una clasificada desde la columna matriz y la(s) otra(s) desde los encabezados de la primera fila, por ejemplo:

Cuadro 2.14

Población, defunciones y tasas de mortalidad por cien mil habitantes, según grupos de edad y sexo, Costa Rica, 2000

Grupos de edad	Población			Defunciones			Tasa de Mortalidad (1)		
	Total	Hombres	Mujeres	Total	Hombres	Mujeres	Total	Hombres	Mujeres
Total	**3 810 179**	**1 902 614**	**1 907 565**	**14 944**	**8 559**	**6 385**	**392**	**17 821**	**12 745**
00 a 04	376 584	192 287	184 297	940	538	402	250	280	218
05 a 09	411 204	210 443	200 761	77	47	30	19	22	15
10 a 14	429 019	219 467	209 552	108	60	48	25	27	23
15 a 19	392 063	198 561	193 502	246	170	76	63	86	39
20 a 24	342 728	171 679	171 049	288	206	82	84	120	48
25 a 29	295 752	146 407	149 345	312	219	93	105	150	62
30 a 34	291 499	146 377	150 361	317	229	88	109	156	59
35 a 39	266 505	141 138	147 652	392	267	125	147	189	85
40 a 44	241 262	118 853	122 409	517	354	163	214	298	133
45 a 49	183 629	90 323	93 306	545	351	194	297	389	208
50 a 54	146 024	71 727	74 297	599	362	237	410	505	319
55 a 59	104 912	51 519	53 393	741	469	272	706	910	509
60 a 64	88 142	43 089	45 053	974	582	392	1 105	1 351	870
65 a 69	71 650	34 555	37 095	1 195	703	492	1 668	2 034	1 326
70 a 74	57 641	27 765	29 876	1 530	890	640	2 654	3 205	2 142
75 y más	84 041	38 424	45 617	6 163	3 112	3 051	7 333	8 099	6 688

Fuente: CCSS (2003)



EJERCICIOS DE REPASO Y AUTOEVALUACIÓN

EJERCICIO 2.13

Encierre en un círculo la opción que se ajusta al enunciado, Falsa (F) o Verdadera (V).

F V 1. Todo proyecto de investigación comienza con la identificación de un problema, esto es, con la búsqueda de una respuesta a la pregunta ¿qué ignoramos?

F V 2. Una buena identificación del problema permite establecer el rumbo de la investigación y especificar una pregunta que demanda respuesta, a la vez que economiza recursos financieros, materiales y humanos, junto con la reducción de esfuerzos y tiempo.

F V 3. Una vez identificado el problema de investigación, no es necesario ni conveniente revisar los escritos acerca del tema o estudios relacionados puesto que contaminan el trabajo y le agregan sesgos.

F V 4. El establecimiento de los objetivos permite señalar si se va a investigar globalmente el problema o solamente una parte, además de definir el grado de profundidad que tendrá el estudio.

F V 5. Los datos son sinónimos de información puesto que son numéricos.

F V 6. Los instrumentos directos de recolección obtienen el dato en la fuente. Entre ellos están los experimentos, el registro y la observación, tanto cuantitativa como cualitativa.

F V 7. Entre los instrumentos directos de recolección se encuentran las entrevistas, las encuestas y los grupos focales.

F V 8. La confiabilidad, la validez y la objetividad deben comprobarse en todos los instrumentos de recolección de datos utilizados.

F V 9. La implementación de las pruebas piloto puede obviarse cuando el investigador ya es muy experimentado. Es más, saltarse esta etapa habla bien del dominio técnico del científico encargado.

F V 10. A los encuestadores debe dárseles cursos de capacitación específica para aclarar múltiples aspectos de la investigación y del instrumento de recolección a fin de eliminar o, al menos, minimizar los errores y sesgos en la captura de los datos.

F V 11. Un buen investigador nunca se ocupará de las tareas administrativas puesto que para eso está todo el personal de apoyo burocrático.

F V 12. Cuando se trabaja con cuestionarios, la recolección de datos puede hacerse de dos maneras: autoadministrada o por entrevista.

F V 13. La entrevista directa es más cara pero reduce las omisiones y errores de interpretación por parte del entrevistado.

F V 14. Los instrumentos estadísticos modernos bastan para una adecuada interpretación de los datos, porque se cuenta con equipos informáticos muy amigables y accesibles.

F V 15. Entre las funciones del cuestionario están el obtener datos para cumplir los objetivos planteados por la investigación y guiar al entrevistador para generar una buena relación con el informante.

F V 16. Según la naturaleza de las respuestas que se solicitan, las preguntas fundamentales del cuestionario pueden ser: de hechos, de comportamiento, de intención o de opinión.

F V 17. Sobre la extensión del cuestionario, la regla general es que si las respuestas son cortas, predeterminadas y fáciles de responder, se pueden plantear muchas; pero si son de respuesta larga, abiertas y de difícil respuesta, el cuestionario deberá ser breve.

F V 18. La contaminación de la respuesta ocurre cuando las preguntas iniciales van llevando, en un "efecto embudo", a una respuesta bastante previsible que altera en un sentido los resultados de la investigación.

F V 19. Es deseable utilizar la negación para la redacción de las preguntas dado que con esto se obliga a reflexionar al entrevistado y, por ello, proporcionará mejores respuestas.

F V 20. La tarea de codificación es más sencilla en las preguntas abiertas porque posibilitan una respuesta corta.

EJERCICIO 2.14

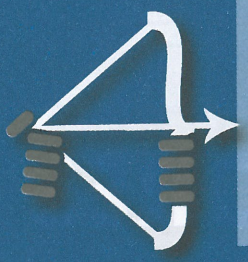
Pareo. Llene los paréntesis con el número que acompaña a cada definición de la izquierda, de forma que cada definición coincida con su concepto. Quedan dos paréntesis en blanco.

1.	Instrumento de presentación de datos, donde estos se ordenan en filas y columnas para ser comprendidos sin mucha explicación textual o gráfica.	( )	Estado de la cuestión
2.	Es el cuerpo de publicaciones que otros investigadores han escrito sobre mi objeto de estudio o tema-problema.	( )	Gráficos estadísticos
3.	Palabras sustitutivas de otra considerada ofensiva, de mal gusto o tabú.	( )	Preguntas bidimensionales
4.	Numero de ocurrencias de un determinado evento o conteo de las observaciones pertenecientes a una categoría.	( )	Cuadro estadístico
5.	Instrumentos estadísticos que comunican la información utilizando distancias, áreas y pendientes o volúmenes.	( )	Validez
6.	Son descripciones de lo que se espera del trabajo de investigación. Se plantean en forma afirmativa y con verbos en infinitivo.	( )	Confianza
7.	Solicitan una sola respuesta en un ítem que, en realidad, contiene dos preguntas.	( )	Pregunta de investigación
8.	Grado de confianza en que un instrumento de recolección de datos mide lo que pretende medir.	( )	Frecuencia
9.	Cuestionamiento general y abstracto del tema por investigar.	( )	Gráfico circular
10.	Representación estadística que informa de la participación porcentual de cada una de las categorías.	( )	Semitabular
		( )	Eufemismos
		( )	Objetivos

CAPÍTULO 3

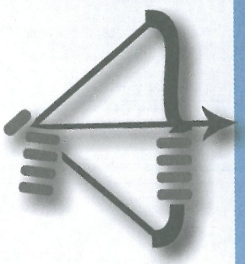
LOS INDICADORES ESTADÍSTICOS EN EL CAMPO DE LA SALUD

OBJETIVO GENERAL



Aplicar el cálculo de razones, proporciones y tasas para evaluar la gestión, como guías para el establecimiento de políticas, patrones de medida y parámetros para la implementación de correctivos en el campo de la salud.

OBJETIVOS ESPECÍFICOS



Al finalizar el estudio de este capítulo usted estará en la capacidad de:

1. Calcular razones, proporciones y tasas para generar indicadores de estado y atención en el sistema de salud.
2. Emplear las razones y proporciones en la evaluación de sensibilidad y especificidad de las pruebas de tamizaje.
3. Medir el valor predictivo de una prueba de tamizaje, tanto en sus resultados positivos como negativos.
4. Aplicar la metodología para el cálculo de tasas brutas, específicas y estandarizadas sobre hechos vitales.
5. Analizar los indicadores negativos de salud: morbilidad, incidencia, prevalencia, casos y letalidad.
6. Interpretar los datos de riesgos y oportunidades de diferentes tratamientos médicos o exposición a sustancias.
7. Ajustar tasas de hechos vitales, directa o indirectamente, para que sirvan como patrones de comparación de poblaciones heterogéneas.

Las razones, proporciones y tasas de hechos vitales se utilizan como indicadores del estado de salud, de la atención, del saneamiento ambiental o de la situación socioeconómica de una población, y sirven como guías para la medición de avances, estancamientos o retrocesos en el campo de la salud.

Mejor es la salud
que nunca se perdió.

SÉNECA

Los indicadores en el campo de la salud buscan medir la importancia cuantitativa y cualitativa de hechos vitales, eventos biodemográficos o fenómenos colectivos, con el fin de prevenir males, respaldar políticas, apoyar acciones e intervenciones, y evaluar el grado de cumplimiento de objetivos, metas y logros en la salud pública.

Los indicadores son cifras que se construyen a partir del registro de los hechos vitales y su posterior procesamiento para informar sobre la situación sanitaria de una sociedad y el desempeño del sistema de atención. El seguimiento de los indicadores permite monitorear los avances, estancamientos o retrocesos, así como el establecimiento de correctivos.

Los más utilizados en el área de la salud se pueden clasificar en indicadores de estado, de atención, de saneamiento ambiental e indicadores socioeconómicos:

- Indicadores de estado
  - Demográficos
    - Natalidad
    - Mortalidad
    - Crecimiento
  - Negativos de salud
    - Morbilidad
      - Incidencia
      - Prevalencia

- Incapacidad
- Letalidad

- Calidad de vida
  - Capacidad funcional
  - Esperanza de vida
  - Adaptación al medio

- Indicadores de atención en salud
  - Estructura de los sistemas de atención
    - Recursos materiales
    - Recursos humanos
    - Recursos financieros
    - Gestión gerencial
  - Servicios según la población
    - De atención individual
      - Urgencias
      - Consulta externa
      - Hospitalización
    - De atención familiar
    - De atención colectiva
  - Calidad del proceso
    - Productividad y eficiencia
    - Índices de hospitalización
    - Seguridad del paciente
    - Salud reproductiva y materno-infantil
    - Auxiliares de diagnóstico médico y tratamiento
  - Resultados
    - Evaluación del desempeño
    - Indicadores centinela
    - Indicadores basados en datos agrupados
    - Acceso y grado de utilización de los servicios
- Indicadores de saneamiento ambiental
  - Acceso al agua potable
  - Eliminación de excretas
  - Manejo de desechos sólidos

- Indicadores socioeconómicos
  - Ingreso nacional
    - Producción
    - Distribución
    - Pobreza
  - Ocupación
    - Desempleo
    - Mercados de trabajo
  - Educación
    - Alfabetización
    - Escolaridad
  - Vivienda y urbanismo

En algunos informes, basta con transcribir las cifras absolutas para dar cuenta de una situación. Estas cifras se obtienen por *enumeración*, es decir, por conteo o registro de los eventos, por ejemplo:

> Durante el año 2011 se abrieron 1495 Historias Clínicas, de las cuales 1209 pacientes tuvieron diagnóstico de cáncer. Las neoplasias (o tumores) más frecuentes fueron cáncer del cuello uterino con 190 casos, seguido del cáncer de mama con 183 casos, cáncer de piel con 135 casos, cáncer de estómago con 105 casos y cáncer de próstata con 82 casos. (Hernández Morón, 2012, p. 1)

Sin embargo, con mayor frecuencia es necesario comparar un número absoluto con otros datos o relacionarlo con la población de la cual forma parte, entonces se producen números relativos que suministran información condensada, los cuales se constituyen en valiosos instrumentos para comparar las tendencias a lo largo del tiempo o para analizar similitudes y diferencias entre dos o más poblaciones.

3.1 LOS NÚMEROS RELATIVOS

Los números relativos permiten transformar grandes volúmenes de datos en cifras manejables, lo cual facilita su uso, permite el establecimiento de relaciones significativas, aumenta la utilidad de los datos y favorece el análisis de las cifras al:

- Resumir grandes volúmenes de datos.
- Poner en evidencia las relaciones entre dos o más números.
- Posibilitar la comparación entre grupos de la misma población.
- Comparar datos tomados de poblaciones diferentes.

Entre los números relativos de uso generalizado están las:

- Razones
- Proporciones
- Tasas

3.1.1. LAS RAZONES

Una *razón* es el dato que resulta de dividir un subgrupo de la población entre otro subgrupo o entre la población total. Por ejemplo, si el personal de enfermería de un centro de salud lo conforman 50 varones y 200 mujeres, la razón de feminidad es:

$$\text{Razón de feminidad} = \frac{\text{Número de enfermeras mujeres}}{\text{Número de enfermeros varones}}$$

$$\text{Razón de feminidad} = \frac{200}{50} = 4$$

En este centro de salud brindan servicios de enfermería 4 mujeres por cada varón.

En términos generales, una razón informa cuantas veces es mayor una clase con respecto de otra. También puede calcularse la razón de masculinidad:

$$\text{Razón de masculinidad} = \frac{\text{Número de enfermeros varones}}{\text{Número de enfermeras mujeres}} \times 100$$

$$\text{Razón de masculinidad} = \left(\frac{50}{200}\right) \times 100 = 25$$

La razón de masculinidad entre el personal de enfermería es de 25 varones por cada 100 mujeres.

Con el fin de hacer más comprensibles los indicadores, pueden ser multiplicados por una *constante de amplificación*, es decir, por un valor arbitrario que usualmente es múltiplo de 10 (los más frecuentes son 100, 1 000 y 100,000).

Muchos indicadores en el campo de la salud son razones, ya que estas permiten informar aspectos de la estructura de atención como: la razón de médicos por cada 1 000 habitantes, la razón de médicos por consultorio o la razón de personal de enfermería por médico.

$$\text{Médicos por habitantes} = \frac{\text{Total de médicos}}{\text{Total de habitantes}} \times 1000$$

$$\text{Médicos por consultorio} = \frac{\text{Médicos en contacto con pacientes}}{\text{Cantidad de consultorios}}$$

$$\text{Personal de enfermería por médico} = \frac{\text{Número de enfermeros}}{\text{Número de médicos}}$$

Las razones sirven como indicadores de proceso; por ejemplo, la razón de morbilidad materna severa que relaciona el número de mujeres que presenta complicaciones severas (hemorragia ante, intra o postparto, embolia obstétrica, complicaciones por anestesia durante el parto, sepsis puerperal o rupturas) con el total de nacidos vivos sería:

$$\text{Razón de morbilidad materna severa} = \frac{\text{Mujeres que presentan complicaciones severas}}{\text{Total de nacidos vivos}}$$

Y también se utilizan como indicadores de productividad, por ejemplo, la razón de consultas por médico, de consultas por especialista o de consultas por odontólogo:

$$\text{Consultas por médico} = \frac{\text{Número de consultas externas}}{\text{Médicos en contacto con pacientes}}$$

EJERCICIO 3.1

a) Un cantón fronterizo cuenta con una población de 25,857 habitantes y tiene destacado el siguiente personal de salud: 8 médicos, 4 enfermeras, 12 auxiliares de enfermería, 1 odontólogo y 15 asistentes de salud rural. Calcule e interprete:

   i La razón de habitantes por cada médico.

   ii La razón de médicos por cada 1 000 habitantes.

   iii La razón de habitantes por trabajador de la salud.

b) Un hospital privado cuenta con los servicios a tiempo completo de 31 médicos y 124 enfermeros, encuentre la razón de enfermeros a médicos.

c) En una sala de traumatismo ingresan diariamente 32 pacientes, de los cuales 12 son operados y 20 no requieren de cirugía. Encuentre la razón de operados por cada 10 no operados.

d) La Unidad de Psiquiatría de un hospital dispone de 24 camas para el internamiento de pacientes y registra 864 egresos por año. Calcule la razón de egresos anuales por cama.

e) Cada uno de los 31 médicos contratados trabaja 252 días al año, con lo cual se logra atender 156,240 consultas externas por año. Con esos datos, calcule la razón consultas por día por médico.

3.1.2. LAS PROPORCIONES

Una *proporción* informa del resultado de dividir un subgrupo entre toda la población. Las proporciones reflejan la importancia relativa de cada clase dentro de la población. De esta forma, la proporción de médicos en labores administrativas se obtiene dividiendo el número de médicos que trabaja administrativamente entre el total de médicos:

$$\text{Proporción de médicos en labores administrativas} = \frac{\text{Médicos en labores administrativas}}{\text{Total de médicos}}$$

Si, por ejemplo, la seguridad social tiene contratados a 3 183 médicos, de los cuales 573 se desempeñan como directores y jefes, entonces la proporción de médicos en labores administrativas es:

$$\frac{573}{3183} = 0,18$$

Obsérvese que una proporción también es una razón, pero con algunas características particulares:

- Relaciona datos de la misma población y de la misma naturaleza.
- Compara una parte con el todo.
- Produce datos en el rango de 0 a 1, al informar el porcentaje que el numerador ocupa del todo.
- Si se amplifica por 100, el rango irá de 0 a 100 %.

Resulta importante mencionar que el término porcentaje no es más que una proporción amplificada por la constante 100; así, la razón 0,18 es equivalente a:

$$\left(\frac{573}{3183}\right) \times 100 = 18\%$$

El 18 % del personal médico está dedicado a tareas administrativas.

Las proporciones son muy útiles como indicadores de proceso, por ejemplo, la proporción de pacientes referidos a otros niveles de atención médica o la proporción de cirugías ambulatorias:

$$\text{Proporción de pacientes referidos} = \frac{\text{Número de pacientes referidos}}{\text{Total de pacientes atendidos}} \times 100$$

$$\text{Proporción de cirugías ambulatorias} = \frac{\text{Número de cirugías ambulatorias}}{\text{Total de intervenciones quirúrgicas}} \times 100$$

Informan sobre la cobertura de los servicios de salud al dividir la población que efectivamente recibe el servicio entre la población que idealmente debería ser atendida, por ejemplo, la proporción de niños vacunados o la cobertura de consulta prenatal:

$$\text{Proporción de niños vacunados} = \frac{\text{Número de niños vacunados con DPT}}{\text{Total de niños de una región}} \times 100$$

$$\text{Cobertura de consulta prenatal} = \frac{\text{Mujeres embarazadas en control}}{\text{Total de embarazadas}} \times 100$$

También, las proporciones sirven como indicadores de hospitalización, como el porcentaje de ingresos hospitalarios por motivo y la proporción de egresos hospitalarios de corta o larga estancia:

$$\text{Porcentaje de ingresos por motivo} = \frac{\text{Ingreso por el motivo X}}{\text{Total de ingresos hospitalarios}} \times 100$$

$$\text{Proporción de egresos de corta estancia} = \frac{\text{Total de egresos de corta estancia}}{\text{Total de egresos hospitalarios}} \times 100$$

EJERCICIO 3.2

a) En el primer trimestre de 2012, por causas no médicas, fueron suspendidas un total de 278 cirugías de 3,475 que estaban programadas. Encuentre la proporción de cirugías suspendidas e indique qué se reconoce con este índice.

b) Para identificar deficiencias en los tratamientos médico-quirúrgicos se calcula el porcentaje de reintervenciones quirúrgicas. En este trimestre, de un total de 5,280 cirugías efectuadas, 264 se realizaron a pacientes operados por la misma dolencia en el mismo periodo. Calcule la proporción de reintervenciones quirúrgicas del trimestre.

c) El Dr. Pedro Hernández Morón (2012, p. 1), jefe del Departamento de Control de Cáncer de la Región Norte de Perú, informa que "durante el año 2011 se abrieron 1,495 Historias Clínicas, de las cuales 1209 pacientes tuvieron diagnóstico de cáncer. Las neoplasias (o tumores) más frecuentes fueron cáncer del cuello uterino con 190 casos, seguido del cáncer de mama con 183 casos, cáncer de piel con 135 casos, cáncer de estómago con 105 casos y cáncer de próstata con 82 casos". A partir de esos datos, construya una tabla que informe las proporciones que estos tipos de cáncer de mayor frecuencia representan del total de diagnosticados con cáncer.

d) 2,568 pacientes de consulta externa son referidos a otras especialidades médicas; de ellos, 463 se envían a consultorios de endocrinólogos. Informe la proporción de pacientes referidos a endocrinología.

3.1.2.1. Aplicación de razones y proporciones: la prueba de tamizaje

El *tamiz* es un cedazo muy tupido que sirve para colar y *tamizar* equivale a separar lo fino de lo grueso. En medicina se aplican muchas pruebas iniciales de diagnóstico para identificar pacientes que padecen o pueden desarrollar una enfermedad y los pacientes que no la padecen. Estas pruebas se conocen como *pruebas de tamizaje*.

Las pruebas médicas pueden ser evaluaciones completas, invasivas y costosas en términos de tiempo, personal, materiales y dinero, o una prueba de tamizaje que, según la OMS, es "una prueba sencilla en una población saludable, para identificar a aquellos individuos que tienen alguna patología, pero que todavía no presentan síntomas" (Galván, p. 2).

Por ejemplo, a 79 pacientes se les efectúa una ecografía para diagnosticar cáncer gástrico, de las cuales se obtienen 8 ecografías positivas y 71 negativas. Gastroscopías y biopsias posteriores proporcionan datos que permiten construir la siguiente tabla de contingencia:

Cuadro 3.1

Prueba de tamizaje (Ecografía gástrica)

	Con cáncer gástrico	Pacientes sanos	TOTAL
Ecografía positiva	6	2	8
Ecografía negativa	1	70	71
	7	72	79

Fuente: Adaptado con propósitos didácticos de Díaz Vélez (2010)

Estos grupos pueden designarse de la siguiente manera:

Cuadro 3.2

Prueba de tamizaje: Nombres de los grupos

	Con cáncer gástrico	Pacientes sanos	TOTAL
Ecografía positiva	Verdadero Positivo (VP)	Falso Positivo (FP)	8
Ecografía negativa	Falso Negativo (FN)	Verdadero Negativo (VN)	71
	7	72	79

La *sensibilidad* de esta prueba de tamizaje se obtiene por la proporción de pacientes con cáncer gástrico que se identifica por medio de una ecografía positiva. En otros términos, se obtiene por la división de los Verdaderos Positivos (VP) entre los realmente enfermos.

$$\frac{VP}{VP + FN} \times 100$$

$$\frac{6}{7} \times 100 = 85,7\%$$

Una ecografía es capaz de identificar al 85,7 % de los pacientes enfermos de cáncer gástrico. La ecografía, en tanto que prueba de tamizaje, demuestra tener alta sensibilidad.

La *especificidad* se obtiene por la proporción de pacientes sanos que resulta negativo en la prueba de tamizaje. En otras palabras, por la división de los Verdaderos Negativos (VN) entre todos los pacientes sanos.

$$\frac{VN}{FP + VN} \times 100$$

$$\frac{70}{72} \times 100 = 97,2\%$$

Una ecografía es capaz de declarar sanos al 97,2 % de los pacientes que realmente no padece cáncer gástrico. Recuerde que la especificidad informa la proporción de pacientes sin la enfermedad que tiene resultado negativo de la prueba de tamizaje. La sensibilidad mide la validez de la prueba de tamizaje entre los enfermos, mientras que la especificidad mide su validez entre los sanos.

Cuero (2012) sugiere que "en términos generales una prueba muy sensible se elige cuando la enfermedad en la que se aplica es grave pero curable o cuando un resultado falso positivo no supone para el individuo ningún perjuicio psicológico ni económico" (p. 27). "En cambio, una prueba muy específica se elige cuando la enfermedad es grave y prácticamente incurable, cuando interesa descartar la enfermedad o cuando un resultado falso positivo supone para el individuo un perjuicio económico o psicológico grave" (p. 28).

Una vez que la prueba de tamizaje haya resultado positiva o negativa, se requiere saber la probabilidad de que el paciente padezca o no de la enfermedad y para ello se calcula el Valor Predictivo de la Prueba. Esto es necesario puesto que, en términos generales, la prueba de tamizaje se efectúa cuando el paciente está sano o, si padece de la enfermedad, es asintomático y se desconoce su estado real.

El *Valor Predictivo de una Prueba Positiva* (VPPP) es el número de pacientes que está realmente enfermo y fue diagnosticado así en la prueba de tamizaje, dividido entre todos los que tuvieron ecografía positiva. Esto es, la proporción de los verdaderos positivos (VP) entre todos los positivos por tamizaje (sospechosos de tener la enfermedad).

$$\frac{VP}{VP + FP} \times 100$$

$$\frac{6}{8} \times 100 = 75\%$$

Únicamente el 75 % de las ecografías positivas corresponden a pacientes realmente enfermos.

El *Valor Predictivo de una Prueba Negativa* (VPPN) es el número de pacientes que está sano y fue diagnosticado sano, dividido entre todos los que presentaron ecografía negativa. Esto es, la proporción de los verdaderos negativos (VN) entre todos los negativos según la prueba de tamizaje.

$$\frac{VN}{FN + VN} \times 100$$

$$\frac{70}{71} \times 100 = 98{,}6\,\%$$

El 98,6 % de los pacientes cuyas ecografías resultaron negativas no tiene realmente el padecimiento.

3.1.2.2. Presunciones para la validación de pruebas de tamizaje

Los presupuestos para el establecimiento de pruebas de tamizaje fueron planteados por Wilson y Jurger (1968), de los cuales Galván Barahona (2009) reproduce el siguiente resumen:

Requisitos exigibles a los programas de tamizaje:

1. Conocimiento de la enfermedad
   - Debe ser un problema importante (como causa de morbilidad o de mortalidad).
   - Las etapas latentes o la sintomatología inicial deben ser detectables.
   - La historia natural de la condición, incluyendo el desarrollo desde la fase de latencia a la de las manifestaciones, debe comprenderse suficientemente.
2. Conocimiento de la prueba
   - La prueba o examen debe ser válida y reproducible.
   - La prueba es aceptable para la población.
   - El proceso de búsqueda de casos debe ser continuo y no único.
3. Tratamiento de la enfermedad
   - Aceptable para los pacientes diagnosticados.
   - Disponibilidad de recursos para el diagnóstico y tratamiento.
   - Acuerdo sobre el tratamiento de los pacientes.
4. Consideraciones económicas
   - El coste de la detección (incluido el del diagnóstico y tratamiento de los positivos) debe ser equilibrado en relación con el conjunto de gasto sanitario (p. 5).

EJERCICIO 3.3

a) 1,400 varones mayores de 40 años son examinados por medio del Tacto Rectal como prueba de tamizaje para detectar cáncer de próstata, de estos resultaron 540 con agrandamiento sospechoso y 860 sin agrandamiento. Dado que la enfermedad es de alta prevalencia y de gran letalidad, posteriormente se les realizan biopsias prostáticas, lo cual genera la siguiente tabla de contingencia:

	Resultado de la biopsia prostática		
Resultado del tacto rectal	Cáncer	Benigno	Total
Agrandamiento	460	80	540
Sin agrandamiento	200	660	860
	660	740	1 400

Calcule:

i. La sensibilidad del tacto rectal como prueba de tamizaje.

ii. La especificidad de la prueba.

iii. El valor predictivo de una prueba positiva (que detecta agrandamiento prostático).

iv. El valor predictivo de una prueba negativa (que no revela agrandamiento de la próstata).

b) A 700 pacientes se les practica una baciloscopía como prueba de tamizaje para la detección de la tuberculosis resultando 20 positivos y 680 negativos. Luego se efectúan cultivos de muestras no pulmonares, generando la siguiente tabla de contingencia:

	Cultivo de muestras no pulmonares		
Baciloscopía	Enfermo	Sano	Total
Positiva	19	1	20
Negativa	41	639	680
	60	640	700

Calcule:

i. La sensibilidad de la baciloscopía como prueba de tamizaje para la detección temprana de la tuberculosis.

ii. La especificidad de la prueba.

iii. El valor predictivo de una prueba positiva (que sospecha paciente enfermo).

iv. El valor predictivo de una prueba negativa (que sospecha paciente sano).

c) 290 mujeres con Síndrome Hipertensivo por Embarazo enfrentan el riesgo de que el feto sea pequeño para la edad de gestación. Para diagnosticarlo, son sometidas a ultrasonografía como prueba de tamizaje, en la cual se detectaron 10 con retraso del crecimiento fetal y 280 sin retraso del crecimiento fetal. Al nacer, los bebes son pesados y se obtienen los datos siguientes:

	Peso del recién nacido		
Ultrasonografía	Bajo Peso	Sanos	Total
Retraso del Crecimiento Fetal	8	2	10
Sin Retraso en el Crecimiento Fetal	40	240	280
	48	242	290

Calcule:

i. La sensibilidad de la ultrasonografía como prueba de tamizaje para detectar el retraso del crecimiento fetal.

ii. La especificidad de la prueba de tamizaje usando ultrasonografía.

iii. El valor predictivo de una prueba positiva (que diagnostica retraso en el crecimiento fetal).

iv. El valor predictivo de una prueba negativa (que no prevé retraso en el crecimiento del feto).

3.1.3. LAS TASAS

Una *tasa* indica la frecuencia relativa con que un evento ocurre en una población y siempre se calcula con referencia a un periodo de análisis y zona geográfica determinada (tiempo y espacio), convencionalmente el tiempo es de un año. Ellas informan el riesgo de que el fenómeno suceda en ese periodo y zona geográfica. Las tasas (de nacer, enfermarse, morir, etc.) pueden ser brutas, específicas o ajustadas.

Las *tasas brutas* se obtienen al dividir el número de elementos que, en el periodo, experimentó las consecuencias del riesgo entre la población total. Por ejemplo, la tasa bruta de natalidad y la tasa bruta de mortalidad.

La población de Costa Rica a mitad del año 2010 fue estimada por el Instituto Nacional de Estadísticas y Censos (INEC) en 4,563,539. Durante ese año se registraron 70,922 nacimientos vivos y 19,077 defunciones, por lo que las tasas brutas de natalidad y mortalidad fueron:

$$\text{Tasa bruta de natalidad} = \left(\frac{\text{Nacidos vivos del año}}{\text{Población a mitad del año}}\right) \times 1000$$

$$\text{Tasa bruta de natalidad} \left(\frac{70\,922}{4\,563\,539}\right) \times 1000 = 15{,}5 =$$

$$\text{Tasa bruta de mortalidad} = \left(\frac{\text{Defunciones durante el año}}{\text{Población a mitad del año}}\right) \times 1000$$

$$\text{Tasa bruta de mortalidad} = \left(\frac{19\,077}{4\,563\,539}\right) \times 1000 = 4{,}2$$

Estos datos indican que hubo 15,5 nacimientos vivos y 4,2 defunciones por cada 1,000 habitantes de Costa Rica en el año 2010.

Las *tasas específicas* se calculan para el subgrupo que generalmente contiene la población expuesta a sufrir el evento. Por ejemplo, la tasa específica de natalidad se obtiene dividiendo el número de nacidos vivos entre las mujeres de 15 a 49 años de edad (puesto que solo las mujeres pueden concebir y, generalmente, lo pueden hacer en ese rango de edad).

$$\text{Tasa específica de natalidad} \left(\frac{\text{Nacidos vivos del año}}{\text{Mujeres de 15 a 49 años}}\right) \times 1000 =$$

En el año 2010, Costa Rica contaba con 1,196,056 mujeres entre 15 y 49 años y se registraron 70,922 nacidos vivos, por lo que la tasa específica de natalidad fue:

$$\text{Tasa específica de natalidad} \left(\frac{70\,922}{1\,196\,056}\right) \times 1000$$

La tasa específica de natalidad en Costa Rica durante 2010 fue de 59,3 nacidos vivos por cada 1000 mujeres en edad fértil.

Esta tasa específica se conoce también como *tasa de fecundidad* e, igualmente, se puede calcular tasas específicas de fecundidad por edad de la madre, estas resultan de dividir el número de nacimientos de madres en un rango de edad, en el periodo, entre la población femenina de ese rango de edad.

$$\text{Tasa específica de fecundidad} = \left(\frac{\text{Nacimientos de madres de edad X en el periodo}}{\text{Población femenina de edad X en el periodo}}\right) \times 100$$

Entre las principales tasas específicas de mortalidad están:

- La tasa de mortalidad infantil
- La tasa de mortalidad por sexo
- La tasa de mortalidad por edad específica

La *tasa de mortalidad infantil* es un indicador muy relevante en el sistema de estadística para la salud pues informa sobre las muertes de niños que, habiendo nacido vivos, mueren antes de cumplir los 12 meses de vida. Para su cálculo, en el numerador irá el número de defunciones infantiles del periodo y en el denominador los nacimientos vivos registrados en este. Por ejemplo:

En el año 2010 se registraron 671 defunciones infantiles y 70 922 nacimientos vivos, por lo que la tasa de mortalidad infantil resultó ser:

$$\left(\frac{\text{Número de defunciones infantiles}}{\text{Nacidos vivos del periodo}}\right) \times 1000$$

$$\text{Tasa de mortalidad infantil}_{2010} = \left(\frac{671}{70\ 922}\right) \times 1000 = 9{,}46$$

Es decir, en Costa Rica hubo 9,46 defunciones infantiles por cada 1,000 nacimientos registrados en el año 2010. Este dato mide la probabilidad de muerte durante el primer año de vida y es muy utilizado no solo como indicador de las condiciones de salud de los lactantes, sino también como reflejo del bienestar económico y social de la población.

Para calcular la *tasa de mortalidad por sexo* se necesitan datos de población y defunciones para hombres y mujeres. Por ejemplo:

En el año 2011, de un total de 18 801 defunciones, se registraron 10 666 defunciones de hombres y 8 135 defunciones de mujeres. La población masculina al 30 de junio se estimaba en 2 340 003 y la femenina en 2 275 643, por lo que las tasas de mortalidad por sexo para ese año fueron:

$$\text{Tasa de mortalidad masculina}_{2011} = \left(\frac{10\ 666}{2\ 340\ 003}\right) \times 100\ 000 = 456$$

$$\text{Tasa de mortalidad femenina}_{2011} = \left(\frac{8135}{2\ 275\ 643}\right) \times 100\ 000 = 357$$

Obsérvese que el factor de amplificación ha sido 100 000. De esta forma, los resultados se interpretan como que en el año 2011, hubo 456 defunciones masculinas por cada 100,000 hombres y 357 defunciones femeninas por cada 100 000 mujeres.

La *tasa de mortalidad por edad específica* informa la probabilidad de morir a las diferentes edades, para ello se divide el número de muertes de un rango de edad entre el total de habitantes con esa misma edad. Por ejemplo:

En el año 2011, las defunciones de personas entre 55 y 64 años de edad fueron 4,030, de una población de 233 865 habitantes en ese mismo rango de edad, por lo que:

$$\text{Tasa de mortalidad de } 55 \text{ y } 64 \text{ años}_{2011} = \left(\frac{4030}{233\ 865}\right) \times 1000 = 17{,}2$$

En Costa Rica en el año 2011 hubo 17,2 defunciones de personas entre 55 y 64 años de edad por cada 1,000 habitantes de ese rango de edad.

EJERCICIO 3.4

a) En Costa Rica, de una población a mitad del año 2000 compuesta de 1,902,614 hombres y 1,907,565 mujeres, se registraron 14,944 defunciones (8,559 de hombres y 6,385 de mujeres). Calcule:

i. La tasa de mortalidad específica para hombres.

ii. La tasa de mortalidad específica para mujeres.

b) Utilizando los datos de la tabla siguiente.

Año	Población	Nacimientos	Defunciones	Crecimiento natural	Tasa Bruta de Natalidad	Tasa Bruta de Mortalidad
2005	4 263 479	71 650	16 034	55 616		
2006	4 401 849	71 292	16 766	54 526		
2007	4 389 139	73 634	17 067	56 567		
2008	4 451 205	70 910	17 845	53 065		
2009	4 509 290	68 407	18 302	50 105		
2010	4 563 539	66 256	18 863	47 393		
2011	4 615 518	67 034	19 469	47 565		
2012	4 667 096	67 700	20 111	47 589		

Calcule:

i. La tasa bruta de natalidad de cada año.

ii. La tasa bruta de mortalidad de 2005 a 2012.

c) En el año 2006 se contabilizaron 692 defunciones de niños menores de 12 meses. Ese mismo año se registraron 71,292 nacimientos vivos. Con esos datos, calcule la tasa de mortalidad infantil.

d) En el año 2003, los registros médicos informaron de 4344 defunciones causadas por enfermedades del sistema circulatorio, 3398 por tumores y 1909 por causas externas (accidentes, homicidios y suicidios). La población a mitad del periodo era 4,136,250 habitantes. Encuentre la tasa de mortalidad por causa de cada 10,000 habitantes:

i. Por enfermedades del sistema circulatorio.

ii. Por tumores.

iii. Por causas externas.

3.2. MORBILIDAD

La *morbilidad* es la ocurrencia de enfermedades en una población, su registro proporciona indicadores negativos del estado de salud de la colectividad.

La *tasa de morbilidad* por causa informa del número de casos de una enfermedad X, dividido entre la población total a mitad del periodo:

$$\left(\frac{\text{Número de casos de una enfermedad X}}{\text{Población total a mitad del periodo}}\right) \times 1000$$

Los indicadores de morbilidad presentan algunas particularidades que es necesario tomar en cuenta:

- Las categorías no siempre son mutuamente excluyentes porque una persona puede tener más de una enfermedad al mismo tiempo.
- Una persona puede ser afectada varias veces por el mismo mal en el periodo de análisis.
- Las categorías difícilmente son exhaustivas puesto que los registros de enfermedad no siempre proporcionan toda la información.
- Se da la existencia de enfermos que no acuden al sistema de salud.

Las fuentes básicas para calcular los indicadores de morbilidad provienen de:

- Los expedientes médicos en el sistema de salud: hospitalizaciones, urgencias o atención ambulatoria.
- El seguimiento que se hace de enfermedades bajo vigilancia epidemiológica: virus AH1N1, tuberculosis, dengue y otras.
- Los registros especiales sobre enfermedades con alta incidencia, prevalencia o letalidad: tumores, VIH/sida, etcétera.
- Estadísticas de muertes violentas: accidentes, homicidios y suicidios.
- Los registros de una enfermedad pueden ser de casos nuevos o de casos existentes. Un caso nuevo puede comenzar y terminar en el período analizado o iniciarse en el período y seguir hasta el siguiente u otros períodos. En estas dos situaciones, se estaría estudiando la *incidencia* de una enfermedad, puesto que interesan los casos nuevos del período en una zona geográfica X.

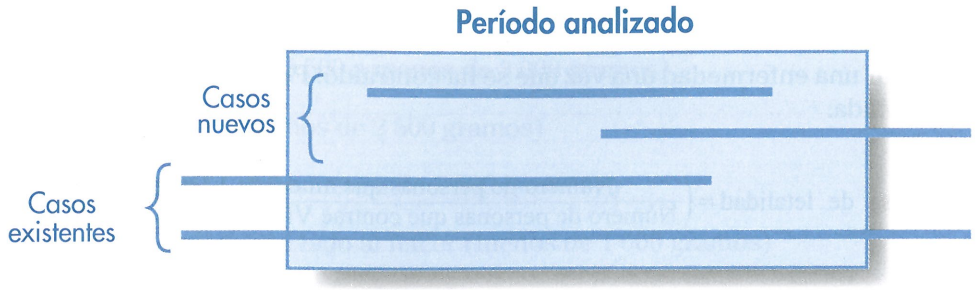


Figura 3.1. **Incidencia o prevalencia de una enfermedad determinada**
Esquema desarrollado con fines didácticos

Algunos casos pueden comenzar antes del período de estudio y terminar en él o incluso continuar durante todo el período de estudio y hasta seguir a otros períodos. Cuando se busca saber el número de individuos padeciendo una enfermedad, sin tomar en cuenta la fecha de su inicio, se está estudiando la *prevalencia* de ese mal. La tasa de prevalencia informa de todos los casos de enfermedad que no han terminado en la muerte, la cura o la remisión, además de los casos nuevos reportados en el período de estudio (Burgos, 2012, p. 108).

La *tasa de incidencia* es el número de *casos nuevos* de una enfermedad durante un periodo por cada 100 000 habitantes en condiciones de sufrir el riesgo. Por ejemplo, la tasa de incidencia de cáncer de próstata es:

$$\text{Tasa de incidencia} = \left(\frac{\text{Casos nuevos de cáncer de próstata en el periodo}}{\text{Población en riesgo}}\right) \times 100\,000$$

La *tasa de prevalencia* informa del número de pacientes que *padece* de una determinada enfermedad, en un periodo de tiempo y zona geográfica determinados, por cada 100 000 habitantes. Por ejemplo, la tasa de prevalencia de la tuberculosis es:

$$\text{Tasa de prevalencia} = \left(\frac{\text{Personas que padecen tuberculosis}}{\text{Población}}\right) \times 100\,000$$

La tasa de prevalencia viene a ser como una fotografía que muestra los casos existentes en un momento determinado.

La *tasa de casos* es un tipo particular de tasa de incidencia, es utilizada cuando existe la posibilidad de que algunas personas contraigan la misma enfermedad más de una vez en el periodo de estudio. Se obtiene al dividir la cantidad de casos declarados de una enfermedad X en el periodo, entre la población expuesta, multiplicado por 10 000 habitantes. Por ejemplo, la tasa de casos de dengue:

$$\text{Tasa de casos} = \left(\frac{\text{Cantidad de casos de dengue en el periodo}}{\text{Población total}}\right) \times 10\,000$$

La *tasa de letalidad* es el porcentaje de personas que contrae una enfermedad X y muere a causa de ella en un periodo determinado, es decir, mide la probabilidad de morir por una enfermedad una vez que se ha contraído. Por ejemplo, la tasa de letalidad del VIH/sida:

$$\text{Tasa de letalidad} = \left(\frac{\text{Número de personas que muere de VIH/sida}}{\text{Número de personas que contrae VIH/sida en el periodo}} \right) \times 100$$

EJERCICIO 3.5

a) Calcule la tasa de incidencia de cáncer de próstata en un año determinado, sabiendo que se presentaron 647 casos nuevos y se estima que hay 1,434,589 varones mayores de edad.

b) Se sabe que 528 personas padecen de tuberculosis de una población de 4,664,311, calcule la tasa de prevalencia de esta enfermedad.

c) Las autoridades de salud reportan 7,821 casos de dengue en una población de 4,664,311, ¿cuál es la tasa de casos de dengue?

d) 13 974 personas contraen el virus del VIH/sida en un periodo, y durante el mismo se registran 3,251 muertes por esa enfermedad, calcule su tasa de letalidad.

3.3. INDICADORES PROPORCIONALES

Para el cálculo de tasas brutas, es necesario tener el dato de la población total a medio periodo y, para tasas específicas, conocer el tamaño de la población en riesgo de enfrentar la situación (nacer, morir o enfermarse). El problema es que, con mucha frecuencia, no se dispone de estas cifras al momento de explicar el estado de la salud colectiva, tampoco se puede delimitar con claridad a la población expuesta o esta resulta ser muy pequeña, razón por la cual se tiene que recurrir a los indicadores proporcionales.

3.3.1. INDICADORES PROPORCIONALES DE NATALIDAD

Un indicador de la calidad de vida y de la eficacia de los sistemas de atención de la salud lo constituye el peso de los recién nacidos. Dice la UNICEF (2001): "El peso al nacer refleja la experiencia intrauterina; no solo es un buen indicador del estado de salud y la nutrición de la madre, sino también de las probabilidades de sobrevivencia, crecimiento, salud a largo plazo y desarrollo psicosocial del recién nacido" (p. 1).

Las categorías de peso al nacer sobre las cuales se debe llamar la atención son:

- Peso insuficiente (de 2 500 a menos de 3 000 gramos)
- Peso bajo (1 500 a menos de 2 500 gramos)
- Peso muy bajo (1 000 a menos de 1 500 gramos) y
- Peso extremadamente bajo al nacer (menos de 1 000 gramos)

Tenga presente que en la literatura médica, cuando se dice *peso bajo al nacer* se están acumulando todos los pesos considerados críticos, es decir, se está hablando de todos los menores de 2,500 gramos. Así lo hace la UNICEF: "Los recién nacidos con bajo peso (aquellos que pesan menos de 2,5 kg) tienen más probabilidades de morir durante los primeros meses o los primeros años. Los que sobreviven tienen disminuidas las funciones del sistema inmunológico y corren mayor riesgo de padecer posteriormente varias enfermedades, incluida la diabetes y diversas cardiopatías. También tienen propensión a seguir malnutridos y a tener menores coeficientes de inteligencia además de discapacidades cognoscitivas" (p. 1).

El cálculo de este índice es muy sencillo, solamente se requiere conocer la cifra de nacimientos del periodo y que se haya registrado su peso al nacer. Si, por ejemplo, en el año 2010 se registraron 70,922 nacimientos de los cuales 4 638 niños tuvieron un peso menor a los 2 500 gramos y 15,043 un peso de 2 500 a menos de 3 000 gramos, entonces:

$$\text{Porcentaje de bajo peso al nacer} = \left(\frac{\text{Niños con peso inferior a 2 500 gr}}{\text{Todos los nacimientos del periodo}}\right) \times 100$$

$$\text{Porcentaje de bajo peso al nacer} \left(\frac{4\,638}{70\,922}\right) \times 100 = 6{,}54\% =$$

$$\text{Porcentaje peso insuficiente} = \left(\frac{\text{Niños con peso de 2500 a menos de 3000 gr}}{\text{Todos los nacimientos del periodo}}\right) \times 100$$

$$\text{Porcentaje de peso insuficiente al nacer} \left(\frac{15\,043}{70\,922}\right) \times 100 = 21{,}21\% =$$

Cuando la proporción de parto intrahospitalario es alta, los datos simplemente se toman de los expedientes. Aún más, el porcentaje de partos con atención profesional es otro indicador de la accesibilidad a los servicios de salud.

El cuadro siguiente corresponde al informe de la Encuesta Nacional sobre Salud reproductiva que realizó el INEC en el año 2010 y, auque está basado en una muestra, señala los diferentes indicadores proporcionales de atención hospitalaria, tanto pública como privada, e indica el porcentaje de partos atendidos por médicos, obstetras o enfermeras.

Cuadro 3.3

Porcentaje de nacimientos según lugar y tipo de persona que atendió el parto –Mujeres de 15 a 49 años con al menos un parto en los últimos 5 años–

Variables	Lugar de Nacimiento		Tipo de persona que atendió el parto			Parto con atención profesional
	Hospital/ Clínica pública	Hospital/ Clínica privada	Médico	Enfermera obstetra	Enfermera	
TOTAL (N=326)	98,1	1,7	51,6	40,5	3,2	95,3
Edad						
15 a 19	100	0	36,5	55,4	6,3	98,2
20 a 24	98,4	0,8	48,5	40,7	3	92,2
25 a 29	100	0	50,5	43,1	0,7	94,3
30 a 34	93,2	6,8	59,6	34,1	6,3	100
35 a 39	97,7	2,3	55,9	38,3	2,3	96,5
Zona de residencia						
Central urbana	97,2	2,8	55,2	40,9	1,8	97,9
Central rural	95,7	3,1	48,9	39,8	6,1	94,8
Resto del país urbano	100	0	51,9	42,4	2,3	96,6
Resto del país rural	100	0	47,5	39,4	4,2	91,1

Fuente: INEC (2010). Encuesta Nacional de Salud Sexual y Reproductiva (p. 29)

3.3.2. INDICADORES PROPORCIONALES DE MORTALIDAD

De los registros de mortalidad es importante identificar la edad en que la muerte ocurre y la situación que la causa.

Los indicadores proporcionales de mortalidad a los que debe dárseles seguimiento son:

- La mortalidad proporcional infantil o de menores de un año
- La mortalidad proporcional de menores de 5 años
- La mortalidad proporcional de 60 años y más

La mortalidad proporcional infantil se calcula dividiendo el número de muertes de menores de un año entre el total de las muertes registradas en el periodo, por ejemplo, con base en los siguientes datos compilados por el Instituto Nacional de Estadísticas y Censos (INEC):

Cuadro 3.4

Total de defunciones generales por grupo de edad y sexo (Costa Rica, 2011)

Edad	Hombres	Mujeres	Total
Menos de 1	374	292	666
1 a 4	54	46	100
5 a 14	97	73	170
15 a 49	2 044	916	2 960
50 a 59	1 299	772	2 071
60 a 69	1 620	1 095	2 715
70 o más	5 172	4 940	10 112
Ignorado	6	1	7
TOTAL	10 666	8 135	18 801

Fuente: INEC (2012) <http://www.inec.go.cr/Web/Home/pagPrincipal.aspx>.

La mortalidad proporcional infantil del año 2011 fue:

$$\left(\frac{\text{Muertes de menores de un año}}{\text{Total de muertes en el periodo}}\right)\times 100$$

$$\text{Mortalidad proporcional infantil} = \left(\frac{666}{18\ 801}\right)\times 100 = 3{,}54\%$$

Este dato es un indicador de las condiciones socioeconómicas y de salubridad de una comunidad y, más que el porcentaje aislado, lo que resulta de mayor interés es observar sus variaciones en el tiempo pues la reducción o el aumento informan acerca de la situación del sistema de salud, el manejo de los partos, la atención perinatal, la salubridad y los aspectos socioeconómicos de la colectividad.

La mortalidad proporcional de menores de 5 años es un indicador del entorno social durante los primeros años de vida y proporciona pistas sobre aspectos de nutrición, factores ambientales, situación económica y educativa de la población. Se calcula dividiendo el número de muertes de menores de 5 años entre el total de defunciones del periodo.

$$\text{Mortalidad proporcional de menores de 5 años} = \left(\frac{666+100}{18\ 801}\right)\times 100 = 4{,}07\%$$

La mortalidad proporcional de 60 años o más es también un indicador del desarrollo socioeconómico y de los sistemas de atención de la salud, puesto que su aumento señala que más personas llegan a la tercera edad. Al igual que los otros indicadores proporcionales de mortalidad, no es necesario conocer la población expuesta, pues es suficiente el registro de todas las defunciones del periodo. En este ejemplo, se toman los datos del Cuadro 3.5:

$$\text{Mortalidad proporcional de 60 años o más} = \left(\frac{2\,715 + 10\,112}{18\,801}\right) \times 100 = 68{,}2\,\%$$

En Costa Rica, durante el año 2011, el 68,2 % de las defunciones fue de personas de 60 años o más.

Cuando el registro de las defunciones consigna la causa de muerte y el porcentaje de "causas indefinidas" es bajo, es de gran utilidad el cálculo de la mortalidad proporcional por causa pues permite saber la importancia relativa de las causas de muerte en una sociedad.

El INEC publica periódicamente las estadísticas vitales y, más importantes que los datos aislados de un año, son las comparaciones a lo largo del tiempo.

Cuadro 3.5

Defunciones por causas de muerte y sexo
(Costa Rica, 2008)

Causas de muerte	Hombres	Mujeres	Total	Porcentaje
Enfermedades del sistema circulatorio	2 843	2 333	5 176	28,72
Tumores (neoplasias)	2 231	1 850	4 081	22,65
Causas externas	2 054	468	2 522	13,99
Enfermedades del sistema respiratorio	813	633	1 446	8,02
Enfermedades del sistema digestivo	761	514	1 275	7,08
Enfermedades endocrinas y nutricionales	371	454	825	4,58
Enfermedades del sistema genitourinario	240	194	434	2,41
Enfermedades del sistema nervioso	279	289	568	3,15
Afecciones perinatales	205	152	357	1,98
Enfermedades infecciosas y parasitarias	223	131	354	1,96
Malformaciones congénitas	125	144	269	1,49
Trastornos mentales	93	103	196	1,09
Enfermedades sistema osteomuscular	49	114	163	0,90
Otros	183	172	355	1,97
TOTAL	10 470	7 551	18 021	100,00

Fuente: INEC (2009). Estadísticas vitales (p. 78)

La forma de calcular la mortalidad proporcional por causa es igualmente sencilla y consiste en dividir el número de defunciones por una causa particular entre el total de defunciones. De esta forma, la mortalidad proporcional por tumores en el año 2008 fue:

$$\left(\frac{4\,081}{18\,021}\right)\times 100 = 22{,}65\,\%$$

En Costa Rica, el 22,65 % de las muertes del año 2008 tuvieron como causa tumores.

3.3.3. INDICADORES PROPORCIONALES DE MORBILIDAD

Las medidas ideales de morbilidad son las tasas de incidencia y de prevalencia cuyo cálculo requiere la cuantificación de la población en riesgo, dato del que no siempre se dispone. Afortunadamente, los registros de egresos hospitalarios permiten construir indicadores proporcionales sin requerir una estimación de la población expuesta.

Los registros de egresos hospitalarios tienen ventajas de precisión en la información puesto que el diagnóstico ha sido hecho por personal médico utilizando laboratorios, imágenes y apoyo administrativo. Además, contienen información del sexo, edad, lugar de residencia, ocupación del paciente, médico tratante, tipo de admisión, tiempo de internamiento, tratamientos recibidos, resultados, eventuales complicaciones, condición de salida o causa de muerte.

A partir de estos datos, se establecen las grandes categorías de diagnóstico y, al dividir la cifra de egresos de una enfermedad entre el gran total, se obtiene un indicador proporcional de morbilidad. Por ejemplo, si en un año se registran 303,789 egresos hospitalarios en los hospitales de la Caja Costarricense del Seguro Social, de los cuales 23,058 fueron diagnosticados como enfermedades del sistema digestivo, entonces:

$$\text{Proporción de egresos}_{\text{Enfermedades sistema digestivo}} = \left(\frac{23\,058}{303\,789}\right)\times 100 = 7{,}59\,\%$$

La proporción de egresos hospitalarios atribuida a enfermedades del sistema digestivo fue de 7,59 %.

EJERCICIO 3.6

a) De 2,134 nacidos vivos en la Sección de Obstetricia de un hospital general, 47 niños pesaron menos de 1,500 gramos y 160 pesaron menos de 2,500 gramos (47 de muy bajo peso y 113 niños pesaron de 1,500 a menos de 2,500 gramos), calcule:

i La proporción de nacidos vivos con bajo peso (menos de 2,500 gramos).

ii. La proporción de nacidos vivos con muy bajo peso (menos de 1,500 gramos).

b) Con base en los datos del cuadro siguiente, calcule los diferentes indicadores proporcionales de mortalidad por edad:

Total de defunciones generales
Por grupo de edad y sexo
(Costa Rica, 2011)

Edad	Hombres	Mujeres	Total	Proporción
Menos de 1	374	292	666	
1 a 4	54	46	100	
5 a 9	32	31	63	
10 a 14	65	42	107	
15 a 19	164	72	236	
20 a 24	225	86	311	
25 a 29	293	91	384	
30 a 34	273	119	392	
35 a 39	266	144	410	
40 a 44	366	175	541	
45 a 49	457	229	686	
50 a 54	630	351	981	
55 a 59	669	421	1 090	
60 a 64	761	512	1 273	
65 a 69	859	583	1 442	
70 a 74	995	710	1 705	
75 a 79	1 148	860	2 008	
80 a 84	1 239	1 126	2 365	
85 ó más	1 790	2 244	4 034	
Ignorado	6	1	7	
TOTAL	10 666	8 135	18 801	

Fuente: INEC (2012)
<http://www.inec.go.cr/Web/Home/pagPrincipal.aspx>.

c) Basado en los datos del cuadro anterior, calcule la mortalidad proporcional de 70 años o más para Costa Rica en el año 2011. Interprete el resultado.

d) Basándose en los datos del cuadro siguiente, calcule los indicadores proporcionales de mortalidad infantil por causa de muerte:

Mortalidad infantil según causa de muerte
(Costa Rica, 2010)

Causa de muerte	Muertes	Porcentaje
Afecciones en el periodo perinatal	330	
Malformaciones congénitas	234	
Enfermedades del sistema respiratorio	38	
Enfermedades del sistema circulatorio	14	
Enfermedades infecciosas o parasitarias	12	
Enfermedades del sistema nervioso	10	
Otras causas	33	
TOTAL	671	

Fuente: INEC, Ministerio de Salud, Dirección de Vigilancia de la Salud.

e) En un año se registran 303,789 egresos hospitalarios, de los cuales 8597 correspondieron a enfermedades osteomusculares. Calcule el indicador proporcional de morbilidad respectivo.

3.4. CÁLCULOS DE RIESGOS Y OPORTUNIDADES

El factor de riesgo es una condición o rasgo detectable (en un individuo, un grupo, una familia o una comunidad) que se asocia con el aumento en la probabilidad de padecer una enfermedad, desarrollar un proceso mórbido, sufrir un accidente o enfrentar la muerte prematuramente.

Los factores de riesgo pueden ser biológicos, genéticos, medioambientales, laborales, hábitos alimenticios, costumbres, comportamientos, aspectos socioculturales, situaciones económicas o sicológicas. La presencia de uno o varios factores de riesgo aumenta la vulnerabilidad y la probabilidad de enfrentar efectos adversos en el campo de la salud, por ello se deben detectar y calcular su influencia utilizando métodos estadísticos. Estos factores de riesgo se usarán para:

- Predecir un eventual brote de enfermedades.
- Prevenir su aparición por medio del combate a los factores asociados.
- Diagnosticar utilizando medidas de prevalencia.
- Mejorar los resultados de los programas de tamizaje.
- Establecer relaciones de causa y efecto.
- Planificar los servicios de salud.

3.4.1. LA OPORTUNIDAD RELATIVA (OR = ODDS RATIO)

Supongamos la existencia de una enfermedad que de 100 pacientes a los que se les aplica un tratamiento con fármacos, 75 sobreviven y 25 mueren del mal. Luego, los científicos desarrollan una cirugía que cuando se intervienen 90 pacientes, 81 sobreviven y 9 mueren.

La oportunidad de sobrevivir utilizando cirugía es la razón de los que sanan entre los que mueren habiendo sido operados:

$$\text{Oportunidad de sobrevivir utilizando cirugía} = \frac{\text{Operados que sanan}}{\text{Operados que mueren}} = \frac{81}{9} = 9$$

Al tratarse quirúrgicamente, por cada paciente que muere, 9 tienen la oportunidad de sanar.

La oportunidad de sobrevivir usando tratamiento con fármacos es la razón de los que sanan entre los que mueren habiendo recibido tratamiento con fármacos:

$$\text{Oportunidad de sobrevivir utilizando fármacos} = \frac{\text{Tratados que sanan}}{\text{Tratados que mueren}} = \frac{75}{25} = 3$$

Utilizando fármacos, por cada paciente que muere, 3 tienen la oportunidad de sanar.

La *oportunidad relativa* (OR) es la razón del éxito terapéutico entre los dos tratamientos, esto es, la oportunidad de sobrevivir utilizando cirugía entre la oportunidad de sobrevivir utilizando fármacos:

$$\text{OR} = \frac{\text{Oportunidad de sobrevivir utilizando cirugía}}{\text{Oportunidad de sobrevivir utilizando fármacos}} = \frac{9}{3} = 3$$

El tratamiento con cirugía ofrece una oportunidad relativa (OR) de sobrevivencia 3 veces mayor que usando fármacos.

3.4.2. EL RIESGO RELATIVO (RR)

El *riesgo relativo* (RR) es una razón que surge de comparar dos tasas. En el ejemplo, se calcula dividiendo la tasa de sobrevivencia con cirugía, entre la tasa de sobrevivencia con fármacos.

Cuadro 3.6

Tabla de contingencia para cálculo de riesgos y oportunidades

	Sobreviven	Mueren	TOTAL
Cirugía	81	9	90
Fármacos	75	25	100
	156	34	190

La tasa de sobrevivencia de los pacientes que se someten a cirugía es:

$$\frac{\text{Sobrevivientes}}{\text{Población tratada}} \times 100 = \frac{81}{90} \times 100 = 90\%$$

La tasa de sobrevivencia de los pacientes que siguen un tratamiento con fármacos es:

$$\frac{\text{Sobrevivientes}}{\text{Población tratada}} \times 100 = \frac{75}{100} \times 100 = 75\%$$

Por lo que el riesgo relativo es:

$$RR = \frac{90}{75} = 1,2$$

El riesgo relativo de sanar con cirugía es 1,2 veces respecto de los pacientes tratados con fármacos.

3.4.3. EL RIESGO ATRIBUIBLE (RA)

El *riesgo atribuible* (RA) es la diferencia entre las tasas que se están comparando. En este ejemplo, a la tasa de sobrevivencia usando cirugía se le resta la tasa de sobrevivencia usando fármacos.

$$RA = 90\,\% - 75\,\% = 15\,\%$$

La probabilidad de morir se reduce en 15 % en un paciente que se somete a cirugía en lugar de ser tratado con fármacos.

3.4.4. LA FRACCIÓN ATRIBUIBLE (FA)

La *fracción atribuible* (FA) también se conoce como fracción etiológica, pues estima el grado de influencia que tiene la cirugía como causa de la curación de los pacientes y que puede ser atribuible al hecho de ser operado. Se calcula por la división del riesgo relativo menos uno (RR – 1) entre el riesgo relativo (RR):

$$FA = \frac{RR - 1}{RR}$$

$$FA = \frac{1{,}2 - 1}{1{,}2} = 0{,}1667$$

El 16,67 % de los sanados se debe al impacto de la cirugía en el tratamiento de esta enfermedad.

EJERCICIO 3.7

a) Se sospecha que la exposición pasiva al humo del tabaco en mujeres embarazadas aumenta el riesgo de que el bebé tenga bajo peso al nacer. En una muestra de 1,323 embarazadas, 426 estaban expuestas al humo del tabaco y 897 no estaban expuestas; al nacer los bebés, estos fueron los resultados:

	Bajo peso	Peso adecuado	TOTAL
Expuestas	10	416	426
No expuestas	7	890	897
	17	1 306	1 323

Calcule e interprete:

i. La oportunidad relativa (OR) de tener bajo peso al nacer.

ii. El riesgo relativo (RR) de bajo peso por exposición al humo del tabaco.

iii. El riesgo atribuible (RA) al humo del tabaco.

iv. La fracción atribuible (FA) a la exposición al humo del tabaco.

b) Una investigación con pacientes hipertensos busca las asociaciones entre sufrir un infarto y el tratamiento con Enalapril. Del seguimiento a 1,183 pacientes se han obtenido los datos siguientes:

	Infarto	No infarto	TOTAL
Sin tratamiento	42	498	540
Con tratamiento	10	633	643
	52	1 131	1 183

Obtenga e interprete:

i. La oportunidad relativa (OR) de sufrir un infarto.

ii. El riesgo relativo (RR) de infartarse por no someterse al tratamiento.

iii. El riesgo atribuible (RA) de no tener tratamiento.

iv. La fracción atribuible (FA) a la falta de tratamiento.

c) Un ensayo clínico con 200 pacientes para probar la eficacia de un antidepresivo informa los datos siguientes:

	Mejora	No mejora	TOTAL
Antidepresivo	66	22	88
Placebo	44	88	132
	110	110	220

Calcule e interprete:

i. La oportunidad relativa (OR) de mejorar en el padecimiento.

ii. El riesgo relativo (RR) de mejorar por el tratamiento con el antidepresivo.

iii. El riesgo atribuible (RA) al efecto del antidepresivo.

iv. La fracción atribuible (FA) de la mejoría por exposición al antidepresivo.

3.5. TASAS AJUSTADAS

Cuando se comparan tasas es necesario tomar en cuenta algunos aspectos, entre ellos, las diferencias en la estructura de edad de las poblaciones, las diferencias económicas o los cambios en las tendencias a través del tiempo.

Tómese, por ejemplo, las siguientes poblaciones hipotéticas:

Cuadro 3.7

	POBLACIÓN A			POBLACIÓN B		
Edad	Población	Porcentaje	Defunciones	Población	Porcentaje	Defunciones
0 a 9	3 000	30,00	450	600	5,00	300
10 a 19	2 500	25,00	250	1 200	10,00	300
20 a 29	2 000	20,00	100	3 000	25,00	150
30 a 39	1 200	12,00	30	4 080	34,00	102
40 a 49	700	7,00	28	1 500	12,50	60
50 a 59	400	4,00	92	1 200	10,00	30
60 y más	200	2,00	50	420	3,50	58
	10 000	100	1 000	12 000	100,00	1 000

La Población A es más joven ya que el 75 % de sus habitantes tiene menos de 30 años, en tanto que en la Población B solamente el 40 % tiene menos de 30 años y el 60 % de sus habitantes tiene 30 años o más. Por tanto, no es conveniente comparar las tasas brutas de mortalidad y concluir que la situación de atención de la salud es mejor en una población que en otra.

Para poder comparar tasas es necesario ajustarlas para tener indicadores estandarizados de comparación, esto es, indicadores que informan la tasa "como si" las estructuras de población fueran semejantes.

En general, se emplean dos métodos de estandarización o ajuste de tasas:

- El método directo
- El método indirecto

Conviene advertir, que el resultado *no* es una tasa real de la población o del periodo, sino que es una tasa hipotética, es decir, una tasa de "lo que sería" si la población fuera como la población tipo.

3.5.1. MÉTODO DIRECTO DE AJUSTE DE TASAS

Teniendo las poblaciones A y B y el informe de las defunciones del periodo, se calculan las tasas de mortalidad específicas por edad y la tasa bruta de mortalidad:

Cuadro 3.8

POBLACIÓN A				POBLACIÓN B		
Edad	Población	Defunciones	Tasa por 1000	Población	Defunciones	Tasa por 1000
0 a 9	3 000	450	150,00	600	300	500,00
10 a 19	2 500	250	100,00	1 200	300	250,00
20 a 29	2 000	100	50,00	3 000	150	50,00
30 a 39	1 200	30	25,00	4 080	102	25,00
40 a 49	700	28	40,00	1 500	60	40,00
50 a 59	400	92	230,00	1 200	30	25,00
60 y más	200	50	250,00	420	58	138,10
	10 000	1 000	100,00	12 000	1 000	83,33

La primera tentación con estos datos es señalar 1,000 defunciones en A y 1,000 defunciones en B y sugerir una igualdad en los patrones de mortalidad.

Puesto que el tamaño de las poblaciones es diferente (10,000 < 12,000) las tasas brutas de mortalidad resultan de 100 por cada 1000 habitantes en A y 83,33 por cada 1000 habitantes en B. Tasas que tampoco son directamente comparables puesto que las estructuras por edad son diferentes.

Para ajustar las tasas, se debe calcular las defunciones esperadas suponiendo que la población es otra. Para hacerlo se puede usar alguna las siguientes poblaciones tipo:

- Una de las dos poblaciones (generalmente la numéricamente mayor)
- La suma de las dos poblaciones
- Una población ideal

Se comienza con el cálculo de las defunciones esperadas de A "como si" tuviera la estructura de edades y el tamaño de la población de B (esto es, usando la primera acepción).

La tasa específica de mortalidad del grupo de 0 a 9 años es de 150 por mil en A y el tamaño de ese grupo de edad en B es de 600 habitantes. Por ello, las muertes esperadas en A si su población fuera igual que la de B es:

$$\text{Muertes esperadas}_{0\text{ a }9\text{ años}} = \frac{(\text{Tasa específica en A}_{0\text{ a }9\text{ años}}) \times (\text{Población por edad en B}_{0\text{ a }9\text{ años}})}{1000}$$

$$\text{Muertes esperadas}_{0\text{ a }9\text{ años}} = \frac{150 \times 600}{1000} = 90$$

Se esperarían 90 defunciones en A si su grupo etáreo de 0 a 9 años fuera de 600 habitantes.

Así se continúa con todas las edades, de lo que resulta:

Cuadro 3.9

	POBLACIÓN A			POBLACIÓN B			
Edad	Población	Defunciones	Tasa por 1000	Población	Defunciones	Tasa por 1000	Defunciones esperadas
0 a 9	3 000	450	150,00	600	300	500,00	90,00
10 a 19	2 500	250	100,00	1 200	300	250,00	120,00
20 a 29	2 000	100	50,00	3 000	150	50,00	150,00
30 a 39	1 200	30	25,00	4 080	102	25,00	102,00
40 a 49	700	28	40,00	1 500	60	40,00	60,00
50 a 59	400	92	230,00	1 200	30	25,00	276,00
60 y más	200	50	250,00	420	58	138,10	105,00
	10 000	1 000	100,00	12 000	1 000	83,33	**903,00**

Al sumar todas las defunciones esperadas de A (si tuviera la estructura de población de B) se obtiene un total de 903 defunciones, por lo que la tasa de mortalidad ajustada de A es:

$$\text{Tasa de mortalidad ajustada} = \left(\frac{\text{Defunciones esperadas}}{\text{Población Tipo}}\right) \times 1000$$

$$\text{Tasa de mortalidad ajustada}_{\text{de A}} = \left(\frac{903}{12\,000}\right) \times 1000 = 75{,}25$$

Resumiendo las tasas de mortalidad para ambas poblaciones, antes y después del ajuste, los datos son los siguientes:

Cuadro 3.10

	ANTES DEL AJUSTE	DESPUÉS DEL AJUSTE
POBLACIÓN A	100,00	75,25
POBLACIÓN B	83,33	83,33
DIFERENCIA	16,67	-8,08
COCIENTE	1,2	0,90

Antes de la estandarización de la tasa de mortalidad de A, la diferencia entre las tasas era de 16,67 por mil (100 – 83,33) o la tasa de A 20 % mayor $\left(\frac{100}{83,33} = 1,2\right)$ que la de B.

Cuando la mortalidad de la Población A se pondera usando la estructura por edades de B, la tasa ajustada de mortalidad disminuye a 75,25 por mil, que al comparar la mortalidad en las dos poblaciones, la de A resulta 8,08 por mil menor que la tasa de mortalidad de B.

Otra forma de decirlo es usando el cociente, entonces, la tasa ajustada de mortalidad de A representa el 90 % de la de B $\left(\frac{75,25}{83,33} = 0,9\right)$.

Recuérdese que la tasa ajustada *solamente* sirve para propósitos comparativos en poblaciones heterogéneas y que variará cuando se utilicen otras poblaciones tipo.

EJERCICIO 3.8

a) Compare las tasas ajustadas de mortalidad tanto de A como de B, estandarizando ambas tasas con la suma de las dos poblaciones:

Edad	Tasa de A por 1000	Tasa de B por 1000	Suma de poblaciones	Muertes esperadas en A	Muertes esperadas en B
0 a 9	150,00	500,00	3 600		
10 a 19	100,00	250,00	3 700		
20 a 29	50,00	50,00	5 000		
30 a 39	25,00	25,00	5 280		
40 a 49	40,00	40,00	2 200		
50 a 59	230,00	25,00	1 600		
60 y más	250,00	138,10	620		
TOTAL	100,00	83,33	22 000		

Complete el cuadro resumen y haga las comparaciones entre tasas de mortalidad, antes y después del ajuste:

	ANTES DEL AJUSTE	DESPUÉS DEL AJUSTE
POBLACIÓN A	100,00	
POBLACIÓN B	83,33	
DIFERENCIA	16,67	
COCIENTE	1,2	

b) Dos comunidades (X y Y) presentan las siguientes distribuciones de grupos laborales entre sus respectivas poblaciones:

i. Calcule la composición porcentual de cada grupo laboral.

Grupos laborales	Población X	Composición porcentual de X	Población Y	Composición porcentual de Y
Empresarios	1 000		3 000	
Profesionales	1 000		6 000	
Técnicos	2 000		6 000	
Obreros	4 000		12 000	
Campesinos	12 000		3 000	
TOTAL	20 000	100	30 000	100

ii. Los trabajadores del sector salud han identificado pacientes que sufren de alcoholismo severo en los diferentes grupos ocupacionales. Los datos se presentan en el cuadro siguiente. Calcule las tasas de prevalencia de alcoholismo.

Grupos laborales	Población X	Alcohólicos	Tasa de prevalencia alcoholismo	Población Y	Alcohólicos	Tasa de prevalencia alcoholismo
Empresarios	1 000	10		3 000	60	
Profesionales	1 000	40		6 000	60	
Técnicos	2 000	30		6 000	120	
Obreros	4 000	200		12 000	300	
Campesinos	12 000	120		3 000	60	
TOTAL	20 000	400		30 000	600	

iii. Dada la estructura laboral de esas dos poblaciones, ¿por qué considera necesario ajustar las tasas de prevalencia de alcoholismo?

iv. Considerando como población tipo la propuesta en la columna 4 del cuadro siguiente, calcule el número esperado de alcohólicos tanto en X como en Y.

Clase social	Tasas en X	Tasas en Y	Población TIPO	Alcohólicos esperados en X	Alcohólicos esperados en Y
Empresarios	10,00	20,00	500		
Profesionales	40,00	10,00	1 000		
Técnicos	15,00	20,00	1 500		
Obreros	50,00	25,00	3 000		
Campesinos	10,00	20,00	4 000		
TOTAL	20,00	20,00	10 000		

v. Calcule las tasas ajustadas de prevalencia de alcoholismo en ambas poblaciones.

vi. Complete el cuadro siguiente y redacte las conclusiones de comparar la prevalencia de alcoholismo en estas dos poblaciones.

	ANTES DEL AJUSTE	DESPUÉS DEL AJUSTE
POBLACIÓN X		
POBLACIÓN Y		
DIFERENCIA		
COCIENTE		

3.5.2. MÉTODO INDIRECTO DE AJUSTE DE TASAS

Para ajustar las tasas por el *método directo*, es indispensable contar con datos que lleven a conocer las tasas específicas; además, se requiere que estas no sean muy pequeñas respecto de las poblaciones por comparar. El problema es que no siempre se cuenta con las tasas específicas y por ello se hace necesario estandarizar por el *método indirecto*.

Si se tienen los datos siguientes por edad de dos poblaciones, A y B:

Cuadro 3.11

Edad	Población A	Población B
0 a 14	14 000	8 000
15 a 24	10 000	7 000
25 a 34	6 000	10 000
35 a 44	4 000	10 000
45 a 54	3 000	6 000
55 a 64	2 000	5 000
65 o más	1 000	4 000
TOTAL	40 000	50 000

Además se sabe que las defunciones ocurridas (DO) en A fueron 300 y en B 400 durante el periodo, por lo que las tasas brutas de mortalidad son:

$$\text{Tasa bruta de mortalidad en A} = \frac{300}{40\,000} \times 1000 = 7{,}5 \text{ por mil}$$

$$\text{Tasa bruta de mortalidad en B} = \frac{400}{50\,000} \times 1000 = 8{,}0 \text{ por mil}$$

El problema es que, con esos datos, no se pueden calcular las tasas específicas de mortalidad. Además, se intuye la necesidad de estandarizar estas tasas de mortalidad porque la estructura de edad de las poblaciones es diferente: una tasa bruta mayor en B parecería lógico pues tiene una población más vieja.

La solución se consigue con el método indirecto de ajuste de tasas. Para emplearlo, se requiere de una tercera población de la cual se conozcan las tasas específicas de mortalidad, por ejemplo la Población C:

Cuadro 3.12

Edad	Población C	Defunciones	Tasas
0 a 14	100 000	300	3,0
15 a 24	90 000	180	2,0
25 a 34	70 000	140	2,0
35 a 44	50 000	250	5,0
45 a 54	40 000	400	10,0
55 a 64	30 000	510	17,0
65 o más	20 000	500	25,0
	400 000	2280	**5,7**

En el método indirecto de ajuste de tasas, se utilizan las tasas específicas de esta tercera población para el cálculo de las muertes esperadas. Recuerde que en el método directo se utilizaba una población tipo, mientras que en el indirecto se usan tasas específicas tipo.

Las defunciones esperadas (DE) en la Población A para las edades de 0 a 14 años es:

$$\text{DE Población } A_{0 \text{ a } 14 \text{ años}} = \frac{(\text{Tasa específicas de } C_{0 \text{ a } 14 \text{ años}}) \times (\text{ Población de } A_{0 \text{ a } 14 \text{ años}})}{1000}$$

$$\text{DE Población } A_{0 \text{ a } 14 \text{ años}} = \frac{3{,}0 \times 14\,000}{1000} = 42 \text{ defunciones}$$

Las defunciones esperadas (DE) en la Población B para las edades de 0 a 14 años es:

$$\text{DE Población } B_{0 \text{ a } 14 \text{ años}} = \frac{(\text{Tasa específicas de } C_{0 \text{ a } 14 \text{ años}}) \times (\text{ Población de } B_{0 \text{ a } 14 \text{ años}})}{1000}$$

$$\text{DE Población } B_{0 \text{ a } 14 \text{ años}} = \frac{3{,}0 \times 8\,000}{1000} = 24 \text{ defunciones}$$

Hasta completar este cuadro:

Cuadro 3.13

Edad	Defunciones esperadas en A	Defunciones esperadas en B
0 a 14	42	24
15 a 24	20	14
25 a 34	12	20
35 a 44	20	50
45 a 54	30	60
55 a 64	34	85
65 o más	25	100
TOTAL	183	353

Si las tasas de mortalidad fueran iguales a las de la Población C, las defunciones esperadas (DE) serían 183 en A y 353 en B.

3.5.3. LA RAZÓN DE ESTANDARIZACIÓN

La razón de estandarización se calcula dividiendo las defunciones ocurridas (DO) entre las defunciones esperadas (DE) de cada población.

$$\text{Razón de estandarización de A} = \frac{\text{D O}}{\text{D E}} = \frac{300}{183} = 1{,}64$$

$$\text{Razón de estandarización de B} = \frac{\text{D O}}{\text{D E}} = \frac{400}{353} = 1{,}13$$

Cuadro 3.14

	Defunciones observadas (DO)	Defunciones esperadas (DE)	$\frac{\text{D O}}{\text{D E}}$
Población A	300	183	1,64
Población B	400	353	1,13

Estas razones de estandarización señalan que, en condiciones de igualdad, la mortalidad es mayor en A que en B (1,64 > 1,13), por lo que se debe proceder a calcular las tasas ajustadas para que las comparaciones sean válidas.

Recuerde que la razón de estandarización *no* mide el riesgo de morir, se calcula *solamente* para comparar tasas provenientes de estructuras diferentes (por edades, composición, acceso a servicios, situación económica).

Para obtener las tasas ajustadas por el método indirecto, se multiplica la tasa bruta tipo (la de la Población C) por la razón de estandarización de cada población (A y B).

$$\text{Tasa Ajustada de A} = (\text{Tasa Bruta de C}) \times (\text{Razón de Estandarización de A})$$

$$\text{Tasa Ajustada de A} = 5{,}7 \times 1{,}64 = 9{,}34$$

$$\text{Tasa Ajustada de B} = (\text{Tasa Bruta de C}) \times (\text{Razón de Estandarización de B})$$

$$\text{Tasa Ajustada de B} = 5{,}7 \times 1{,}13 = 6{,}46$$

Cuadro 3.15

	Tasa bruta observada	Tasa bruta ajustada
Población A	7,50	9,34
Población B	8,00	6,46

En el Cuadro 3.15 puede verse que la mortalidad ajustada es mayor en la Población A. Los datos iniciales hacían creer lo contrario, pero al estandarizar el efecto de las diferencias de estructura por edades se evidencia una mayor mortalidad en A que en B (9,34 > 6,46).

Las tasas ajustadas se calculan con el fin de *comparar* la frecuencia con que ocurre un fenómeno en varias poblaciones con diferencias (de edad, geografía, tiempo o condición socioeconómica). Sirven también para comparar resultados de diferentes periodos, pero *no* son medidas reales sino tasas hipotéticas.

Para el análisis de los fenómenos vitales es mejor la utilización de tasas específicas, las tasas ajustadas solamente comparan tasas totales usando un solo indicador.

EJERCICIO 3.9

a) Dos poblaciones cuentan con la siguiente estructura de población por edad:

Edad	Población X	Población Y
0 a 14	4 000	7 000
15 a 24	3 500	5 000
25 a 34	5 000	3 000
35 a 44	5 000	2 000
45 a 54	3 000	1 500
55 a 64	2 500	1 000
65 o más	2 000	500
TOTAL	25 000	20 000

En un periodo se han registrado 301 muertes en la Población X y 156 muertes en la Población Y. Se requiere comparar la mortalidad de estas poblaciones y no se dispone de datos de las tasas específicas por edad, razón por la cual se sugiere ajustar las tasas utilizando la estructura de edades de la Población C:

Edad	Población C	Defunciones	Tasas específicas
0 a 14	100 000	300	3,0
15 a 24	90 000	180	2,0
25 a 34	70 000	140	2,0
35 a 44	50 000	250	5,0
45 a 54	40 000	400	10,0
55 a 64	30 000	510	17,0
65 o más	20 000	500	25,0
TOTAL	400 000	2 280	5,700

Encuentre:

1. Las tasas brutas de mortalidad de X y de Y.

ii. Las defunciones esperadas utilizando las tasas específicas de la Población C.

	DEFUNCIONES ESPERADAS	
Edad	En X	En Y
0 a 14		
15 a 24		
25 a 34		
35 a 44		
45 a 54		
55 a 64		
65 o más		
Total		

iii. La razón de estandarización de las dos poblaciones a comparar.

iv. Ajuste las tasas de mortalidad de las dos poblaciones, elabore un cuadro comparativo e interprete los resultados.

b) Una investigación sobre mortalidad por causa informó que, de 710 muertes por enfermedad hipertensiva, 343 casos fueron de hombres y 367 de mujeres. Con esos datos se generaron tasas de mortalidad de 14,66 por cien mil habitantes para los hombres y 16,13 por cien mil habitantes para las mujeres.

Sexo	DO	Tasa
Masculino	343	14,66
Femenino	367	16,13

Se quiere saber si estas tasas están afectadas por la estructura de edad de hombres y mujeres, para ello se decide ajustarlas utilizando la información siguiente:

Edad	Tasas de mortalidad por 100 000	Hombres	Mujeres	Población	Defunciones esperadas de hombres	Defunciones esperadas de mujeres
0 a 9	129,1	402 730	385 058	787 788		
10 a 19	43,1	418 028	403 054	821 082		
20 a 29	94,0	318 086	320 394	638 480		
30 a 39	121,1	287 515	298 013	585 528		
40 a 49	249,9	209 176	215 715	424 891		
50 a 59	534,0	123 246	127 690	250 936		
60 a 69	1 357,4	77 644	82 148	159 792		
70 y más	5 429,8	66 189	75 493	141 682		
Total	392,2	1 902 614	1 907 565	3 810 179		

Encuentre:

i. Las defunciones esperadas de hombres y de mujeres.

ii. Calcule la razón de estandarización para hombres y para mujeres.

	Defunciones observadas (DO)	Defunciones esperadas (DE)	$\frac{DO}{DE}$
Hombres	343		
Mujeres	367		

iii. Calcule e interprete las tasas ajustadas.

	Tasa bruta observada	Tasa bruta ajustada
Hombres	14,66	
Mujeres	16,13	

3.5.4. LA RAZÓN DE ESTANDARIZACIÓN COMO INDICADOR DE SALUD

Se ha visto que la razón de estandarización resulta de la división de los casos observados entre los casos esperados, esto es, de la división de eventos reales entre eventos hipotéticos. El supuesto básico de los eventos hipotéticos es que sus riesgos son equivalentes a los de una población tipo, de la cual se utilizan sus tasas específicas para hacer el ajuste y poder comparar poblaciones heterogéneas.

La razón de estandarización (generalmente amplificada por 100) se utiliza frecuentemente como un índice de comparación directo. Por ejemplo, la Caja Costarricense del Seguro Social publica las siguientes razones de estandarización por 100 defunciones por cáncer de útero, reproducida por Moya (2009, p. 189):

Cuadro 3.16

Razones de estandarización por 100 defunciones por cáncer de útero Según provincia de residencia. Costa Rica: 1956-2003

Quinquenio	Costa Rica	San José	Alajuela	Cartago	Heredia	Guanacaste	Puntarenas	Limón
1956 a 1960	100	114	96	98	129	40	95	82
1961 a 1964	100	95	88	131	10	75	109	138
1965 a 1969	100	116	83	79	105	66	122	100
1070 a 1974	100	101	87	77	105	79	128	159
1975 a 1979	100	95	98	86	113	100	113	135
1980 a 1984	100	94	104	111	76	68	143	136
1985 a 1989	100	96	81	81	89	123	131	163
1990 a 1994	100	91	77	95	73	164	122	176
1995 a 1999	100	95	92	86	88	123	111	157
2000 a 2003	100	92	88	96	81	114	120	184

De la observación de las razones por provincia, se colige que entre 1956 y 1960 los niveles de riesgo de cáncer del útero eran mayores en Heredia (129) y San José (114); pero entre el año 2000 y 2003 este riesgo es mayor en Limón (184), Puntarenas (120) y Guanacaste (114).



EJERCICIOS DE REPASO Y AUTOEVALUACIÓN

Esta sección pretende que haga un repaso general de los contenidos del capítulo.

EJERCICIO 3.10

Encierre en un círculo la opción que se ajusta al enunciado, Falsa (F) o Verdadera (V).

F V 1. Los indicadores son cifras que se construyen a partir del registro de los hechos vitales y su posterior procesamiento para informar sobre la situación sanitaria de una sociedad y el desempeño del sistema de atención.

F V 2. Los números relativos permiten transformar cifras grandes en cifras manejables, lo cual facilita su uso, permite el establecimiento de relaciones significativas, aumenta la utilidad de los datos y favorece el análisis de las cifras.

F V 3. Una proporción explica la relación estadística que resulta de la división de la población entre un subgrupo.

F V 4. Una razón es el dato que resulta de dividir un subgrupo de la población entre otro subgrupo o entre la población total.

F V 5. Las tasas brutas se calculan dividiendo el número de elementos que, en el periodo, experimentó las consecuencias del riesgo entre la población expuesta a sufrir el evento.

F V 6. Las pruebas de tamizaje son exámenes que solamente se realizan a bebés en su primera semana de vida.

F V 7. La sensibilidad de una prueba de tamizaje la informa el porcentaje de pacientes sanos que resulta negativo.

F V 8. La especificidad de una prueba de tamizaje se obtiene por la proporción de pacientes enfermos que se identifican inicialmente.

F V 9. El valor predictivo de una prueba positiva es el número de pacientes realmente sano dividido entre todos los libres de sospecha en el diagnóstico inicial.

F V 10. El valor predictivo de una prueba negativa es el número de pacientes que está realmente enfermo dividido entre los sospechosos de padecer la enfermedad.

F V 11. Los informes de morbilidad de una población proporcionan indicadores positivos del estado de salud de la colectividad.

F V 12. Los indicadores de morbilidad siempre se construyen con categorías mutuamente excluyentes y colectivamente exhaustivas.

F V 13. La tasa de incidencia de una enfermedad es el número de casos nuevos en un periodo por cada 100,000 habitantes en condiciones de sufrir el riesgo.

F V 14. La tasa de prevalencia informa de todos los casos de enfermedad que no han terminado en la muerte, la cura o la remisión, además de los casos nuevos reportados en el periodo de estudio.

F V 15. La tasa de incidencia viene a ser como una fotografía que muestra los casos existentes en un momento determinado.

F V 16. La tasa de casos es un tipo particular de tasa de incidencia, es utilizada cuando existe la posibilidad de que algunas personas contraigan la misma enfermedad más de una vez en el periodo de estudio.

F V 17. La oportunidad relativa (OR) de sobrevivir utilizando cirugía es la razón de los que sanan entre los que mueren habiendo sido operados.

F V 18. La mayor ventaja de utilizar tasas ajustadas es que sus índices no variarán aunque cambien las poblaciones tipo.

F V 19. En el método directo de ajuste de tasas se utiliza una población tipo, mientras que en el método indirecto se usan tasas específicas tipo.

F V 20. La razón de estandarización resulta de la división de los casos observados entre los casos esperados, esto es, de la división de eventos reales entre eventos hipotéticos.

EJERCICIO 3.11

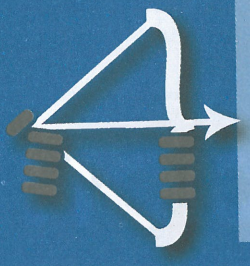
Pareo. Llene los paréntesis con el número que acompaña a cada definición de la izquierda, de forma que cada definición coincida con su concepto. Quedan dos paréntesis en blanco.

1.	Medida que indica la proporción de pacientes sanos que resulta negativo en la prueba de tamizaje.	( )	Morbilidad
2.	Condición o rasgo detectable que se asocia con el aumento en la probabilidad de padecer una enfermedad, desarrollar un proceso mórbido, sufrir un accidente o enfrentar la muerte prematuramente.	( )	Proporción
3.	Ocurrencia de enfermedades en una población, su registro proporciona indicadores negativos del estado de salud de la colectividad.	( )	Tasa de letalidad
4.	Es la razón del éxito terapéutico entre los dos tratamientos.	( )	Constante de amplificación
5.	Indicador que se calcula dividiendo las defunciones ocurridas entre las defunciones esperadas de cada población, esto es, de la división de eventos reales entre eventos hipotéticos.	( )	Factor de riesgo
6.	Medida que indica la proporción de pacientes con un padecimiento que se identifica por medio de una prueba de tamizaje.	( )	Fracción atribuible
7.	Número relativo que mide la probabilidad de morir por una enfermedad una vez que se ha contraído.	( )	Oportunidad relativa
8.	Razón que estima el grado de influencia que tiene un tratamiento como causa de la curación de los pacientes y que puede ser atribuible al hecho de seguirlo.	( )	Especificidad de la prueba
9.	Valor arbitrario por el cual se multiplican los indicadores para hacerlos más comprensibles para el gran público.	( )	Sensibilidad de la prueba
10.	Número relativo que informa del resultado de dividir un subgrupo entre toda la población.	( )	Tasas brutas
		( )	Razón de estandarización
		( )	Valor predictivo de la prueba

CAPÍTULO 4

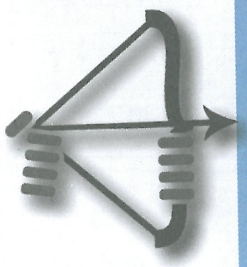
MEDIDAS DE POSICIÓN Y DE VARIABILIDAD

OBJETIVO GENERAL



Utilizar las técnicas adecuadas para el cálculo de medidas de posición, forma y distribución, como insumos que mejoran la evaluación y la práctica de la salud.

OBJETIVOS ESPECÍFICOS



Al finalizar el estudio de este capítulo usted estará en la capacidad de:

1. Explicar el significado de las principales medidas de posición, forma y variabilidad.
2. Calcular e interpretar las principales medidas de posición y de dispersión para datos no agrupados.
3. Realizar cálculos de las medidas de posición y de dispersión para datos agrupados e interpretarlos.
4. Seleccionar las medidas de posición y de dispersión que expliquen en mejor forma una serie de datos.
5. Calcular el intervalo intercuartil y utilizarlo para definir límites de tolerancia.
6. Indicar las alertas sanitarias a partir de la observación de registros del índice endémico.

En este capítulo se estudian los métodos de cálculo e interpretación de las principales medidas de posición, forma y variabilidad, así como la construcción de intervalos de control, límites de tolerancia, tanto para distribuciones simétricas como asimétricas, y la interpretación del índice endémico en una comunidad..

Determina primero los hechos,
después puedes tergiversarlos como te plazca.

MARK TWAIN

En los capítulos previos se ha trabajado la recolección de datos, la distribución de frecuencias y el cálculo de indicadores estadísticos en el campo de la salud. En este capítulo corresponde abordar la elaboración de medidas descriptivas sobre la tendencia de los datos, la forma de la distribución y las variaciones alrededor de cifras de posición.

El análisis y la interpretación de los datos se facilita con la obtención de medidas que resumen las principales características de una distribución como:

- La posición
- La forma
- La variabilidad

Lo que se pretende es calcular valores representativos de todos los datos, que expliquen el comportamiento general de la distribución, además de describir esta distribución y compararla para llegar a mejores conclusiones. De estas medidas, las más comunes son el promedio y el rango, que ya se utilizaron para construir la distribución de frecuencias.

4.1. MEDIDAS DE POSICIÓN

Las *medidas de posición* también se conocen como medidas de tendencia central y son de gran utilidad para comprender fenómenos del campo de la salud. Entre ellas están:

- La moda
- El promedio o media aritmética
- La mediana
- Los percentiles

Gracias al avance de la informática, se puede procesar un gran volumen de datos y obtener estas medidas con mucha precisión. Sin embargo, los informes ya pueden estar organizados en una distribución de frecuencias, por lo que es necesario conocer la técnica para calcular estos valores desde cifras agrupadas. Para comenzar, se trabajará con pocos datos, luego con datos agrupados y en el Anexo 2 se presentará una forma práctica de cálculo utilizando herramientas de *Excel*.

4.1.1. LA MODA (Mo)

La *moda* es la cantidad o cualidad que más se repite en un conjunto de datos, técnicamente se define como el valor más frecuente. Si un estudiante obtuvo las siguientes calificaciones en las once materias que cursó:

$$70,\ 75,\ 75,\ 80,\ 85,\ 85,\ \underbrace{90,\ 90,\ 90,\ 90}_{\text{valor más frecuente}},\ 100$$

La moda de sus calificaciones es 90, puesto que es el valor que más se repite.

Así como puede calcularse la moda para cantidades, también puede definirse para cualidades. En el siguiente informe:

CLASE DE PARTO
Vaginal
Cesaria
Episiotomía
Vaginal
Vaginal
Cesaria
Episiotomía
Cesaria
Vaginal
Cesaria
Vaginal
Cesaria
Vaginal
Vaginal
Vaginal

La moda es el parto vaginal por contar con el mayor número de observaciones: 8. Otros ejemplos de moda son: el salario más frecuente de los asegurados, el analgésico que más se prescribe, la enfermedad que genera el mayor número de consultas, el periodo de hospitalización más común.

Además de poder usarse tanto para datos cuantitativos como cualitativos y su facilidad de cálculo e interpretación, la moda tiene la ventaja de que no se ve afectada por la presencia de valores extremos, sean muy altos o muy bajos. Como desventajas se señala que en algunos casos puede no existir, en otros puede darse la presencia de varias modas y siempre sus aplicaciones matemáticas son limitadas.

4.1.2. EL PROMEDIO ($\bar{x}$)

El *promedio* o media aritmética es la medida de tendencia central más utilizada y resulta de dividir la suma de todos los valores entre el número de ellos. De las once materias cursadas el promedio simple es:

$$\bar{x} = \frac{\text{Suma de todos los valores}}{\text{Número de valores}}$$

$$\bar{x} = \frac{70 + 75 + 75 + 80 + 85 + 85 + 90 + 90 + 90 + 90 + 100}{11}$$

$$\bar{x} = \frac{930}{11} = 84{,}54$$

En general, el promedio o media aritmética de una muestra se define como:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

El uso del promedio tiene las siguientes ventajas:

- Siempre se puede calcular para un conjunto de datos.
- Existe un único promedio para un conjunto de datos.
- El promedio se ve afectado por todos y cada uno de los valores del conjunto.
- Si se multiplica el promedio por el número de observaciones, se obtiene la sumatoria de las observaciones.
- Si se resta el promedio a cada una de las observaciones, y luego se suman sus resultados, la sumatoria siempre será cero.
- Si se suma (resta) una constante a cada valor del conjunto, el promedio aumentará (disminuirá) en la misma cantidad.
- Si se multiplica (o divide) una constante a cada valor del conjunto, el promedio se multiplicará (o dividirá) por esa constante.

La mayor desventaja del promedio es que va a ser influido por la presencia de valores extremos.

El cálculo del promedio simple partió del supuesto de que cada materia tiene igual importancia, pero este no siempre es el caso. En el mismo ejemplo de las once calificaciones, agréguese la información de que las primeras diez materias corresponden a cursos teóricos de tres créditos cada uno y la última materia es un seminario teórico-práctico de cinco créditos, para un total de 35 créditos cursados. Ahora el promedio por calcular deberá ser ponderado, este se obtiene sumando las multiplicaciones de cada calificación por sus créditos y luego dividiendo esa suma entre todos los créditos matriculados (35):

$$\frac{70(3) + 75(3) + 75(3) + 80(3) + 85(3) + 85(3) + 90(3) + 90(3) + 90(3) + 90(3) + 100(5)}{35}$$

$$= \frac{2\,990}{35} = 85{,}42$$

La fórmula general para calcular el promedio ponderado es:

$$\bar{x}_p = \frac{\sum_{i=1}^{n} x_i f_i}{\sum_{i=1}^{n} f_i}$$

4.1.3. LA MEDIANA (Me)

La *mediana* es el valor que divide en dos grupos iguales a una serie ordenada, la mitad de las observaciones tendrá valores menores que la mediana y la otra mitad tendrá valores mayores. Técnicamente es *el valor central de un conjunto ordenado de datos*, es decir, aquel que ocupa la posición $\frac{n+1}{2}$. En el ejemplo de las once calificaciones, será la $\frac{11+1}{2} = 6$ (sexta) posición, que es de 85.

$$70,\ 75,\ 75,\ 80,\ 85,\ \overbrace{85}^{\text{sexta posición}},\ 90,\ 90,\ 90,\ 90,\ 100$$

En el caso de que n sea par, el valor de la mediana es el promedio de los dos valores centrales, pero siempre el 50 % de las observaciones será menor y el 50 % será mayor.

EJERCICIO 4.1

a) En la farmacia de un dispensario, se toma una muestra de los analgésicos despachados durante la mañana de un lunes. Los resultados se presentan seguidamente y se le pide que indique cuál es la moda en ese informe.

Paracetamol (Tyenol)	Paracetamol (Tyenol)
Ácido acetilsalicílico (Aspirina)	Paracetamol (Tyenol)
Ácido acetilsalicílico (Aspirina)	Ácido acetilsalicílico (Aspirina)
Naproxeno (Aleve)	Naproxeno (Aleve)
Ibuprofeno (Motrin)	Ácido acetilsalicílico (Aspirina)
Paracetamol (Tyenol)	Ibuprofeno (Motrin)
Ibuprofeno (Motrin)	Naproxeno (Aleve)
Naproxeno (Aleve)	Ibuprofeno (Motrin)

b) Los registros médicos de una muestra de 100 pacientes informan que para el tratamiento de rinitis se les ha prescrito, con regularidad, uno de los siguientes antihistamínicos:

ANTIHISTAMÍNICO	Pacientes que lo usan
Bromfeniramina (Dimetapp)	5
Cetrizina (Zyrtec)	4
Clorfeniramina (Chlor-Trimetron)	24
Difenhidramina (Benadryl)	17
Fexofenadina (Allegra)	21
Loratadina (Claritin)	24
Meclastina (Tavist)	5
	100

En la prescripción de antihistamínicos, ¿cuál es la moda?

c) Entre los años 2002 y 2009 se registraron todos los casos de rubéola reportados mes a mes en una jurisdicción sanitaria. Dada la entrada de la época lluviosa, interesan los datos del mes de mayo, del cual se sabe lo siguiente:

AÑO	Casos en Mayo
2002	16
2003	4
2005	7
2006	19
2007	5
2008	6
2009	6

Calcule el promedio, la moda y la mediana de los casos de rubéola del mes de mayo.

d) Se tienen las siguientes observaciones sobre el contenido de glucosa en la sangre de un paciente en ayunas:

Glucosa -mg/dl-
75
100
92
100
76
80
70
75
82
100

Calcule el promedio, la moda y la mediana del contenido de glucosa en la sangre de este paciente.

4.2. CÁLCULO DE LAS MEDIDAS DE POSICIÓN PARA DATOS AGRUPADOS

Muchas veces los datos con que se hará un estudio, ya se presentan agrupados en distribuciones de frecuencias y hay que trabajarlos aplicando algunas variantes de los métodos básicos.

4.2.1. LA MODA (Mo)

Para calcular la moda de datos agrupados, primero se busca la clase que presenta la mayor frecuencia o "clase modal", luego se identifica el límite inferior real de esta clase y se suma el cociente de las diferencias de la frecuencia de la clase modal con las frecuencias de la clase anterior y posterior multiplicada por el ancho de clase, así:

$$Mo = L_i + \left(\frac{d_1}{d_1 + d_2}\right) c$$

Donde:

L_i = Límite inferior real de la clase modal

d_1 = Diferencia entre la frecuencia de la clase modal y la frecuencia de la clase anterior

d_2 = Diferencia entre la frecuencia de la clase modal y la frecuencia de la clase posterior

c = Intervalo de clase.

Por ejemplo, se tiene la distribución de frecuencia de los pesos de una muestra de 85 alumnos matriculados en la carrera de Registros Médicos:

CLASES Peso en libras	Punto medio x_i	Frecuencia f_i
85,5 a 100,5	93	1
100,5 a 115,5	108	3
115,5 a 130,5	123	20
130,5 a 145,5	138	21
145,5 a 160,5	153	17
160,5 a 175,5	168	10
175,5 a 190,5	183	8
190,5 a 205,5	198	5
		85

En esta distribución se observa que la clase modal va de 130,5 a 145,5 libras, con frecuencia de 21 observaciones, por lo que la moda se calcula así:

$$Mo = 130{,}5 + \left(\frac{1}{1+4}\right) 15$$

$$Mo = 130{,}5 + 3 = 133{,}5 \text{ libras}$$

4.2.2. EL PROMEDIO ($\bar{x}$)

Para el cálculo del promedio de una serie de datos agrupados, primero se multiplican los puntos medios de clase por sus respectivas frecuencias, después se suman y este total se divide entre la sumatoria de las frecuencias, así:

CLASES Peso en libras	Punto medio x_i	Frecuencia f_i	Multiplicación $x_i f_a$
85,5 a 100,5	93	1	93
100,5 a 115,5	108	3	324
115,5 a 130,5	123	20	2 460
130,5 a 145,5	138	21	2 898
145,5 a 160,5	153	17	2 601
160,5 a 175,5	168	10	1 680
175,5 a 190,5	183	8	1 464
190,5 a 205,5	198	5	990
		85	12 510

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i f_i}{\sum_{i=1}^{n} f_i} = \frac{12\,510}{85} = 147{,}18 \text{ libras}$$

Por el hecho de utilizar los puntos medios de cada clase, la media aritmética para datos agrupados puede diferir ligeramente del resultado considerando todos los datos.

4.2.3. LA MEDIANA (Me)

Para calcular la mediana en datos agrupados, primero se busca la clase en donde está el valor $\frac{n}{2}$, luego a su límite inferior se suma el resultado de la interpolación lineal de las frecuencias por el ancho de clase, así:

$$Me = L_i + \left(\frac{\left(\frac{n}{2}\right) - F_{aa}}{f_i} \times c \right)$$

Donde:

L_i = Límite inferior real de la clase en donde está la mediana

n = Total de las observaciones

F_{aa} = Frecuencia acumulada "X o menos" de la clase anterior a la clase mediana

f_i =Frecuencia absoluta de la clase en donde está la mediana

c = Ancho de clase

Siguiendo con la distribución de frecuencias de los pesos de 85 alumnos de la carrera de Registros Médicos:

CLASES Peso en libras	Punto medio x_i	Frecuencia f_i	Acumulada "X o menos" f_a
85,5 a 100,5	93	1	1
100,5 a 115,5	108	3	4
115,5 a 130,5	123	20	24
130,5 a 145,5	138	21	45
145,5 a 160,5	153	17	62
160,5 a 175,5	168	10	72
175,5 a 190,5	183	8	80
190,5 a 205,5	198	5	85
		85	

El valor $\frac{n}{2} = 42,5$ se ubica en la clase con el límite inferior de 130,5, la frecuencia absoluta de la clase en donde está la mediana es 21, la frecuencia acumulada "X o menos" de la clase anterior es 24 y el ancho de clase es 15:

$$Me = 130,5 + \left(\frac{42,5 - 24}{21} \times 15 \right)$$

$$Me = 130,5 + 13,2 = 143,7 \text{ libras}$$

Lo cual informa que la mitad de los estudiantes matriculados en Registros Médicos pesa menos de 143,7 libras y la otra mitad pesa más de 143,7 libras.

EJERCICIO 4.2

a) El cuadro siguiente muestra la distribución de frecuencias del peso de 40 recién nacidos en una clínica rural.

CLASES	Puntos medios	FRECUENCIAS ABSOLUTAS		
(Peso en gramos)	x_i	f_i	F_a	$x_i f_i$
499,5 – 999,5	749,5	1	1	749,5
999,5 – 1 499,5	1 249,5	2	3	2 499,0
1 499,5 – 1 999,5	1 749,5	2	5	3 499,0
1 999,5 – 2 499,5	2 249,5	4	9	8 998,0
2 499,5 – 2 999,5	2 749,5	5	14	13 747,5
2 999,5 – 3 499,5	3 249,5	7	21	22 746,5
3 499,5 – 3 999,5	3 749,5	10	31	37 495,0
3 999,5 – 4 499,5	4 249,5	6	37	25 497,0
4 499,5 – 4 999,5	4 749,5	3	40	14 248,5
		40		129 480,0

Con esos datos agrupados, obtenga:

i. La moda del peso de los recién nacidos.

ii. El promedio del peso de los recién nacidos.

iii. La mediana del peso de los recién nacidos.

b) Se cuenta con la siguiente distribución acerca de la edad en años cumplidos de 50 mujeres al momento de dar a luz en el Hospital de la Anexión, durante diciembre de 2014.

CLASES	Puntos medios	FRECUENCIA ABSOLUTA	FRECUENCIA ACUMULADA	
(Edades)	x_i	f_i	F_a	$x_i f_i$
10 a menos de15	12,5	2	2	25
15 a menos de 20	17,5	4	6	70
20 a menos de 25	22,5	5	11	112,5
25 a menos de 30	27,5	12	23	330
30 a menos de 35	32,5	14	37	455
35 a menos de 40	37,5	9	46	337,5
40 a menos de 45	42,5	3	49	127,5
45 a menos de 50	47,5	1	50	47,5
		50		1 505

Con esos datos agrupados, obtenga:

i. La moda de la edad de las parturientas.

ii. El promedio de edad de las parturientas.

iii. La edad mediana de las parturientas.

4.3. USO DE LAS MEDIDAS DE POSICIÓN

Ya se ha dicho que con las medidas de posición se pretende calcular valores representativos de una distribución de frecuencias, para que expliquen su comportamiento general.

Cada medida tiene sus ventajas: la moda se calcula y se comprende con mucha facilidad, el promedio es la medida más popular, en tanto que la mediana no se ve afectada por la presencia de valores extremos. En algunas situaciones, una de estas medidas puede ser la que mejor explique las características de la distribución, pero nunca se podrá decir que compiten entre sí, antes bien, se complementan y juntas advierten mejor acerca de la *forma* de la distribución de los datos.

Cuando los datos se distribuyen en forma simétrica, como en la figura siguiente:

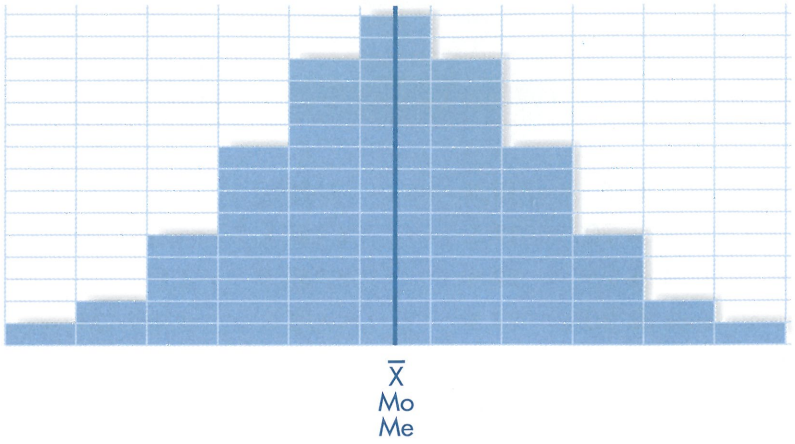


Figura 4.1. **Distribución simétrica de una serie de datos**

Entonces, el promedio, la moda y la mediana coinciden en el centro de la distribución, de manera que cualquiera de ellas puede utilizarse para explicar al grupo de datos.

Para medir la asimetría, Karl Pearson desarrolló un coeficiente útil cuando las distribuciones son unimodales, con forma de campana y moderadamente asimétricas: al promedio ($\bar{x}$) se le resta la moda (Mo) y el resultado se divide entre la desviación estándar (que siempre tendrá valor positivo y cuyo cálculo se estudiará más adelante).

$$A_p = \frac{\bar{x} - Mo}{s}$$

En una distribución simétrica, el promedio ($\bar{x}$) y la moda (Mo) son iguales, resultando cero el denominador, por lo tanto, si el coeficiente es cero la distribución se asume como *simétrica*.

Cuando el promedio es mayor que la moda ($\bar{x}$ > Mo) el coeficiente de asimetría es positivo y una cola de la distribución se prolonga hacia la derecha, entonces se dice que la distribución tiene *asimetría positiva* o hacia la derecha.

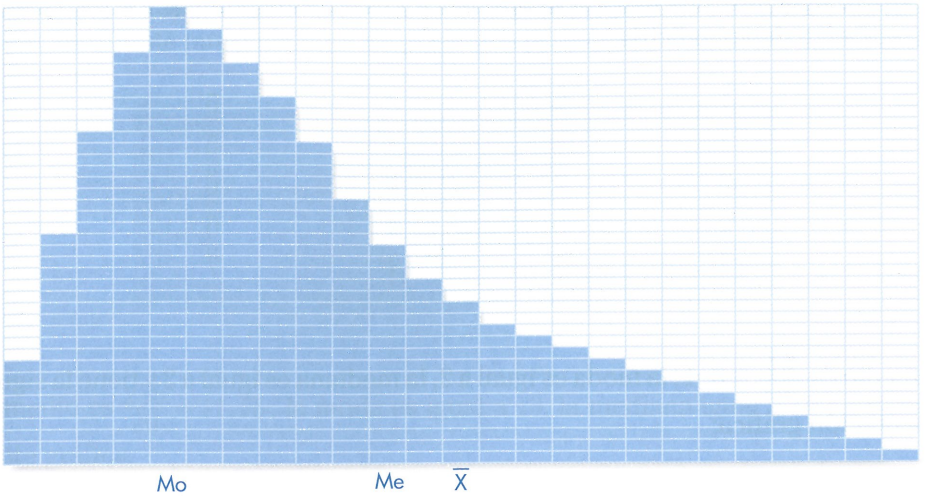


Figura 4.2. **Distribución con asimetría positiva**

Cuando el promedio es menor que la moda ($\bar{x}$ < Mo) el coeficiente de asimetría es negativo y una cola de la distribución se prolonga hacia la izquierda, entonces se dice que la distribución tiene *asimetría negativa* o hacia la izquierda.

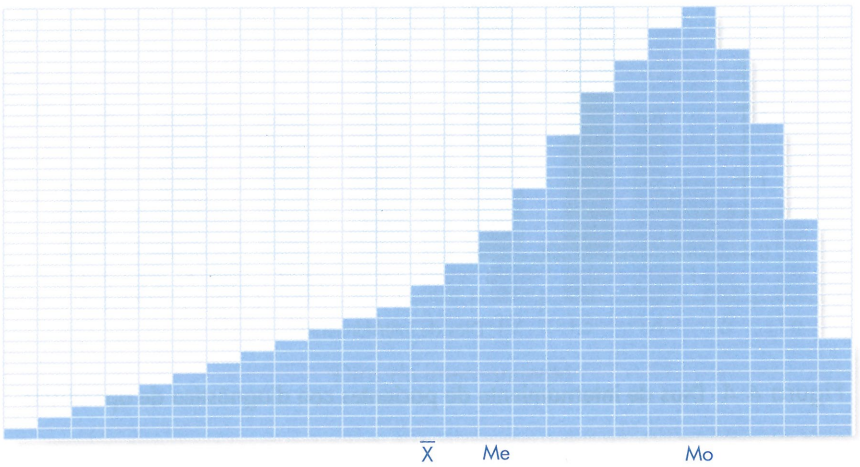


Figura 4.3. **Distribución con asimetría negativa**

Considere la siguiente información referente a los días de hospitalización para pacientes con diagnóstico de apendicitis:

1	2	3	5
2	2	3	5
2	2	4	6
2	3	4	6
2	3	4	7

El promedio es $\frac{68}{20} = 3,4$, la moda es igual a 2, la mediana es igual a 3 y tiene la siguiente representación gráfica:

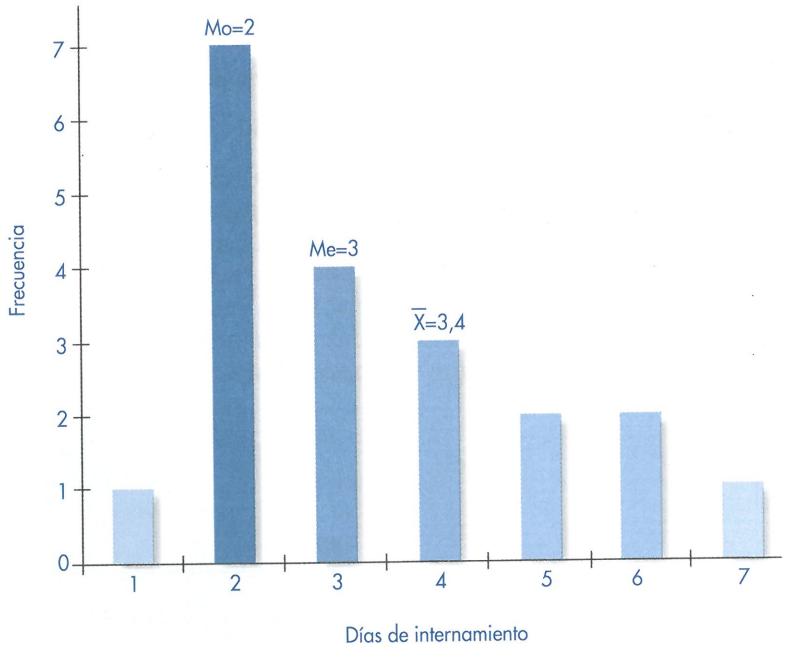


Figura 4.4. **Días de internamiento de pacientes con diagnóstico de apendicitis**

En esta distribución el promedio es mayor que la moda ($\bar{x} > \text{Mo} \rightarrow 3,4 > 2$), por lo que el coeficiente de asimetría es positivo.

Ahora observe el comportamiento en el pabellón de psiquiatría de un hospital general, en donde los días de internamiento de pacientes presentan la siguiente distribución:

7	10	12	13
8	11	13	14
9	11	13	14
9	12	13	15
10	12	13	16

El promedio es $\frac{235}{20} = 11,75$, la moda es igual a 13, la mediana igual a 12 y presenta la siguiente gráfica:

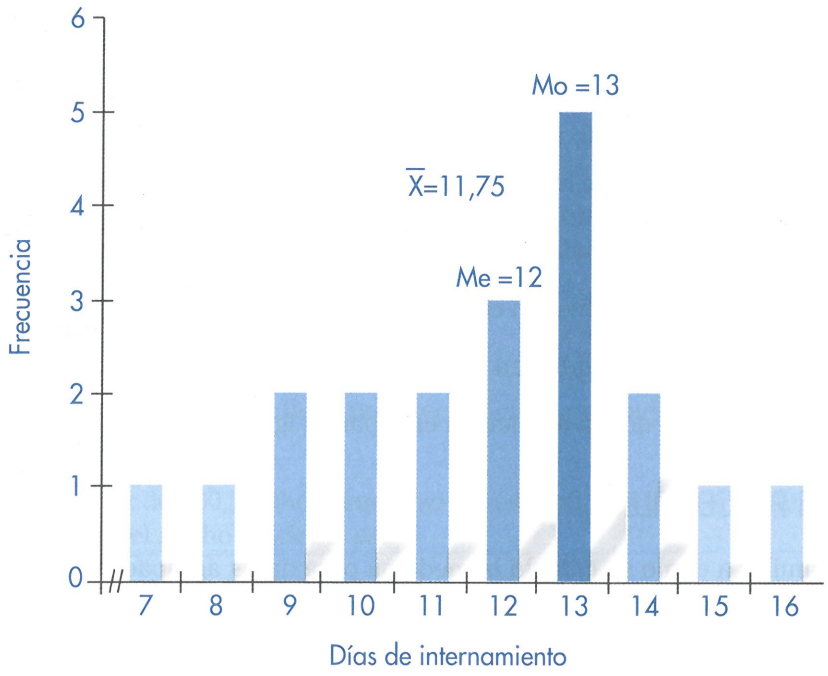


Figura 4.5. **Días de internamiento de pacientes en el pabellón de psiquiatría**

En esta distribución el promedio es menor que la moda ($\bar{x} < \text{Mo} \rightarrow 11,75 < 13$ 2), por lo que el coeficiente de asimetría es negativo.

EJERCICIO 4.3

a) En un país vecino, la moda de las edades de sus pobladores es 13 años, en tanto que la edad promedio es de 28 años. ¿Considera que la distribución por edades de esta población es simétrica o asimétrica? En caso de ser asimétrica, ¿de qué tipo es?

b) Se sabe que la moda de los pesos de los bebés de un mes de nacidos en una comunidad rural es 3713,8 gramos, la mediana es 3428,1 gramos y el peso promedio es 3237 gramos. ¿Qué tipo de asimetría tiene esta distribución?

c) Según los funcionarios actuariales de la CCSS, la moda de los salarios de los asegurados es de 301 225 colones, pero las autoridades políticas sostienen que los salarios promedios son del orden de los 600 000 colones. Ambos datos se basan en cotizaciones efectivas, por lo que su diferencia no puede ser adjudicada a evasión ni a morosidad. ¿Cómo podría explicar esta situación apoyado en la teoría estadística?

4.4. LOS CUANTILES

Otras medidas explicativas de los datos son los *cuantiles*, estos dividen la distribución en fracciones de igual magnitud. De ellos ya se ha visto la mediana, que parte la distribución en dos partes iguales. Entre los cuantiles, además de la mediana, están:

- Los cuartiles, que la dividen en cuatro
- Los deciles, que la dividen en diez
- Los percentiles, que la dividen en cien partes iguales

4.4.1. LOS PERCENTILES

De forma similar a cómo se calculó la mediana para datos agrupados, se puede variar ligeramente el procedimiento para calcular los percentiles.

Tómese la distribución de frecuencias de los pesos de 85 alumnos de la carrera de Registros Médicos:

CLASES Peso en libras	Punto medio *xi*	Frecuencia *fi*	Acumulada "X o menos" *Fa*
85,5 a 100,5	93	1	1
100,5 a 115,5	108	3	4
115,5 a 130,5	123	20	24
130,5 a 145,5	138	21	45
145,5 a 160,5	153	17	62
160,5 a 175,5	168	10	72
175,5 a 190,5	183	8	80
190,5 a 205,5	198	5	85
		85	

Si, por ejemplo, se necesita conocer el percentil 50, se multiplicará el porcentaje (%) por el número de datos (n):

$$\%n = 0,50 \times 85 = 42,5$$

Con esto se identifica la clase en donde está el percentil 50 ($P_{0,50}$) es decir, el dato en la posición 42,5.

Este dato se encuentra en la clase con un límite inferior de 130,5, una frecuencia absoluta de 21, la clase anterior tiene una frecuencia acumulada "X o menos" de 24 y el ancho de clase es igual a 15.

A ese límite inferior de 130,5 deberá sumarse el resultado de la interpolación lineal de las frecuencias por el ancho de clase, así:

$$P_{0,50} = 130,5 + \left(\frac{42,5 - 24}{21} \times 15\right)$$

$$P_{0,50} = 130,5 + 13,2 = 143,7 \quad \text{libras}$$

Ya usted notó que la mediana y el percentil 50 son lo mismo. A partir del cálculo de la mediana, con ligeras modificaciones, derivamos la fórmula de los percentiles, así:

$$P_{\%} = L_i + \left(\frac{(\%n) - F_{aa}}{f_i} \times c\right)$$

Donde:

- $P_{\%}$ = Percentil buscado
- L_i = Límite inferior real de la clase donde está el percentil buscado
- n = Total de las observaciones
- f_{aa} = Frecuencia acumulada "X o menos" de la clase anterior donde está el percentil buscado
- f_i =Frecuencia absoluta de la clase donde está el percentil
- c = Intervalo de clase

Si se busca el percentil 75, primero se identifica la posición multiplicando 0,75 por el total de los datos (n), así:

$$\%n = 0,75 \times 85 = 63,75$$

Con esto se ubica la clase donde está el percentil 75 ($P_{0,75}$), es decir, el dato en la posición 63,75.

Este dato se encuentra en la clase con un límite inferior de 160,5, una frecuencia absoluta de 10, la clase anterior tiene una frecuencia acumulada "X o menos" de 62 y el ancho de clase es igual a 15.

A ese límite inferior de 160,5 deberá sumarse el resultado de la interpolación lineal de las frecuencias por el ancho de clase, así:

$$P_{0,75} = 160,5 + \left(\frac{(0,75 \times 85) - 62}{10} \times 15\right)$$

$$P_{0,75} = 160,5 + \left(\frac{63,75 - 62}{10} \times 15\right)$$

$$P_{0,75} = 160,5 + 2,625 = 163,125 \text{ libras}$$

Este cálculo informa que el 75 % de los estudiantes matriculados en la carrera de Registros Médicos pesa menos de 163,125 libras y el 25 % tiene un peso superior a 163,125 libras.

4.4.2. LOS CUARTILES

Los cuartiles (Q_x) informan los límites que dividen la distribución en cuartas partes, donde cada una contiene el 25 % del conjunto de datos.

Así, el primer cuartil (Q_1) contiene la cuarta parte de los datos contenidos entre el valor mínimo de la distribución (85,5 libras) y el valor Q_1, que se calculará seguidamente.

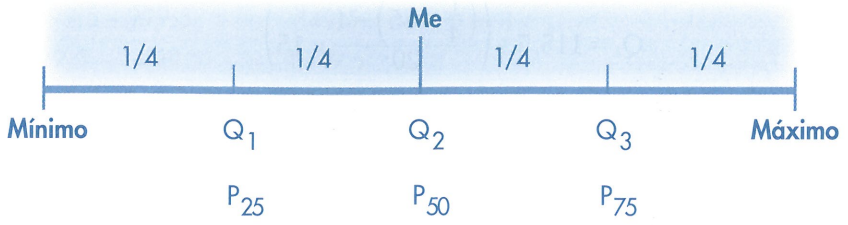


Figura 4.6. **Representación gráfica de los cuantiles: cuartiles y percentiles.**

Sabemos que los cuartiles son un tipo particular de percentiles, el cuartil 1 equivale al percentil 25 ($Q_1 = P_{0,25}$), el cuartil 2 es igual al percentil 50 y a la mediana ($Q_2 = P_{0,50}$), el cuartil 3 es igual al percentil 75 ($Q_3 = P_{0,75}$).

CLASES Peso en libras	Punto medio x_i	Frecuencia f_i	Acumulada "X o menos" f_a
85,5 a 100,5	93	1	1
100,5 a 115,5	108	3	4
115,5 a 130,5	123	20	24
130,5 a 145,5	138	21	45
145,5 a 160,5	153	17	62
160,5 a 175,5	168	10	72
175,5 a 190,5	183	8	80
190,5 a 205,5	198	5	85
		85	

Así, para el cálculo del cuartil 1 (Q_1), primero se ubicará la posición $\frac{1}{4}$ de todos los datos:

$$\frac{1}{4}\,n = \frac{1}{4} \times 85 = 21,25$$

Con esto se busca la clase donde está el límite del primer cuartil, es decir, el dato en la posición 21,725.

Este dato se encuentra en la clase con un límite inferior de 115,5, una frecuencia absoluta de 20, la clase anterior tiene una frecuencia acumulada "X o menos" de 4 y el ancho de clase es igual a 15.

A ese límite inferior de 115,5 deberá sumarse el resultado de la interpolación lineal de las frecuencias por el ancho de clase, así:

$$Q_1 = 115,5 + \left(\frac{\left(\frac{1}{4} \times 85 \right) - 4}{20} \times 15 \right)$$

$$Q_1 = 115,5 + \left(\frac{21,25 - 4}{20} \times 15 \right)$$

$$Q_1 = 115,5 + 12,94 = 128,44 \text{ libras}$$

Este resultado informa que el 25 % de los estudiantes matriculados en la carrera de Registros Médicos pesa menos de 128,44 libras y el 25 % tiene un peso superior a 128,44 libras.

EJERCICIO 4.4

a) El cuadro siguiente muestra la distribución de frecuencias del peso de 40 recién nacidos en una clínica rural.

CLASES	Puntos medios	FRECUENCIAS ABSOLUTAS		
(Peso en gramos)	x_i	f_i	F_a	$x_i f_i$
499,5 – 999,5	749,5	1	1	749,5
999,5 – 1 499,5	1249,5	2	3	2 499,0
1 499,5 – 1 999,5	1749,5	2	5	3 499,0
1 999,5 – 2 499,5	2249,5	4	9	8 998,0
2 499,5 – 2 999,5	2749,5	5	14	13 747,5
2 999,5 – 3 499,5	3249,5	7	21	22 746,5
3 499,5 – 3 999,5	3749,5	10	31	37 495,0
3 999,5 – 4 499,5	4249,5	6	37	25 497,0
4 499,5 – 4 999,5	4749,5	3	40	14 248,5
		40		129 480,0

Con esos datos agrupados, obtenga:

i. El percentil 75 $(P_{0,75})$.

ii. El primer cuartil (Q_1).

b) Se cuenta con la siguiente distribución acerca de la edad en años cumplidos de 50 mujeres al momento de dar a luz en el Hospital de la Anexión, durante diciembre de 2014.

CLASES	Puntos medios	FRECUENCIA ABSOLUTA	FRECUENCIA ACUMULADA	
(Edades)	x_i	f_i	F_a	$x_i f_i$
10 a menos de15	12,5	2	2	25
15 a menos de 20	17,5	4	6	70
20 a menos de 25	22,5	5	11	112,5
25 a menos de 30	27,5	12	23	330
30 a menos de 35	32,5	14	37	455
35 a menos de 40	37,5	9	46	337,5
40 a menos de 45	42,5	3	49	127,5
45 a menos de 50	47,5	1	50	47,5
		50		1 505

Con esos datos agrupados, obtenga:

i. El percentil 20 $(P_{0,20})$.

ii. El percentil 80 $(P_{0,80})$.

4.4.3. LOS PERCENTILES COMO MEDIDAS DE POSICIÓN

Cuando las frecuencias se presentan acumuladas en el sentido "X o menos" y las variables son continuas, también se dice que las frecuencias están acumuladas "al límite superior", esto porque interesa la proporción de la distribución que está por debajo o es menor que un valor límite.

El valor que se señala es el límite superior de cada clase o categoría y sus frecuencias relativas se consideran estimaciones de probabilidad de que se presenten valores menores o iguales que ese límite. En el ejemplo de los estudiantes de la carrera de Registros Médicos:

- Se estima que el 25 % de los estudiantes pesa 128,44 libras o menos. Donde 128,44 libras es el límite superior.
- El 50 % de los estudiantes pesa 143,7 libras (límite superior) o menos.
- El 75 % pesa 163,125 libras (límite superior) o menos.
- El 100 % pesa 205,5 libras o menos.

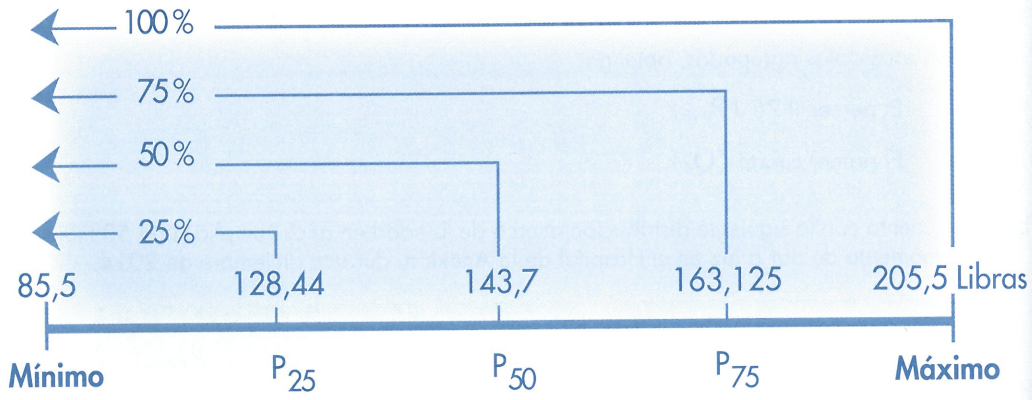


Figura 4.7. **Percentiles utilizados como medidas de posición o límites superiores de una acumulación de frecuencias**

En el caso de variables continuas con un número grande de observaciones, se pueden establecer intervalos muy pequeños y calcular el porcentaje de la distribución que se encuentra por debajo del límite superior. Si se corta la distribución en 100 segmentos, cada valor posicional será llamado *percentil*.

4.4.4. RANGO DEL PERCENTIL

En otras ocasiones se tiene un valor y se quiere saber cuál es su posición en la distribución de frecuencias, es decir, se busca el porcentaje que está por debajo de ese dato. En estos casos lo que se busca es el rango del percentil (r).

Llámese $P_{\%} = X$ que es el valor conocido y del cuál se quiere saber el rango en la distribución. Sustitúyase $\% = \frac{r}{100}$, con lo que se obtiene:

$$X = L_i + \left(\frac{\left(\frac{r}{100} n\right) - F_{aa}}{f_i} \times c \right)$$

Al despejar el valor de r (rango del percentil) se llega a la siguiente expresión:

$$r = \left(\frac{(X - L_i) \times f_i}{c} + F_{aa} \right) \times \frac{100}{n}$$

Se tiene la distribución de frecuencia de los pesos de 85 estudiantes de la carrera de Registros Médicos:

CLASES Peso en libras	Punto medio x_i	Frecuencia f_i	Acumulada "X o menos" F_a
85,5 a 100,5	93	1	1
100,5 a 115,5	108	3	4
115,5 a 130,5	123	20	24
130,5 a 145,5	138	21	45
145,5 a 160,5	153	17	62
160,5 a 175,5	168	10	72
175,5 a 190,5	183	8	80
190,5 a 205,5	198	5	85
		85	

Se quiere saber cuál es el rango del percentil de un estudiante que pesa 163,125 libras.

Puede verse en la tabla que el valor 163,125 libras se encuentra en la sexta clase.

$X = 163,125$ $L_i = 160,5$ $f_i = 10$

$F_{aa} = 62$ $c = 15$ $n = 85$

Sustituyendo en:

$$r = \left(\frac{(X - L_i) \times f_i}{c} + F_{aa}\right) \times \frac{100}{n}$$

$$r = \left(\frac{(163,125 - 160,5) \times 10}{15} + 62\right) \times \frac{100}{85}$$

$$r = \left(\frac{(2,625) \times 10}{15} + 62\right) \times \frac{100}{85}$$

$$r = (1,75 + 62) \times \frac{100}{85}$$

$$r = 63,75 \times \frac{100}{85}$$

$$r = 75$$

Un estudiante que pesa 163,125 libras se encuentra en el percentil 75 ($P_{0,75}$), es decir, el 75% de sus compañeros pesa menos que él.

EJERCICIO 4.5

a) El cuadro siguiente muestra la distribución de frecuencias del peso de 40 recién nacidos en una clínica rural.

CLASES	Puntos medios	FRECUENCIAS ABSOLUTAS		
(Peso en gramos)	x_i	f_i	F_a	$x_i f_i$
499,5 – 999,5	749,5	1	1	749,5
999,5 – 1 499,5	1 249,5	2	3	2 499,0
1 499,5 – 1 999,5	1 749,5	2	5	3 499,0
1 999,5 – 2 499,5	2 249,5	4	9	8 998,0
2 499,5 – 2 999,5	2 749,5	5	14	13 747,5
2 999,5 – 3 499,5	3 249,5	7	21	22 746,5
3 499,5 – 3 999,5	3 749,5	10	31	37 495,0
3 999,5 – 4 499,5	4 249,5	6	37	25 497,0
4 499,5 – 4 999,5	4 749,5	3	40	14 248,5
		40		129 480,0

Calcule:

i. El rango del percentil que ocupa un bebé que pesó 3 949,5 gramos al nacer.

ii. El rango del percentil que ocupa un bebé que pesa 2 599,5 gramos al nacer.

b) Se cuenta con la siguiente distribución acerca de la edad en años cumplidos de 50 mujeres al momento de dar a luz en el Hospital de la Anexión, durante diciembre de 2014.

CLASES	Puntos medios	FRECUENCIA ABSOLUTA	FRECUENCIA ACUMULADA	
(Edades)	x_i	fi	Fa	$x_i f_i$
10 a menos de15	12,5	2	2	25
15 a menos de 20	17,5	4	6	70
20 a menos de 25	22,5	5	11	112,5
25 a menos de 30	27,5	12	23	330
30 a menos de 35	32,5	14	37	455
35 a menos de 40	37,5	9	46	337,5
40 a menos de 45	42,5	3	49	127,5
45 a menos de 50	47,5	1	50	47,5
		50		1 505

Obtenga:

i. El rango percentil que ocupa una mujer que tiene 24 años al momento de dar a luz.

ii. El rango percentil que ocupa una mujer que tiene 40 años al momento de dar a luz.

4.5. LA VARIABILIDAD

La variabilidad es un concepto fundamental del análisis estadístico, puesto que los hechos se repiten pero lo hacen con alguna variación. Las medidas de posición brindan información útil, pero no es suficiente para hacer inferencias válidas.

Saber que los dos hijos de la familia Pérez festejarán un promedio de 10 años, no es suficiente para inferir que son gemelos o que el mayor cumplirá 19 años y el menor apenas un año. Si son gemelos la variabilidad es cero, en el segundo supuesto, la variabilidad es alta.

Conocer que la temperatura promedio de la ciudad es de 17 °C proporciona un dato útil, pero no suficiente porque no informa sobre la temperatura máxima en el verano o la mínima en el invierno. Es más, esa ciudad podría estar en el desierto, donde la arena no retiene el calor, y esas variaciones de temperatura podrían ocurrir cada 12 horas alcanzando temperaturas infernales al mediodía y gélidas por la madrugada.

En cada uno de estos ejemplos, el promedio es el mismo (10 años o 17 °C), pero la dispersión puede ser muy diferente, de allí que sea tan importante conocer las medidas de posición como la variabilidad alrededor de ellas.

4.5.1. MEDICIÓN DE LA VARIABILIDAD

En ejercicios previos, ya se han efectuado mediciones de variabilidad pues siempre se ha tomado en cuenta el *rango* de los datos (también llamado recorrido o amplitud). Otras mediciones de variabilidad son:

- La desviación media
- La desviación estándar
- La variancia
- El coeficiente de variación

El *rango* o *recorrido* solamente considera los valores extremos de un conjunto de datos y es la diferencia entre el valor mayor y el menor de la distribución. En el ejemplo de las calificaciones de las once materias que cursó un estudiante en su primer año:

$$70, 75, 75, 80, 85, 85, 90, 90, 90, 90, 100$$

Se sabe que las calificaciones van del 70 al 100, también puede decirse que el recorrido es $R = L_s - L_i = 100 - 70 = 30$.

El rango o recorrido tiene como ventajas la sencillez y rapidez de cálculo, aún cuando se trate de un número grande de datos, también es de muy fácil compresión. Como desventajas pueden anotarse sus limitaciones para el análisis matemático, luego, que se ve afectado por los valores extremos y que no brinda mayor información acerca de los demás valores del intervalo. Estas limitaciones pueden visualizarse en la Figura 4.8, que presenta tres series diferentes con el mismo rango o recorrido pero con diferencias de concentración.

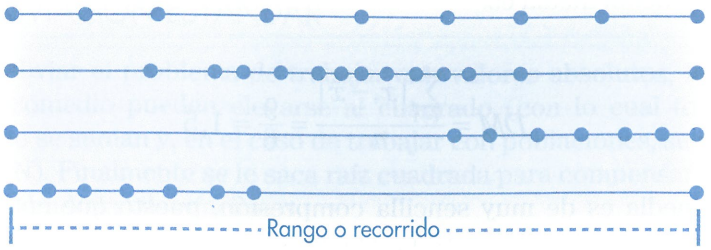


Figura 4.8. **Diferentes series con el mismo rango o recorrido**

4.5.2. LA DESVIACIÓN MEDIA

La *desviación media* incorpora en su cálculo todos los datos y se basa en las desviaciones o diferencias de cada dato individual con respecto de la media geométrica o promedio. Dado que la sumatoria de las desviaciones con respecto del promedio siempre dará cero, se hace necesario utilizar los valores absolutos para el cálculo. La sumatoria de los valores absolutos de las desviaciones respecto del promedio se divide entre el número de datos, así:

$$\text{DM} = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n} = \frac{\textit{suma de los valores absolutos de las desviaciones}}{\textit{número de datos}}$$

La desviación media es la media aritmética de los valores absolutos de las desviaciones respecto del promedio. Si se tienen los valores siguientes:

x_1	$x_1 - 7{,}5$	$\lvert x_1 - 7{,}5 \rvert$
8	0,5	0,5
5	–2,5	2,5
10	2,5	2,5
9	1,5	1,5
7	–0,5	0,5
6	–1,5	1,5
$\sum$ = 45	0	9

Su promedio es:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{45}{6} = 7{,}5$$

Entonces la desviación media es:

$$DM = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n} = \frac{9}{6} = 1,5$$

La desviación media es de muy sencilla compresión, puesto que entre más alto su valor, se sabe que la dispersión es mayor; sin embargo, el hecho de trabajar con valores absolutos le establece muchas limitaciones para posteriores desarrollos.

EJERCICIO 4.6

Se tienen las siguientes observaciones sobre el nivel de glucosa en la sangre de un paciente en ayunas:

x_i	$x_i - \bar{x}$	$\lvert x_i - \bar{x} \rvert$
75		
100		
92		
100		
76		
80		
70		
75		
82		
100		
SUMA		

Calcule:

i. El promedio del contenido de glucosa en la sangre de este paciente en ayunas.

ii. El rango del contenido de glucosa en la sangre.

ii. La desviación media.

4.5.3. LA DESVIACIÓN ESTÁNDAR

Con el fin de obviar el problema de trabajar con valores absolutos, las desviaciones respecto del promedio pueden elevarse al cuadrado (con lo cual todas se vuelven positivas), luego se suman y, en el caso de trabajar con poblaciones, se dividen entre el total de datos (N). Finalmente se le saca raíz cuadrada para compensar el hecho inicial de elevar al cuadrado las desviaciones.

$$\sigma = \text{Desviación estándar poblacional} = \sqrt{\frac{\sum (\textit{desviaciones})^2}{\textit{número de datos}}} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{N}}$$

Cuando se trabaja con muestras, para el cálculo de la *desviación estándar*, la sumatoria del cuadrado de las desviaciones se divide entre el tamaño de la muestra menos uno (n – 1) y luego se saca su raíz cuadrada, así:

$$s = \text{Desviación estándar de la muestra} = \sqrt{\frac{\sum (\textit{desviaciones})^2}{\textit{número de datos menos 1}}} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

De la muestra siguiente:

x_i	$x_i - 7{,}5$	$(x_i - 7{,}5)^2$
8	0,5	0,25
5	–2,5	6,25
10	2,5	6,25
9	1,5	2,25
7	–0,5	0,25
6	–1,5	2,25
$\sum = 45$	0	17,5

La desviación estándar de la muestra resulta:

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{17{,}5}{6-1}}$$

$$s = \sqrt{\frac{17{,}5}{5}} = \sqrt{3{,}5} = 1{,}87$$

La desviación estándar de 1,87 indica cuánto se alejan las observaciones, globalmente, del promedio 7,5.

La desviación estándar y el promedio son las medidas de resumen más utilizadas para describir o analizar las distribuciones, y las unidades en que se expresan son las mismas. Por ejemplo, si el promedio está expresado en gramos, la desviación estándar también estará expresada en gramos; si el promedio se mide en centímetros, la desviación estándar se medirá en centímetros; si el promedio se indica en grados centígrados, lo mismo ocurrirá con la desviación estándar.

EJERCICIO 4.7

Se tiene una muestra de observaciones acerca del nivel de glucosa en la sangre de un paciente en ayunas:

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
75		
100		
92		
100		
76		
80		
70		
75		
82		
100		
SUMA		

Calcule la desviación estándar de esta muestra.

4.5.4. LA VARIANCIA

Puesto que es más cómodo trabajar sin la raíz cuadrada, se ha desarrollado el concepto de *variancia*, que es el cuadrado de la desviación estándar.

En el caso de trabajar con poblaciones, la suma del cuadrado de las desviaciones se divide entre el número de datos:

$$\sigma^2 = \text{Variancia poblacional} = \frac{\sum (desviaciones)^2}{número\ de\ datos} = \frac{\sum (x_i - \bar{x})^2}{N}$$

Entonces, la variancia es la media aritmética del cuadrado de las desviaciones respecto del promedio de una distribución.

Cuando se trabaja con muestras, para el cálculo de la variancia, la sumatoria del cuadrado de las desviaciones se divide entre el tamaño de la muestra menos uno (n – 1), así:

$$s^2 = \text{Variancia de la muestra} = \frac{\sum (desviaciones)^2}{número\ de\ datos\ menos\ 1} = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

De la muestra siguiente:

x_i	$x_i - 7{,}5$	$(x_i - 7{,}5)^2$
8	0,5	0,25
5	-2,5	6,25
10	2,5	6,25
9	1,5	2,25
7	-0,5	0,25
6	-1,5	2,25
$\sum$ = 45	0	17,5

La variancia muestral es:

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{17{,}5}{6-1}$$

$$s^2 = \frac{17{,}5}{5} = 3{,}5$$

4.5.5. CÁLCULO DE LA VARIANCIA PARA DATOS AGRUPADOS

La variancia de una muestra con datos agrupados necesita utilizar el punto medio de las clases, puesto que es a este al que se le restará el promedio, para luego multiplicar esa desviación por la frecuencia de cada clase.

Las desviaciones multiplicadas se suman y luego se dividen por n – 1.

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2 f_i}{n-1}$$

Esta expresión puede convertirse algebraicamente en una fórmula que resulta más práctica para efectuar los cálculos:

$$s^2 = \frac{\left[\sum_{i=1}^{n} (x_i^2 f_i)\right] - \left[\frac{\left(\sum_{i=1}^{n} x_i f_i\right)^2}{n}\right]}{n-1}$$

Con los datos siguientes:

CLASES Peso en libras	Punto medio x_i	Frecuencia f_i	$x_i f_i$	$x_i^2 f_i$
85,5 a 100,5	93	1	93	8 649
100,5 a 115,5	108	3	324	34 992
115,5 a 130,5	123	20	2 460	302 580
130,5 a 145,5	138	21	2 898	399 924
145,5 a 160,5	153	17	2 601	397 953
160,5 a 175,5	168	10	1 680	282 240
175,5 a 190,5	183	8	1 464	267 912
190,5 a 205,5	198	5	990	196 020
		85	12 510	1 890 270

Se sustituye:

$$s^2 = \frac{1\ 890\ 270 - \left[\frac{(12\ 510)^2}{85}\right]}{85-1}$$

$$s^2 = \frac{1\ 890\ 270 - 1\ 841\ 177{,}65}{84}$$

$$s^2 = 584{,}4$$

La variancia es 584,4. Como la desviación estándar es la raíz cuadrada de la variancia, de allí se tiene que:

$$s = \sqrt{584{,}4} = 24{,}17 \text{ libras}$$

EJERCICIO 4.8

a) El cuadro siguiente informa la distribución de frecuencias del peso de una muestra de 40 recién nacidos en una clínica rural.

CLASES	Puntos medios			
(Peso en gramos)	x_1	f_1	$x_1 f_1$	$x_i^2 f_1$
499,5 – 999,5	749,5	1	749,5	561 750,3
999,5 – 1 499,5	1 249,5	2	2 499,0	3 122 500,5
1 499,5 – 1 999,5	1 749,5	2	3 499,0	6 121 500,5
1 999,5 – 2 499,5	2 249,5	4	8 998,0	20 241 001,0
2 499,5 – 2 999,5	2 749,5	5	13 747,5	37 798 751,3
2 999,5 – 3 499,5	3 249,5	7	22 746,5	73 914 751,8
3 499,5 – 3 999,5	3 749,5	10	37 495,0	140 587 502,5
3 999,5 – 4 499,5	4 249,5	6	25 497,0	108 349 501,5
4 499,5 – 4 999,5	4 749,5	3	14 248,5	67 673 250,8
		40	129 480,0	458 370 510,0

Con esa información calcule:

i. La variancia.

ii. La desviación estándar del peso de los recién nacidos.

b) Se cuenta con la siguiente distribución acerca de la edad en años cumplidos de una muestra de 50 mujeres al momento de dar a luz en el Hospital de la Anexión, durante diciembre de 2014.

CLASES	Puntos medios			
(Edades)	x_i	f_i	$x_i f_i$	$x_i^2 f_i$
10 a menos de15	12,5	2	25	312,50
15 a menos de 20	17,5	4	70	1 225,00
20 a menos de 25	22,5	5	112,5	2 531,25
25 a menos de 30	27,5	12	330	9 075,00
30 a menos de 35	32,5	14	455	14 787,50
35 a menos de 40	37,5	9	337,5	12 656,25
40 a menos de 45	42,5	3	127,5	5 418,75
45 a menos de 50	47,5	1	47,5	2 256,25
		50	1 505	48 262,50

Con esa información calcule:

i. La variancia.

ii. La desviación estándar del peso de los recién nacidos.

4.5.6. COEFICIENTE DE VARIACIÓN

El coeficiente de variación indica la importancia de la desviación estándar en relación con el promedio y permite comparar dos o más conjuntos de datos en cuanto a su variabilidad:

$$\text{CV} = \left(\frac{\text{Desviación estándar}}{\text{Promedio}}\right) \times 100$$

Cuando se trabaja con muestras el coeficiente de variación es:

$$\text{CV} = \left(\frac{s}{\bar{x}}\right) \times 100$$

Y si se trabaja con poblaciones es:

$$\text{CV} = \left(\frac{\sigma}{\mu}\right) \times 100$$

Este coeficiente permite comparar magnitudes de diferente orden, también posibilita el comparar la variabilidad de distribuciones expresadas en diferentes unidades de medida y corrige el efecto que sobre la desviación estándar tiene la magnitud general de los datos.

Si se tiene un promedio de $\bar{x} = 7{,}5$ y una desviación estándar de $s = 1{,}87$, el coeficiente de variación es:

$$\text{CV} = \left(\frac{1{,}87}{7{,}5}\right) \times 100 = 24{,}9\,\%$$

Este resultado indica que la desviación estándar representa un 24,9 % del promedio.

4.5.7. PERCENTILES COMO MEDIDAS DE VARIABILIDAD

Antes se ha visto que los percentiles y el rango percentil informan sobre la posición que una observación ocupa en una distribución de frecuencias; ahora se pasará a analizar varios percentiles juntos para apreciar la variabilidad o dispersión de los datos.

Con dos percentiles diferentes se delimita un intervalo y luego se analiza el porcentaje de observaciones que contiene. La construcción que se hace con mayor frecuencia es el *intervalo intercuartil*, que va del primero al tercer cuartil (Q_1 *al* Q_3) o, lo que es igual, del percentil 25 al percentil 75 ($P_{0,25}$ *al* $P_{0,75}$).

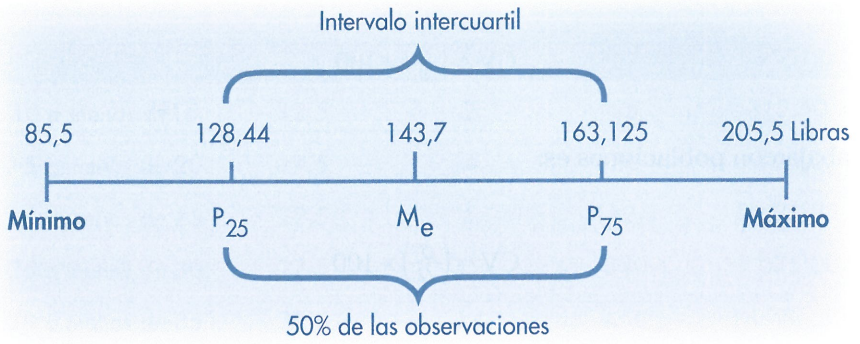


Figura 4.9. **Representación del intervalo intercuartil**

Calcular el intervalo intercuartil es de mucha importancia puesto que contiene a la mediana (Me) y al 50 % de las observaciones centrales, esto es, que no toma en cuenta al 25 % de las observaciones muy bajas ni al 25 % de las observaciones muy altas (valores extremos). Visto así, el intervalo intercuartil es una medida ajustada de dispersión alrededor de la mediana.

Por supuesto que pueden construirse otros intervalos, como el que va del percentil 10 al percentil 90 ($P_{0,10}$ *al* $P_{0,90}$), pues en él se encuentra el 80 % de las observaciones centrales.

4.6. FIJACIÓN DE LÍMITES DE TOLERANCIA

Para controlar o evaluar indicadores de salud, conviene establecer límites aceptables de variabilidad, conocidos como límites de tolerancia, por ejemplo:

Análisis	Límite inferior mg/dL	Límite superior mg/dL	Observaciones
Colesterol total	140	200	173
Triglicéridos	35	150	54
Colesterol HDL	30	85	69
Colesterol LDLc	0	140	94
Colesterol VLDL	20	42	25

Esto ayudará al médico tratante para trabajar sobre las desviaciones con el fin de corregirlas, ya sea buscando un aumento, una reducción o su eliminación.

La determinación de los límites de tolerancia marcan la diferencia entre un paciente sano y uno enfermo, una situación controlada o una epidemia, políticas adecuadas o incorrectas, servicio al asegurado eficiente o deficiente.

Para la fijación de los límites de tolerancia, es importante conocer si la distribución tiene forma simétrica o asimétrica.

- Cuando la distribución es simétrica, el intervalo se establece partiendo del promedio y sumando o restando algún múltiplo de la desviación estándar.

$$\text{Límite inferior } \bar{x} - m\sigma$$

$$\text{Límite superior } \bar{x} + m\sigma$$

$$\text{Donde } m = 1;\ 1{,}5;\ 2;\ 2{,}5 \text{ o } 3$$

- Cuando la distribución es asimétrica, los límites se establecen partiendo de los percentiles 25 y 75. El límite inferior es el percentil 25 $(P_{0,25})$ menos algún múltiplo del intervalo intercuartil. El límite superior es el percentil 75 $(P_{0,25})$ más algún múltiplo del intervalo intercuartil.

$$\text{Límite inferior } P_{0,25} - [m(P_{0,75} - P_{0,25})]$$

$$\text{Límite superior } P_{0,75} - [m(P_{0,75} + P_{0,25})]$$

Ejemplo:

Con respecto a la edad de las mujeres que dan a luz en el Hospital de la Anexión, se ha calculado el promedio en 30,1 años, la desviación estándar en 7,8 años y se sabe que la distribución de edades es bastante simétrica. Estableciendo un múltiplo de 1,7 desviaciones estándar, ¿cuáles son los límites de edad entre los cuales un embarazo es conveniente desde el punto de vista de la salud de la madre y del niño?

$$\text{Límite inferior} = \bar{x} - m\sigma$$

$$30,1 - (1,7 \times 7,8)$$

$$30,1 - 13,26 = 16,84 \text{ años}$$

$$\text{Límite superior} = \bar{x} + m\sigma$$

$$30,1 + (1,7 \times 7,8)$$

$$30,1 + 13,26 = 43,36 \text{ años}$$

La edad conveniente para los embarazos está entre 16,84 y 43,36 años. Se aclara que esto es así desde una perspectiva estadística, puesto que en este caso criterios sociales, culturales, religiosos o de política demográfica podrían modificarlos.

Estadísticamente, los embarazos y partos de mujeres menores de 16,84 años se consideran de riesgo, igualmente lo son los embarazos y partos de mujeres mayores de 43,36 años.

Ejemplo:

Las incapacidades por psiquiatría de los maestros de secundaria tienen un promedio de 7 días. Estudios indican que el percentil 25 es 6 días de incapacidad y el percentil 75 informa 11 días de incapacidad, para un intervalo intercuartil de 5 días. Las autoridades buscan establecer límites de tolerancia para auditar una muestra de casos que esté fuera de esos límites de tolerancia considerando una m = 0,8.

$$\text{Límite inferior} = P_{0,25} - [m(P_{0,75} - P_{0,25})]$$

$$L_i = 6 - [0,8(11 - 6)]$$

$$L_i = 6 - [0,8 \times 5]$$

$$L_i = 6 - 4 = 2 \text{ días}$$

$$\text{Límite superior } P_{0,75} + [m(P_{0,75} - P_{0,25})]$$

$$L_s = 11 + [0,8(11 - 6)]$$

$$L_i = 11 + [0,8 \times 5]$$

$$L_i = 11 + 4 = 15 \text{ días}$$

Las incapacidades fuera de los límites de tolerancia son los menores de 2 días y las mayores de 15 días.

EJERCICIO 4.9

i. El peso al nacer de una muestra de recién nacidos ha permitido establecer un peso promedio de 3 237 gramos, con desviación estándar de 1 003,12 gramos. A una distancia del promedio de m = 1,23 desviaciones estándar, establezca el peso mínimo límite, a partir del cual se activarán todas las alarmas médicas para dedicarles máxima atención.

ii. Los niveles de colesterol en la sangre de un paciente sano indican que en el percentil 25 se presentan 140 mg/dL y en el percentil 75 un total de 200 mg/dL. Utilizando como múltiplo m = 1,8, calcule el límite superior de colesterol en el cual un paciente tendrá que ser intervenido clínicamente.

4.7. LA VARIABILIDAD EN EL CASO DE PROPORCIONES

Los conceptos de media, variancia y desviación estándar son también aplicables para casos de proporciones. Estas pueden provenir de variables dicotómicas como estar enfermo o sano, masculino o femenino, asegurado o sin seguro, y también pueden calcularse a partir de variables cualitativas multinomiales, toda vez que se transformen en dicotómicas, por ejemplo: costarricense o no costarricense, soltero o no soltero.

Si $x_i = N_1$ indica el número de pacientes con una característica, entonces su proporción es:

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{N_1}{N} = P$$

$(1 - P)$ representará, entonces, la proporción de pacientes sin esa característica.

La variancia de la población se calculará así:

$$\sigma^2 = P\,(1-P)$$

Y cuando se trabaje con muestras pequeñas se deberá utilizar un factor de corrección:

$$s^2 = \left(\frac{n}{n-1}\right) P\,(1-P)$$

A medida que el tamaño de la muestra aumenta, el factor $\left(\frac{n}{n-1}\right)$ se hace más cercano a 1, y podría usarse simplemente $s^2 = P\,(1-P)$.

Si, por ejemplo, en una muestra de 200 pacientes, 60 tienen problemas de sobrepeso, la proporción con esa característica será:

$$P = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{N_1}{N} = \frac{60}{200} = 0,30$$

El 0,30 de los pacientes padece de sobrepeso. La variancia es:

$$s^2 = P\,(1-P) = 0,3 \times (1-0,3) = 0,3 \times 0,7 = 0,21$$

Se sabe que la desviación estándar es la raíz cuadrada de la variancia, por lo que:

$$s = \sqrt{P\,(1-P)}$$

$$s = \sqrt{0,3 \times (0,7)} = \sqrt{0,21} = 0,458$$

0,458 es la desviación estándar de la proporción de pacientes con sobrepeso.

EJERCICIO 4.10

i. De una muestra de 300 pacientes atendidos en el servicio de emergencias en el Hospital Nacional de Niños, 96 fueron atendidos con problemas respiratorios. Con esos datos calcule:

   a) La proporción de pacientes atendidos por problemas respiratorios.

   b) La variancia de esa proporción.

   c) La desviación estándar.

ii. De una muestra de 1 000 personas, se determinó que 600 tenían sangre tipo O. Calcule:

   a) La proporción de pacientes con sangre tipo O.

   b) La variancia de esa proporción.

   c) La desviación estándar.

iii. De una muestra de 10 escolares de sexto grado, 8 presentan caries dentales o secuelas de las mismas. Calcule:

   a) La proporción de escolares que padece caries o secuelas de esta enfermedad.

   b) La variancia de esa proporción.

   c) La desviación estándar.

4.8. APLICACIONES AL ÁREA DE SALUD: EL ÍNDICE ENDÉMICO

El registro de los casos de una enfermedad permite construir un índice endémico, con el cual se puede detectar a tiempo un número alto durante una época del año para prevenir epidemias e implementar medidas oportunas de control. Un registro es como el siguiente:

Casos mensuales de rubéola en el cantón de Turrucalito
Registrados del año 2002 al 2009

AÑO	Enero	Feb.	Marzo	Abril	Mayo	Junio	Julio	Agosto	Set.	Oct.	Nov.	Dic.	TOTAL
2002	3	12	18	22	16	7	7	4	5	4	5	0	103
2003	3	6	12	16	19	7	11	2	1	2	1	0	80
2005	7	11	8	8	4	8	4	4	5	4	3	0	66
2006	2	2	7	2	7	4	0	2	2	0	3	0	31
2007	0	4	2	7	5	1	1	0	0	1	0	0	21
2008	5	0	0	0	6	0	0	1	2	2	0	0	16

Fuente: **Datos hipotéticos con propósitos didácticos**

Luego estos datos se ordenan de menor a mayor número de casos en el mes:

Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Set.	Oct.	Nov.	Dic.	
0	0	0	0	4	0	0	0	0	0	0	0	
2	**2**	**2**	**2**	**5**	**1**	**0**	**1**	**1**	**0**	**0**	**0**	**Primer Cuartil**
3	4	7	7	6	4	1	2	1	1	1	0	
3	**6**	**8**	**8**	**6**	**7**	**4**	**2**	**2**	**2**	**2**	**0**	**Mediana**
3	7	10	12	7	7	4	2	2	2	3	0	
5	**11**	**12**	**16**	**16**	**8**	**7**	**4**	**5**	**4**	**3**	**0**	**Tercer cuartil**
7	12	18	22	19	10	11	4	5	4	5	0	

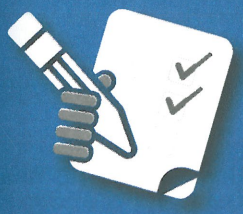
Se calcula la mediana (Me) y los cuartiles 1 y 3. El índice endémico es la mediana de cada mes, se tiene como variaciones toleradas las que se dan al interior del intervalo intercuartil de cada mes.

Cuando los casos del mes superan el tercer cuartil, se activarán las medidas para combatir una epidemia. Si se llega a niveles menores del primer cuartil, se dice que las medidas de control sanitario son eficaces.

EJERCICIO 4.11

En el cuadro anterior, ¿qué número de casos activarán las alarmas sanitarias al finalizar abril?

En mayo, ¿en qué situación se considerará que la epidemia fue controlada?



EJERCICIOS DE REPASO Y AUTOEVALUACIÓN

Esta sección pretende que haga un repaso general de los contenidos del Capítulo.

EJERCICIO 4.12

Encierre en un círculo la opción que se ajusta al enunciado, Falso (F) o Verdadero (V).

F V 1. Las medidas de posición, forma y variabilidad buscan calcular valores representativos de todos los datos, que expliquen el comportamiento general de la distribución, además de describir esta distribución y compararla para llegar a mejores conclusiones.

F V 2. La mediana es la cantidad o cualidad que más se repite en un conjunto de datos, técnicamente se define como el valor más frecuente.

F V 3. Una distribución de frecuencias determinada puede tener varios promedios.

F V 4. Si se resta el promedio a cada una de las observaciones y luego se suman sus resultados, la sumatoria siempre será uno.

F V 5. La mayor desventaja del promedio es que va a ser influido por la presencia de valores extremos.

F V 6. El cálculo del promedio en datos agrupados será exactamente igual al promedio de todos y cada uno de los datos, puesto que al usar los puntos medios de cada clase se garantiza exactitud.

F V 7. Cuando los datos presentan una distribución simétrica, el promedio se encontrará entre la media y la mediana.

F V 8. Si el promedio es mayor que la moda ($\bar{x} > Mo$) el coeficiente de asimetría es positivo y una cola de la distribución se prolonga hacia la derecha, entonces se dice que la distribución tiene *asimetría positiva* o hacia la derecha.

F V 9. Cuando el coeficiente de asimetría es positivo, esto indica que la situación de salud de una comunidad anda por buen camino.

F V 10. La mediana y el percentil cincuenta ($P_{0,50}$) producen el mismo dato.

F V 11. El percentil 75 ($P_{0,75}$) y el primer cuartil (Q_1) también generan valores coincidentes.

F V 12. En aspectos de salud comunitaria, no es relevante la variabilidad puesto que los hechos se repiten con frecuencia.

F V 13. El rango o recorrido analiza todos y cada uno de los valores de una distribución.

F V 14. El rango o recorrido tiene como desventaja que cuando se trata de un número grande de datos, se complica su cálculo y dificulta la comprensión.

F V 15. La desviación media es la media aritmética de los valores absolutos de las desviaciones respecto del promedio.

F V 16. El mayor problema de utilizar la desviación estándar es cuando su valor resulta negativo.

F V 17. La desviación estándar y el promedio son las medidas de resumen más utilizadas para describir o analizar las distribuciones, y las unidades en que se expresan son las mismas.

F V 18. La variancia es la mediana del cuadrado de las desviaciones respecto de la moda de una distribución.

F V 19. El coeficiente de variación indica la importancia de la desviación estándar en relación con el promedio.

F V 20. Los límites de tolerancia permiten obtener los valores mínimo y máximo que diferencian salud y enfermedad y buenas o malas prácticas.

EJERCICIO 4.13

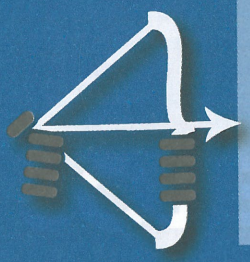
Pareo. Llene los paréntesis con el número que acompaña a cada definición de la izquierda, de forma que cada definición coincida con el concepto en la derecha. Quedan dos paréntesis en blanco.

1.	Explica la importancia de la desviación estándar de una serie de datos en relación con su promedio.	( )	La desviación media
2.	Son los valores que dividen a un conjunto ordenado de datos en fracciones de igual magnitud.	( )	Los percentiles
3.	Es la media aritmética de los valores absolutos de las desviaciones respecto del promedio.	( )	La moda
4.	Es el valor central de una serie de datos ordenados, el 50 % de los datos está por debajo de ese valor y el 50 % por encima.	( )	El rango
5.	Es la cantidad o cualidad que más se repite en un conjunto de datos; técnicamente se define como el valor más frecuente.	( )	El promedio
6.	Son los valores que dividen un conjunto ordenado de datos en cien partes iguales.	( )	La variancia
7.	Es el valor que se obtiene de la sumatoria de todos los datos dividida entre el número de observaciones.	( )	El coeficiente de variación
8.	Valor que informa del intervalo contenido entre el valor mayor y el menor de los datos.	( )	Los deciles
9.	Es la media aritmética del cuadrado de las desviaciones respecto del promedio de una distribución.	( )	La asimetría
10.	Son los valores que dividen un conjunto ordenado de datos en diez partes iguales.	( )	El apuntamiento
		( )	Los cuantiles
		( )	La mediana

CAPÍTULO 5

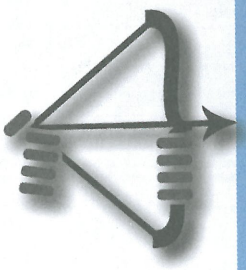
ESTADÍSTICAS DE LOS SERVICIOS DE SALUD Y EL ANÁLISIS GRÁFICO

OBJETIVO GENERAL



Manejar conocimientos sobre la representación gráfica de datos y series estadísticas para su apropiada utilización en el análisis de los datos obtenidos y su aplicación en el campo de la salud.

OBJETIVOS ESPECÍFICOS



Al finalizar el estudio de este capítulo usted estará en la capacidad de:

1. Identificar los elementos esenciales de todo gráfico y utilizarlos adecuadamente para su confección.
2. Redactar los títulos de los gráficos de forma que indiquen qué es y cómo, cuándo y dónde se recolectaron los datos.
3. Utilizar las dimensiones apropiadas para el cuerpo de los gráficos, de forma que la información no se exagere ni se atenúe.
4. Dibujar los diferentes tipos de gráfico a partir de tablas con datos estadísticos.
5. Escoger el gráfico conveniente para el tipo de serie estadística que se está informando.
6. Combinar en forma oportuna el instrumental gráfico cuando presente sus informes de investigación.

Este capítulo presenta las técnicas de representación gráfica de los datos estadísticos, proporcionando herramientas tanto para su confección como para la selección del tipo de gráfico a utilizar en los informes de investigación.

Quien no comprende una mirada,
tampoco comprenderá una larga explicación.
PROVERBIO ÁRABE

Los gráficos estadísticos comunican la información utilizando:

- Distancias (para una dimensión)
- Áreas (para dos dimensiones)
- Pendientes (para dos dimensiones)
- Volúmenes (para tres dimensiones)

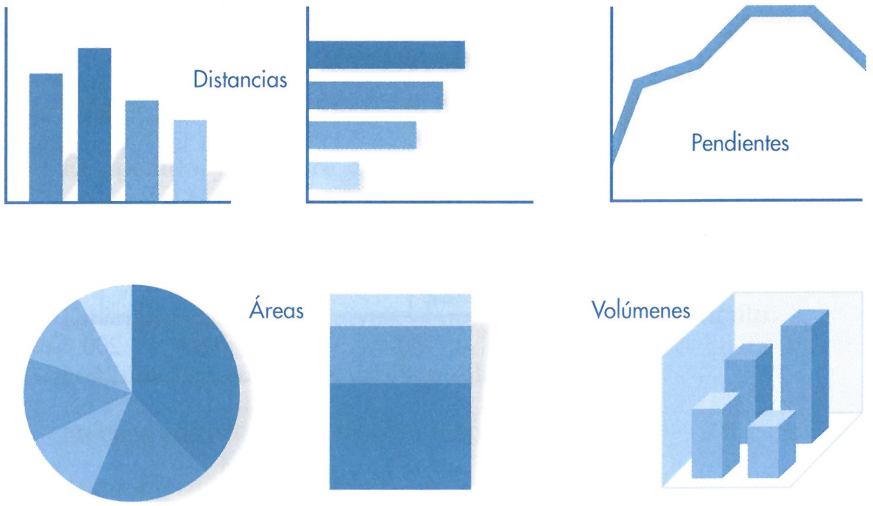


Figura 5.1. **Magnitudes usadas para la elaboración de gráficos**
Fuente: Elaboración propia con fines didácticos

Estas magnitudes (distancias, pendientes, áreas o volúmenes) representan los datos numéricos y, como ventaja, se señala la relativa facilidad para ser interpretadas, además de atraer la atención de los lectores.

5.1. ELEMENTOS DE LOS GRÁFICOS

Para ayudar a su comprensión, los gráficos requieren la presencia de elementos fundamentales tales como:

- El *título*, que debe ser redactado con detalles similares a los de los cuadros (¿qué? ¿dónde? ¿cómo? y ¿cuándo?), para que se describa con claridad su contenido.
- El *cuerpo* del gráfico, cuya relación base por altura no debe sobrepasar la proporción 1,5 a 1, pues eso alteraría las pendientes, las distancias o las áreas, distorsionando la información.
- El nombre de *las categorías* que se presentan. Dependiendo del tipo de gráfico, estas pueden ir en los ejes o en el cuerpo del diagrama.
- La *escala* y otras unidades geométricas o de tiempo empleadas.

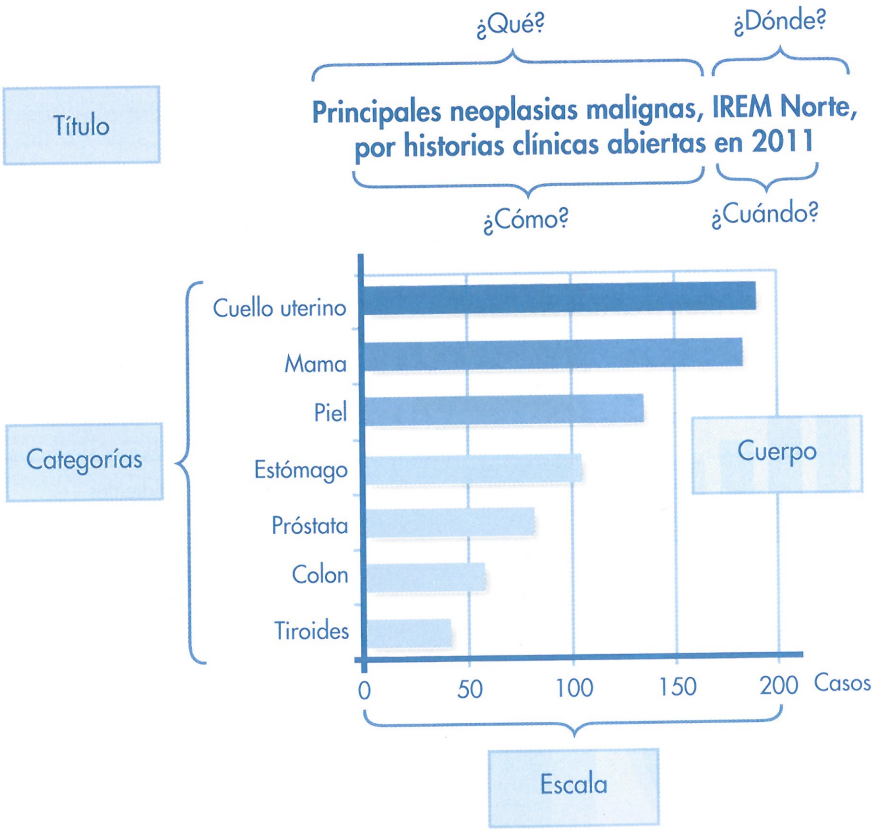


Figura 5.2. **Elementos fundamentales que debe contener todo gráfico**

Fuente: Uso didáctico de un gráfico de IREM (2012)

Adicionalmente, los gráficos pueden incluir:

- El *número* de gráfico, que es muy útil cuando un informe utiliza varios
- La *nota introductoria* que aclara una parte o todo el contenido del gráfico; se coloca centrada entre paréntesis o guiones debajo del título y en minúscula.

  Una nota introductoria precisa:

  - Las unidades en que están los datos: (En dólares), (En miles de colones), (Tasas por mil habitantes)
  - Una mayor especificación del título: (Muestra de estudiantes), (Datos censales)
  - Limitaciones de los datos: (Cifras preliminares), (Proyecciones)
  - En el caso de presentar números índices, indica la base con la que se realizan las comparaciones: (Enero 1995 = 100)

- La *simbología*, que puede ir en un recuadro y es indispensable cuando se presenta más de una serie
- La *fuente*, la cual es obligatoria cuando los datos que se utilizan para su confección han sido obtenidos de otras investigaciones o cuando el gráfico completo ha sido tomado de otros estudios.

  Para no saturar el gráfico, se estila que la referencia de la fuente solamente lleve el apellido del autor o institución responsable y el año de la publicación, que va entre paréntesis, por ejemplo:

  Fuente: Durán, F. (2001)

  En la bibliografía, se hará una cita más completa que contenga:

  - El apellido y nombre del autor, autores o el nombre de la institución responsable del trabajo utilizado
  - El año del estudio o de su publicación, que va entre paréntesis
  - En letras itálicas o cursivas, el nombre del libro o artículo
  - La ciudad en que se realizó la publicación y
  - La casa editorial o institución responsable

  Por ejemplo:

  Durán, Fabio. (2001). *Equidad y acceso a los servicios de salud en Costa Rica.* San José: CCSS.

- Los *cortes de escala en los ejes*, si los hubiere. Generalmente se usan cortes cuando la base es cero, pero las cifras informativas comienzan muy distantes del origen. Algunas formas adecuadas de representar los cortes de escala se muestran en la Figura 5.3.

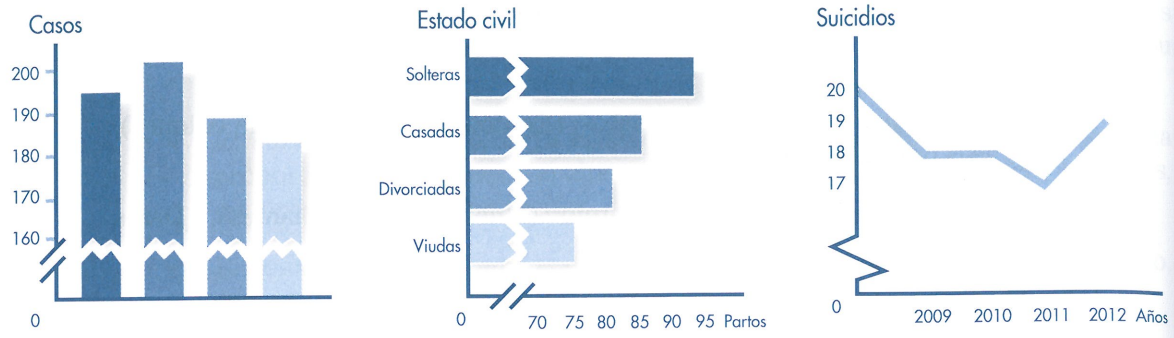
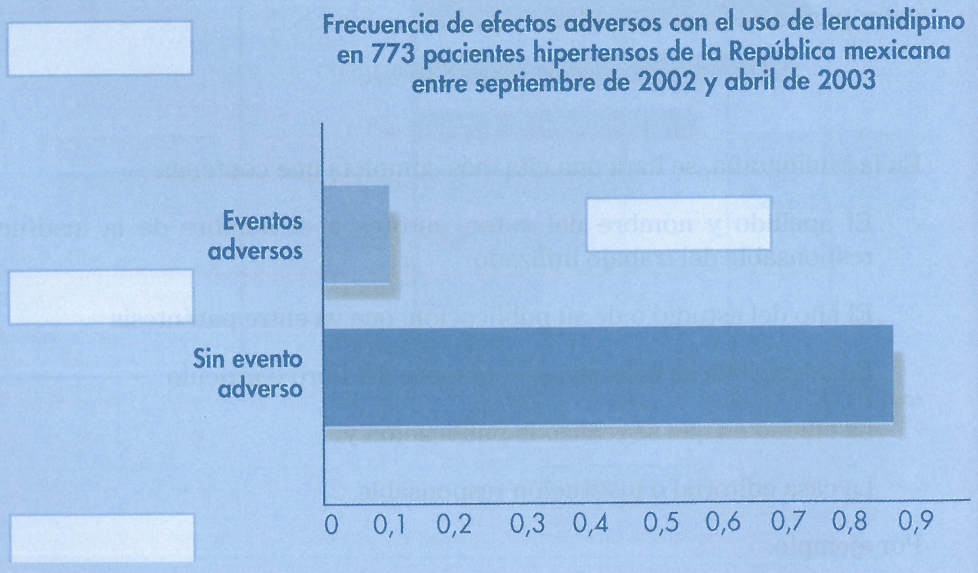


Figura 5.3. **Formas usuales de indicar los cortes de escala en los ejes**

Fuente: Elaboración propia con fines didácticos

EJERCICIO 5.1

a) En el siguiente gráfico, señale los elementos fundamentales que ayudan a su adecuada comprensión.



Fuente: Lepe Montoya (2004)

b) Señale los elementos explicativos en la redacción del siguiente título (¿qué? ¿dónde? ¿cómo? y ¿cuándo?).

> Frecuencia de efectos adversos con el uso de lercanidipino en 773 pacientes hipertensos de la República mexicana entre septiembre de 2002 y abril de 2003

c) Liste los elementos adicionales que puede contener un gráfico para mejorar la calidad de los informes.

5.2. LIMITACIONES DE LOS GRÁFICOS

Entre las limitaciones de los gráficos se puede anotar que:

- No permiten la inclusión de muchas variables, pues lo recargarían volviéndolo confuso.
- La información se da en valores aproximados.
- Su confección requiere habilidades para dibujo o manejar paquetes informáticos [4].

Adicionalmente, tome en cuenta que los gráficos no pueden utilizarse como sustitutos de los cuadros porque reducirían los informes a panfletos de divulgación o propaganda. Recuerde que los buenos informes de investigación utilizan tanto cuadros como gráficos, por ser herramientas complementarias para la presentación de los hallazgos.

5.3. TIPOS DE GRÁFICOS

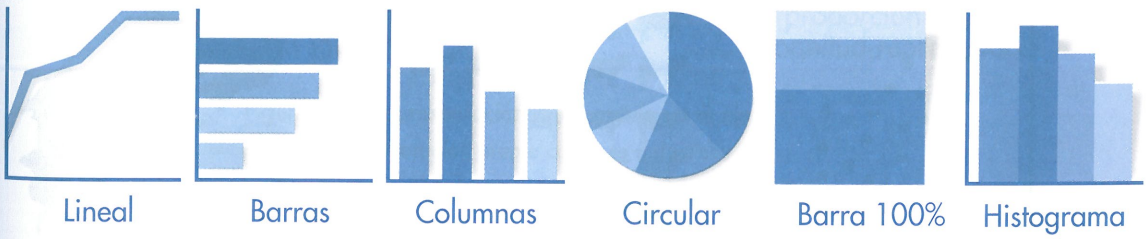


Figura 5.4. **Gráficos más utilizados para la presentación de informes generales**

Fuente: Elaboración propia con fines didácticos

4 En el ANEXO 3 encontrará algunos pasos sencillos para la confección de gráficos utilizando Excel de Microsoft.

Cuando se redacta un informe, se utilizan con mayor frecuencia los gráficos:

- Lineales
- De barras simples, compuestas o comparativas
- De columnas (igualmente pueden ser: simples, compuestas o comparativas)
- Circulares
- Columna cien por ciento
- Histogramas

También se utilizan otros gráficos especializados como:

- Barras de dos direcciones
- Ojivas de frecuencias acumuladas
- Pirámides poblacionales
- Diagramas de dispersión
- Pictogramas
- Mapas estadísticos

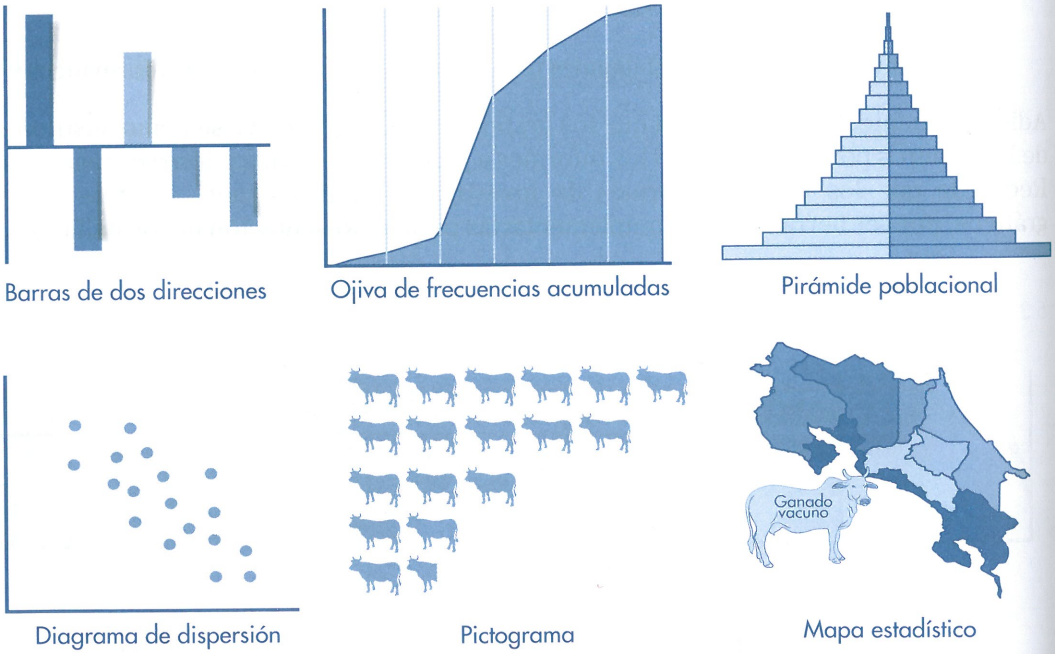


Figura 5.5. **Gráficos estadísticos especializados**
Fuente: Elaboración propia con fines didácticos

5.3.1. GRÁFICO LINEAL

Los gráficos lineales son muy útiles para la representación de series cronológicas. En el eje horizontal se ubica el tiempo (meses, años, décadas, siglos) y en el eje vertical se incluye la escala en la que se informan las ocurrencias.

Las claves que permiten interpretar adecuadamente al gráfico lineal son:

- La distancia
- La tendencia
- Los cambios de rumbo
- La inclinación de las líneas que unen los diferentes puntos.

Para que estas claves se interpreten adecuadamente, al momento de dibujar la escala debe respetarse rigurosamente el principio de que "distancias iguales representan siempre magnitudes iguales". En la Figura 5.6, los dos gráficos de la izquierda respetan este principio, por lo que se puede afirmar que las escalas están bien señaladas; en cambio, los dos gráficos de la derecha tienen escalas incorrectas.

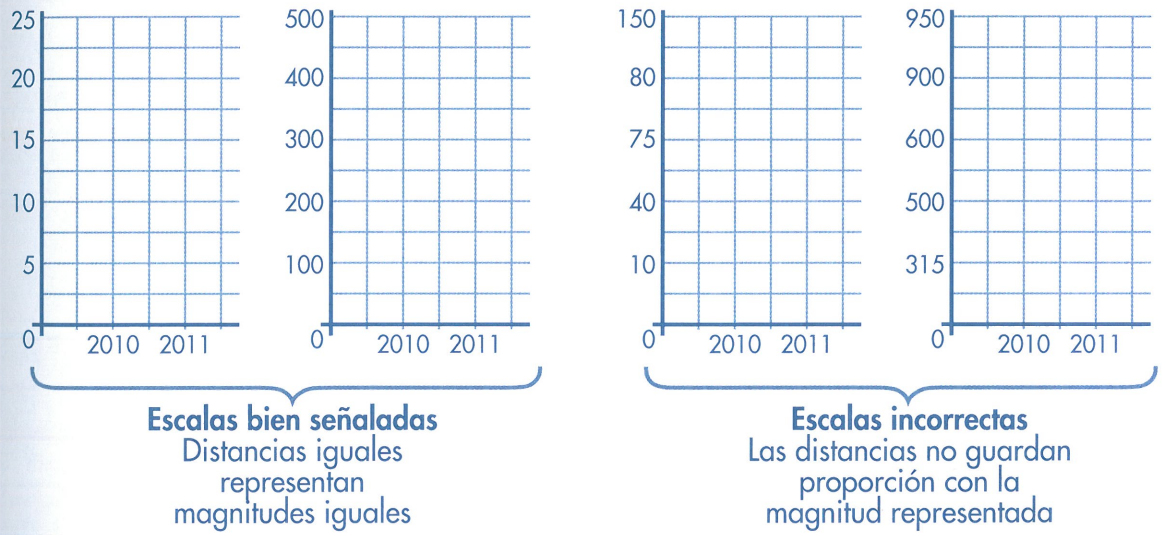


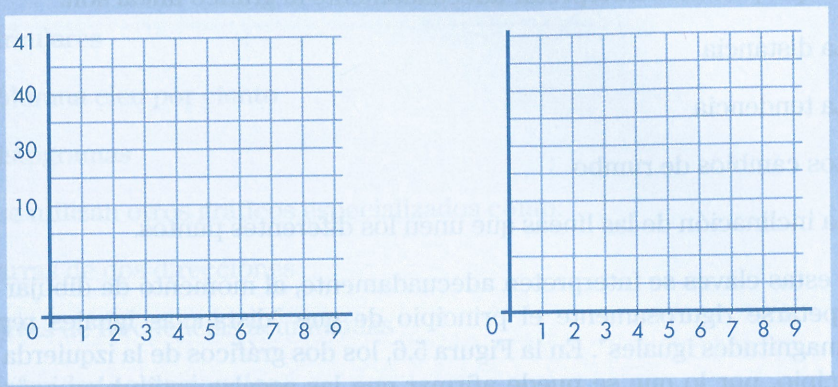
Figura 5.6. **Representación de la escala en el eje Y**

Fuente: Elaboración propia con fines didácticos

EJERCICIO 5.2

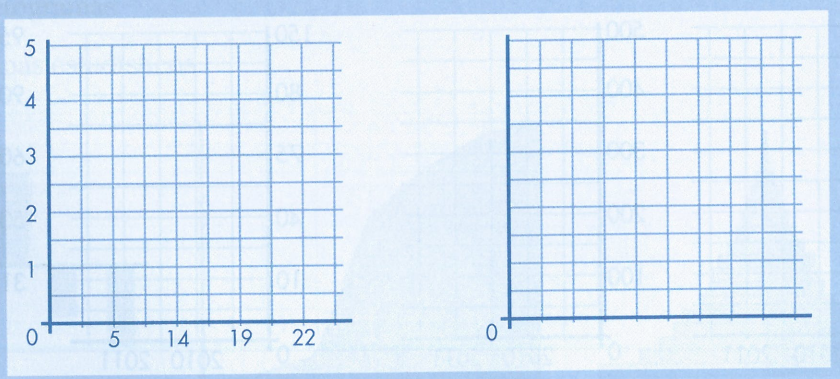
Corrija los defectos de la escala que se pretende utilizar para los siguientes gráficos:

a)

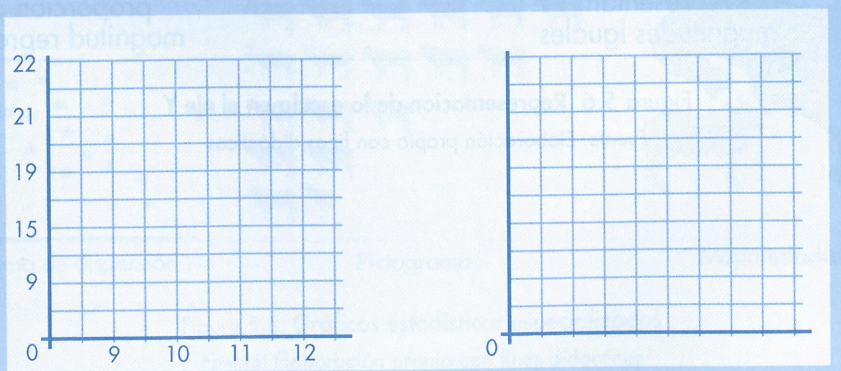


:

b)



c)



En el boletín "Indicadores Básicos de Salud 2010", del Ministerio de Salud, se informa que el número de homicidios desde el año 2000 hasta el 2009 en Costa Rica se comportaba según el siguiente cuadro:

Cuadro 5.1. **Mortalidad por homicidios en Costa Rica de acuerdo con Informes de Medicatura Forense, de 2000 a 2009**

Años	Muertes
2000	241
2001	206
2002	236
2003	285
2004	251
2005	310
2006	318
2007	333
2008	483
2009	525*

* Cifras Preliminares

Fuente: INEC-Dirección de Vigilancia de la Salud – Ministerio de Salud (2011)

A partir de esos datos, se construye un gráfico lineal para que informe al gran público del comportamiento de los homicidios en ese espacio temporal, así:



Figura 5.7. **Mortalidad por homicidios en Costa Rica de acuerdo con Informes de Medicatura Forense, de 2000 a 2009**

Fuente: INEC-Dirección de Vigilancia de la Salud – Ministerio de Salud (2011)

Puesto que la inclinación de las líneas que unen los diferentes puntos ayuda a la interpretación, es indispensable que el área que ocupa el cuerpo del gráfico no sobrepase una relación base por altura de 1 a 1,5. Observe lo que ocurre cuando esta regla se irrespeta:

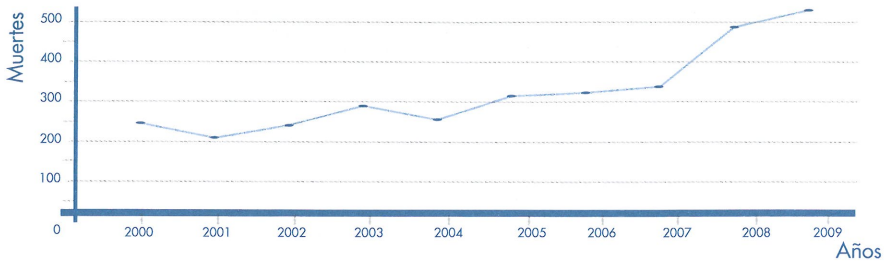


Figura 5.8. **Mortalidad por homicidios en Costa Rica de acuerdo con Informes de Medicatura Forense, de 2000 a 2009**

Si el gráfico es muy ancho en la base y corto en la altura, las tendencias se amortiguan y parecen menores.

En caso contrario, si la base del diagrama es pequeña y la altura es grande, entonces las tendencias se exageran:

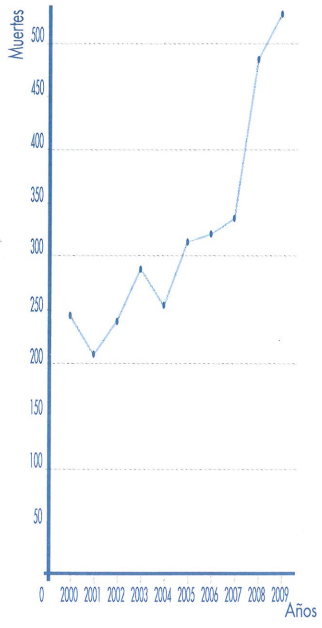


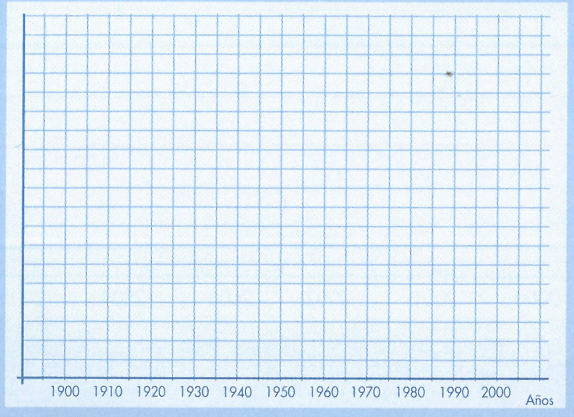
Figura 5.9. **Mortalidad por homicidios en Costa Rica de acuerdo con Informes de Medicatura Forense, de 2000 a 2009**

EJERCICIO 5.3

a) Los datos de esperanza de vida al nacimiento del cuadro siguiente han sido tomados del trabajo *Evolución de la Esperanza de Vida al Nacimiento en Costa Rica 1900–2000,* realizado por las investigadoras Floribel Méndez Fonseca y Olga Martha Araya Umaña para el INEC y publicado el año 2003. A partir de esos datos, confeccione un gráfico lineal que muestre la tendencia durante el siglo XX.

Año	Esperanza de vida al nacimiento
1900	34,7
1910	35,1
1920	35,1
1930	42,23
1940	46,93
1950	55,59
1960	62,6
1970	65,48
1980	74,18
1990	76,7
2000	77,4

Fuente: Méndez y Araya (2003)



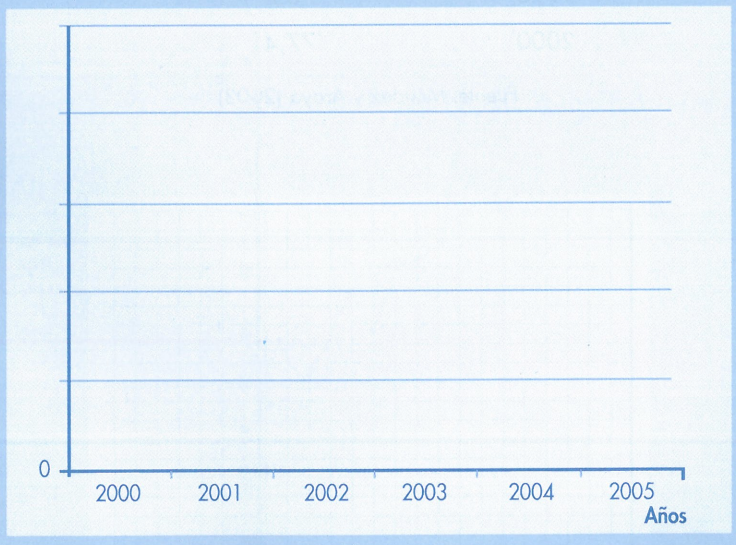
b) El Centro Centroamericano de Población calcula la tasa de crecimiento natural de Costa Rica, datos que se presentan seguidamente. A partir de ese informe, elabore una representación lineal que contenga los elementos esenciales de un gráfico.

Tasa de crecimiento natural de Costa Rica, 1960-2005

Año	Tasa de crecimiento natural (1000s)
1960	38,24
1965	33,71
1970	25,39
1975	24,30
1980	26,69
1985	27,98
1990	23,36
1995	19,14
2000	16,30
2005	13,18

Fuente: Centro Centroamericano del Población (2002)

Resumido para propósitos didácticos



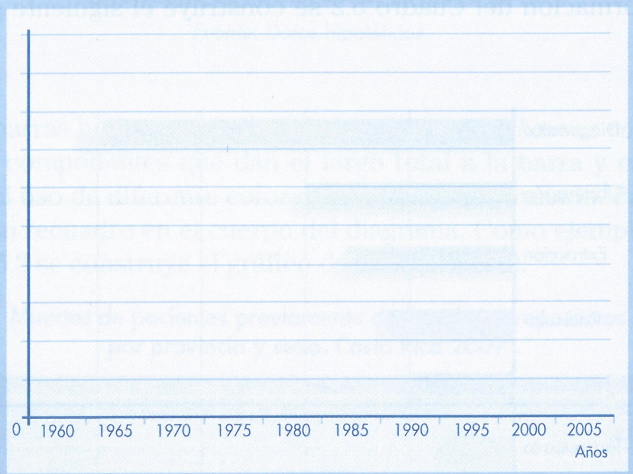
c) Con base en los siguientes datos del Centro Centroamericano de Población sobre tasas de natalidad en Costa Rica, de 1960 a 2005, elabore un gráfico lineal.

Tasas brutas de natalidad y mortalidad en Costa Rica (1960-2005)

Año	Tasa bruta de natalidad (1000s)
1960	46,89
1965	41,99
1970	32,71
1975	29,25
1980	30,87
1985	31,96
1990	27,09
1995	23,21
2000	20,14
2005	16,94

Fuente: Centro Centroamericano de Población (2002)

Resumido para propósitos didácticos



5.3.2. GRÁFICOS DE BARRAS HORIZONTALES

Este tipo de gráfico se utiliza para representar series cualitativas y series geográficas, dada la posibilidad de escribir el nombre completo de las categorías a lo largo sin tener que recurrir a abreviaturas ni a códigos, que podrían hacer confusa la lectura o interpretación del diagrama. La información se expresa por la longitud de las barras y no por el área ocupada ni por el ancho de clase. A manera de ejemplo, tómense los siguientes datos hipotéticos:

Cuadro 5.2. **Número de procedimientos odontológicos realizados a pacientes de la Clínica Universitaria de Acción Social en mayo 2012**

Procedimiento	Pacientes
Diagnóstico	199
Obturación	100
Extracción	77
Radicular	60
Placa bacteriana	50
Tartrectomía	39
Ortodoncia	15
TOTAL	540

Fuente: Datos hipotéticos

A partir de la información del Cuadro 5.2 se construye el siguiente gráfico de barras horizontales:

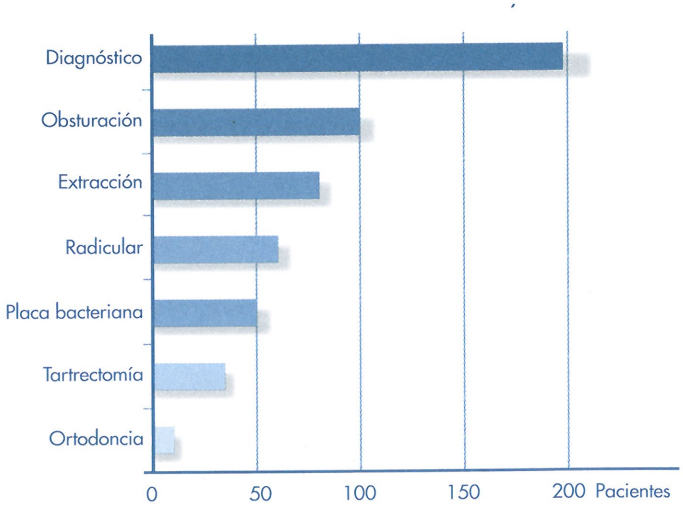


Figura 5.10. **Número de Procedimientos odontológicos realizados a pacientes de la Clínica Universitaria de Acción Social en mayo 2012**

Fuente: Datos hipotéticos

Se recomienda que las barras se ubiquen en orden de magnitud, de mayor a menor para mostrar con claridad cuáles categorías tienen frecuencias mayores y cuáles menores.

5.3.3. GRÁFICO DE BASTONES

Cuando las categorías son muchas, se pueden presentar solamente las más frecuentes. También es posible disminuir el ancho de las barras hasta convertir la ilustración en un gráfico de bastones, como el siguiente:

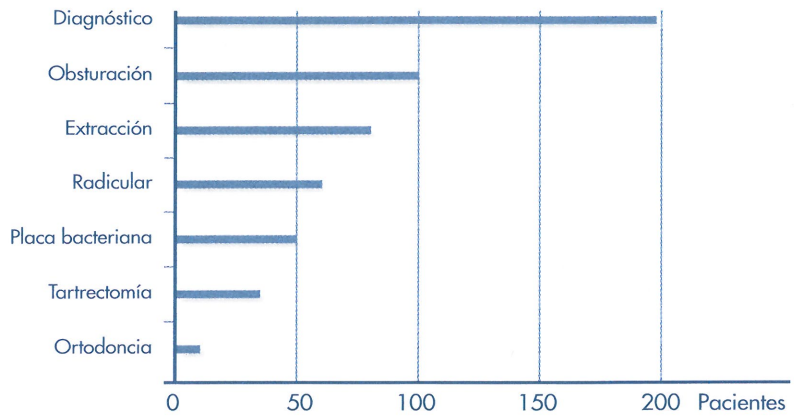


Figura 5.11. **Número de Procedimientos odontológicos realizados a pacientes de la Clínica Universitaria de Acción Social en mayo 2012**

Fuente: Datos hipotéticos

Otra variedad de barras horizontales se usa en el gráfico de *barras compuestas*. Estas tienen dos o más componentes que dan el largo total a la barra y cada categoría se individualiza por el uso de diferente color o distinto tipo de rallado. Para identificarlas mejor, se coloca un recuadro en el cuerpo del diagrama. Como ejemplo, a partir de los datos del Cuadro 5.3 se construye el gráfico de la Figura 5.12.

Cuadro 5.3. **Muertes de pacientes previamente declarados seropositivos (HIV-SIDA) por provincia y sexo, Costa Rica 2009**

	Hombres	Mujeres	Total
San José	47	14	61
Alajuela	23	8	31
Cartago	20	4	24
Puntarenas	16	7	23
Limón	13	4	17
Heredia	8	7	15
Guanacaste	10	2	12
TOTAL	137	46	183

Fuente: Datos hipotéticos

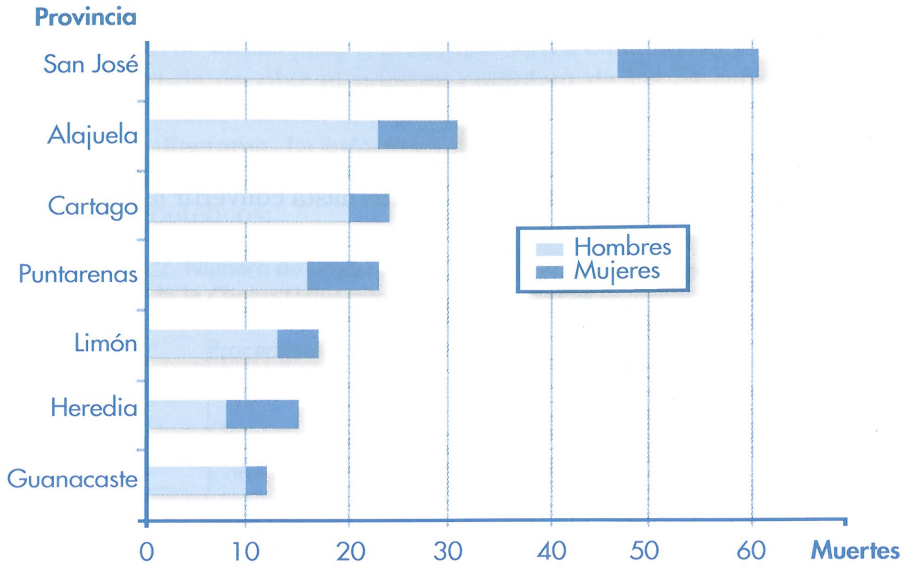


Figura 5.12. **Muertes de pacientes previamente declarados seropositivos (HIV-SIDA) por provincia y sexo, Costa Rica 2009**

Fuente: Datos hipotéticos

Los gráficos de barras compuestas permiten observar la importancia que el fenómeno de interés tiene en cada provincia o zona geográfica, pero no facilita la comparación entre sexos. Esta limitación puede subsanarse utilizando el gráfico de barras comparativas.

En el gráfico de *barras comparativas* se separan los subgrupos de cada categoría. Esto permite al lector hacer las comparaciones en forma rápida, aunque se limita la apreciación del total por categoría.

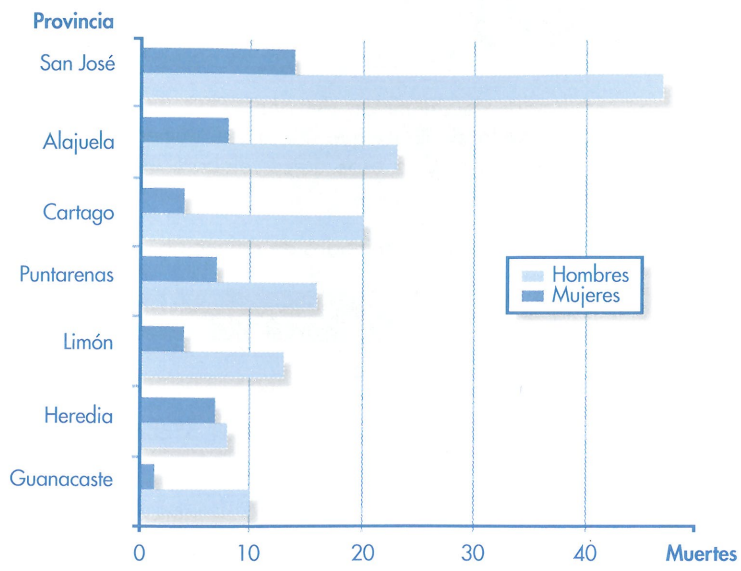


Figura 5.13. **Muertes de pacientes previamente declarados seropositivos (HIV-SIDA) por provincia y sexo, Costa Rica, 2009**

Fuente: Datos hipotéticos

Los gráficos de barras compuestas y los de barras comparativas no necesariamente compiten entre sí, antes bien, se complementan o pueden usarse alternamente aprovechando las ventajas comunicativas de unos y otros. Si se busca mostrar los totales por categoría, será más adecuado el gráfico de barras compuestas; en tanto que si interesa comparar las categorías, será de mayor utilidad el gráfico de barras comparativas.

Para ambos tipos, se recomienda no utilizar más de tres componentes puesto que cargan al gráfico y dificultan la comparación.

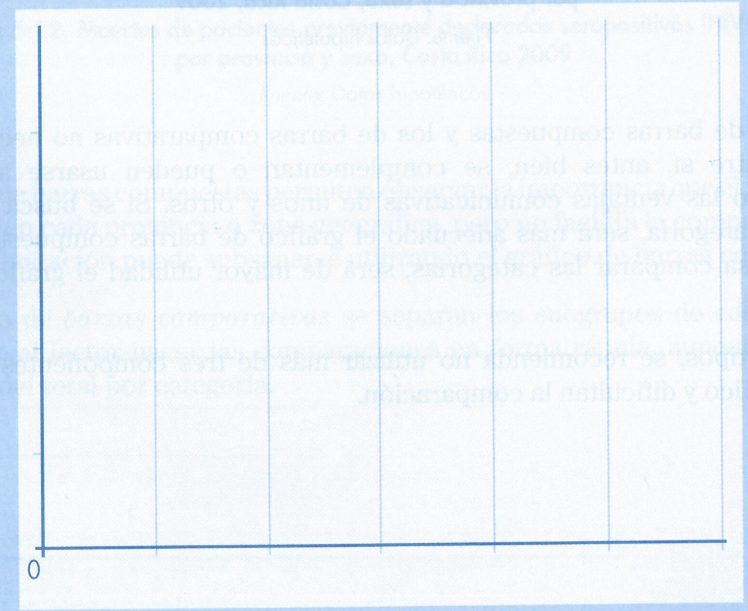
EJERCICIO 5.4

a) Con base en los certificados de defunción del año 2003 en El Salvador, Méndez y otros (2006) presentan la tabla siguiente que informa el número de defunciones según las principales causas de muerte. Con esos datos, elabore un gráfico explicativo del fenómeno.

Principales causas de muerte según certificados de defunción, El Salvador 2003

	Causa de muerte	Muertes
1	Enfermedades del sistema circulatorio	5 308
2	Causa externa	4 973
3	Sin clasificar	3 913
4	Tumores	3 229
5	Enfermedades del sistema genitourinario	1 647

Fuente: Méndez y otros (2006)

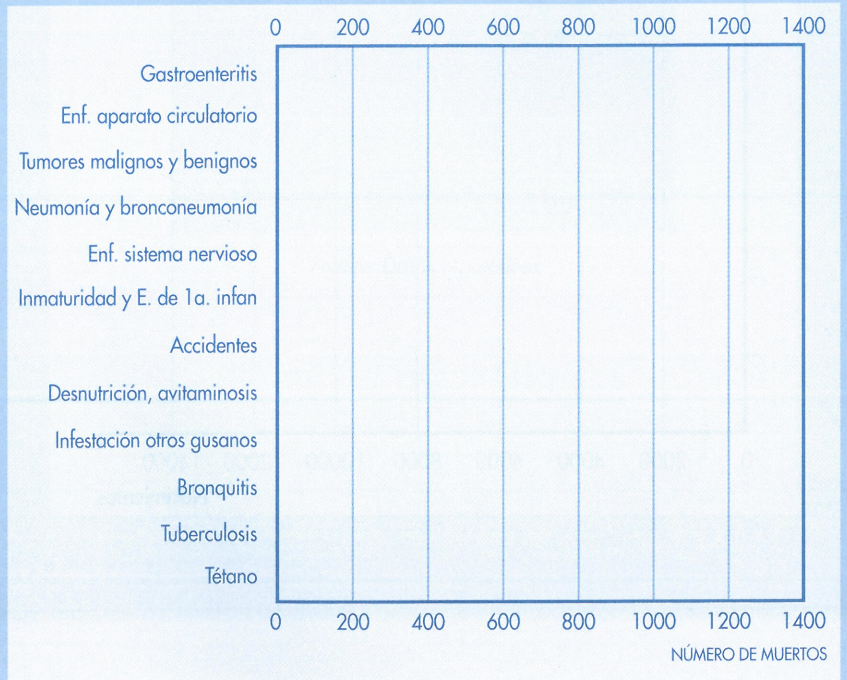


b) Según datos del Centro Centroamericano de Población, en Costa Rica, durante 1956, las causas de muerte tanto con asistencia médica como sin ella, fueron las siguientes:

	Con asistencia médica	Sin asistencia médica	Número de Muertes
Gastroenteritis	680	590	1270
Enfermedades del aparato circulatorio	600	203	803
Tumores malignos y benignos	679	115	794
Neumonía y bronconeumonía	396	195	591
Enfermedades del sistema nervioso	382	63	445
Inmaturidad y E. de la 1a infancia	372	25	397
Accidentes	148	223	371
Desnutrición, avitaminosis	232	136	368
Infestación otros gusanos	70	229	299
Bronquitis	40	160	200
Tuberculosis	179	21	200
Tétano	80	118	198

Fuente: <http://ccp.ucr.ac.cr/bvp/pdf/anuariocr/an1956/anuario-56_03.pdf>

Con esos datos, elabore un gráfico de barras compuestas.

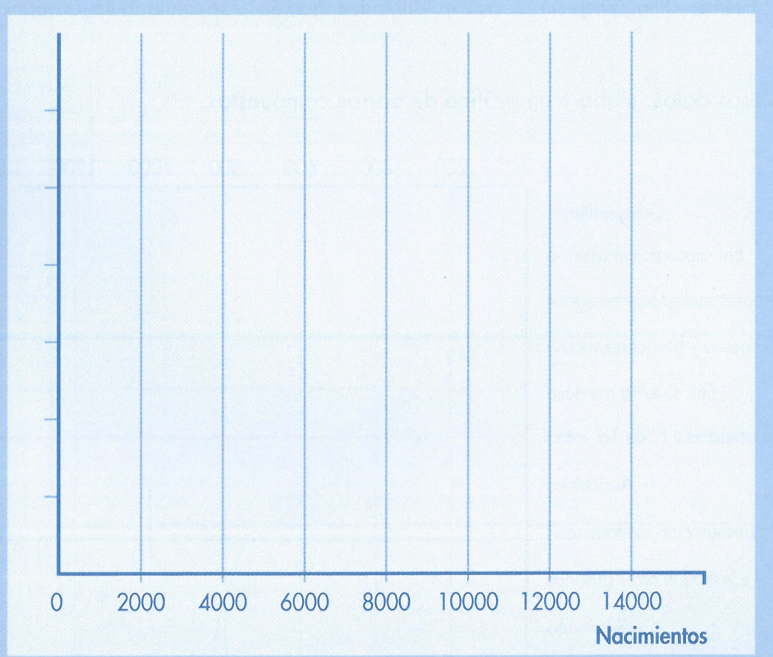


c) Elabore un gráfico de barras comparativas a partir del informe sobre nacimientos dentro y fuera del matrimonio, por provincias, durante el año 2010, realizado por el Instituto Nacional de Estadísticas y Censos (INEC).

Nacimientos inscritos dentro y fuera del matrimonio por provincias de Costa Rica en el año 2010

PROVINCIA	Dentro del matrimonio	Fuera del matrimonio	Total de nacimientos
San José	7 692	14 726	22 418
Alajuela	5 038	9 300	14 338
Cartago	3 025	4 180	7 205
Heredia	2 452	3 954	6 406
Guanacaste	1 325	4 375	5 700
Puntarenas	1 832	5 396	7 228
Limón	1 717	5 910	7 627
			70 922

Fuente: INEC (2011)



5.3.4. LOS GRÁFICOS DE COLUMNAS

El uso de columnas o barras verticales es apropiado tanto para series cronológicas como para variables cuantitativas discretas o categorizadas por intervalos. Al igual que los gráficos de barras horizontales, lo que informa es la longitud de las barras y no su ancho. En el caso de las columnas, estas no tienen que ordenarse por magnitud sino conforme la secuencia lógica del tiempo o del aumento en la variable cuantitativa.

Dependiendo del tipo de datos que se quiera representar, la columnas también pueden ser simples (una categoría), compuestas y comparativas (dos o más categorías).

Por ejemplo, se tiene la siguiente información sobre el número de hijos de un grupo:

Cuadro 5.4. **Número de hijos de los estudiantes del Posgrado en Registros Médicos UNED, Segundo Ciclo 2015**

CLASES N. ° de hijos	FRECUENCIAS ABSOLUTAS
0	1
1	2
2	7
3	8
4	4
5	2
6	1
	25

Fuente: Datos hipotéticos

Para confeccionar el gráfico de columnas, en el eje de las X se ubica el número de hijos. Observe que, en este caso, existen estudiantes que no tienen hijos, de forma que el cero no es equivalente al origen, sino que es una información acerca del número de hijos. En el eje Y va la escala de las frecuencias absolutas, de manera que si 7 alumnos tienen 2 hijos, a partir del 2 del eje X se dibujará una columna de alto 7, resultando el siguiente gráfico:

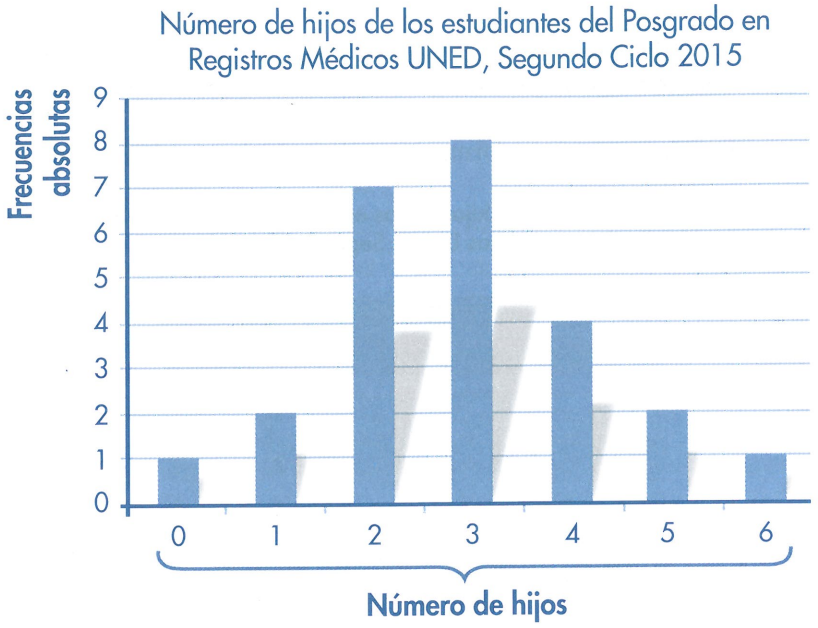


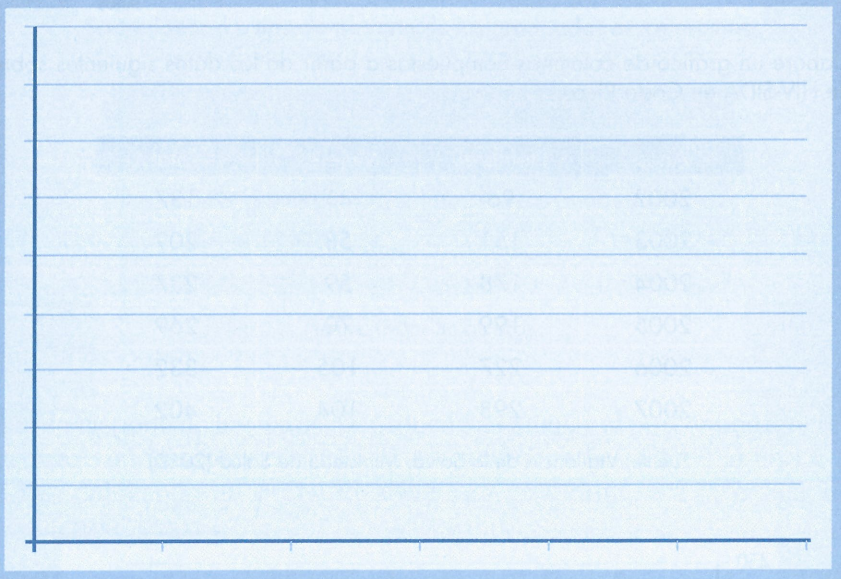
Figura 5.14. **Número de hijos de los estudiantes del Posgrado en Registros Médicos UNED, Segundo Ciclo 2015**

EJERCICIO 5.5

a) Elabore un gráfico de columnas a partir de los siguientes datos sobre defunciones infantiles en Costa Rica.

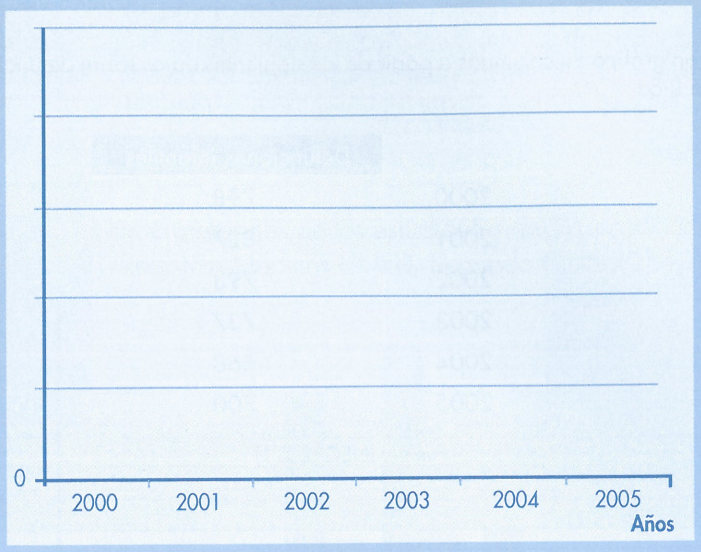
	Defunciones infantiles
2000	798
2001	827
2002	793
2003	737
2004	668
2005	700

Fuente: INEC (2011)



b) A partir de los siguientes datos hipotéticos sobre suicidios de hombres y de mujeres, elabore un gráfico de columnas comparativas.

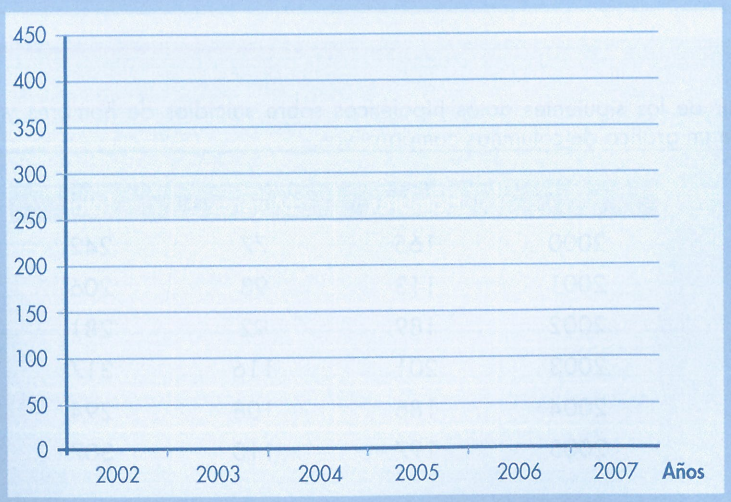
	Hombres	Mujeres	Total
2000	165	77	242
2001	113	93	206
2002	189	92	281
2003	201	116	317
2004	186	108	294
2005	197	110	307



c) Elabore un gráfico de columnas compuestas a partir de los datos siguientes sobre casos de HIV-SIDA en Costa Rica.

Años	Hombres	Mujeres	Total
2002	96	41	137
2003	151	58	209
2004	178	59	237
2005	199	70	269
2006	227	105	332
2007	298	104	402

Fuente: Vigilancia de la Salud, Ministerio de Salud (2010)



5.3.5. REPRESENTACIÓN DE LA PARTICIPACIÓN RELATIVA

Cuando lo que interesa representar es la participación porcentual de las categorías con respecto del total, se puede utilizar:

- La barra 100 %
- El gráfico circular

La *barra 100 %* se compone de un rectángulo suficientemente ancho cuya altura constituye el 100%, luego se marcan las diferentes áreas para indicar los porcentajes que cada clase representa del total. En este tipo de gráfico, lo que se destaca es la importancia relativa de cada una de las categorías y no el valor absoluto de cada una de ellas.

Si, por ejemplo, se cuenta con los siguientes datos hipotéticos acerca de la ocupación principal de un grupo de graduados:

Cuadro 5.5

Sector laboral o tipo de patrono de los graduados de Enfermería (Promociones desde el año 2000 hasta el 2010)

CCSS	195	65%
Cuenta propia	60	20%
Sector privado	30	10%
Desempleado	15	5%
	300	100%

Fuente: Datos hipotéticos

Una vez que se dispone de los datos, se procede a calcular el porcentaje que representa cada categoría. Aquella con el porcentaje mayor irá en la base de la barra, luego las demás se van colocando en orden de magnitud, la última será la menor o, cuando existe, la categoría "otros".

Las convenciones para la construcción de este tipo de gráfico son de sentido común; por ejemplo, la barra no tiene que ser muy alta ni muy baja, tampoco muy ancha ni muy delgada, de forma que comunique adecuadamente la información. Adicionalmente, puede incluirse un recuadro explicativo de las categorías o rotularse sobre el gráfico, también pueden incluirse los porcentajes o los datos absolutos. En general se puede decir que la barra 100 % tolera la creatividad, siempre que se respeten las proporciones de las áreas.

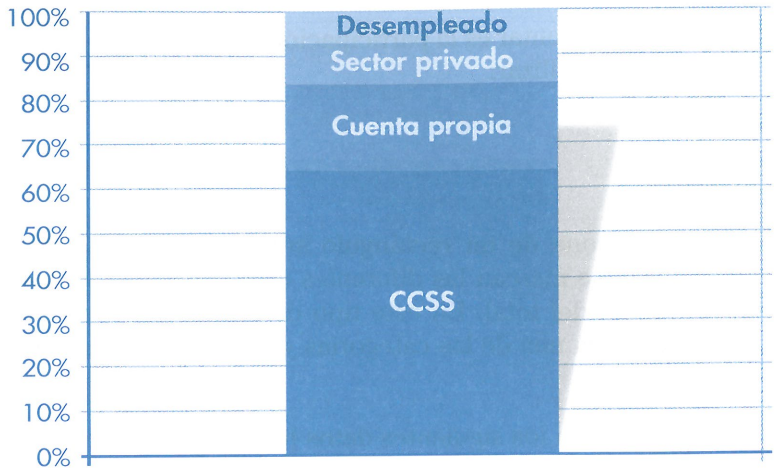


Figura 5.15. **Sector laboral o tipo de patrono de los graduados de Enfermería (Promociones desde el año 2000 hasta el 2010)**

Fuente: Datos hipotéticos

Los *gráficos circulares* también informan la composición porcentual de cada una de las categorías. Al igual que en la barra 100 %, es el área de cada una de las secciones la que indica la participación relativa de la categoría. Por ejemplo:

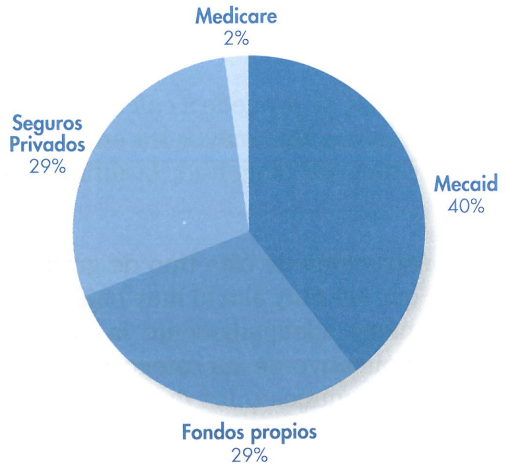


Figura 5.16. **¿Quién paga por los servicios médicos generados por el SIDA en EE. UU.? (Agosto de 1991, con base en datos de la U.S. National Commission on AIDS)**

Fuente: Reproducido de Mann (1992)

Gracias a los múltiples programas informáticos disponibles en el mercado, la confección de este tipo de gráfico no presenta grandes dificultades técnicas y se puede seccionar o separar una porción para destacar su participación relativa, hacerlo en tres dimensiones y manipular los colores.

Si el gráfico tuviera que hacerse en forma manual, el porcentaje de cada categoría se tiene que multiplicar por los 360 grados de un círculo y luego, con un transportador, marcar los grados que ocupa cada clase, comenzando por la mayor. Posteriormente, siguiendo el sentido de las agujas del reloj, se van agregando las otras clases en orden de mayor a menor, dejando de último la clase "otros", si existiere.

Como ejemplo, se utilizarán los mismos porcentajes del gráfico de barra 100 %. Estos de multiplican por 360 grados, con lo que resulta:

Cuadro 5.6

Sector laboral o tipo de patrono de los graduados de Enfermería (Promociones desde el año 2000 hasta el 2010)

CCSS	0,65 X 360	234 grados
Cuenta propia	0,20 X 360	72 grados
Sector privado	0,10 X 360	36 grados
Desempleado	0,05 X 360	18 grados
		360 grados

Fuente: Datos hipotéticos

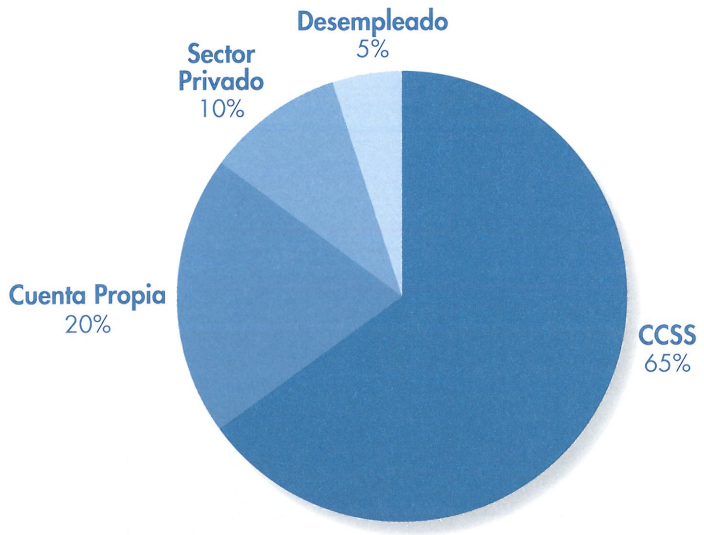


Figura 5.17. **Sector laboral o tipo de patrono de los graduados de Enfermería (Promociones desde el año 2000 hasta el 2010)**

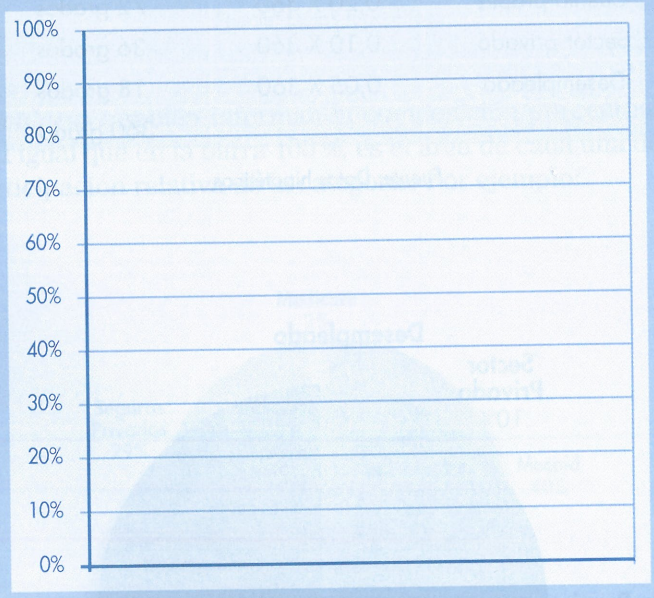
EJERCICIO 5.6

a) Represente los datos del cuadro siguiente en una barra 100 %.

Apertura de historias clínicas según servicios en el año 2011

TOTAL	1 495	100
Abdomen	378	25,3
Ginecología	339	22,7
Cabeza y cuello	308	20,6
Senos y tumores mixtos	262	17,5
Urología	162	10,8
Medicina oncológica	46	3,1

Fuente: IREN (2011)

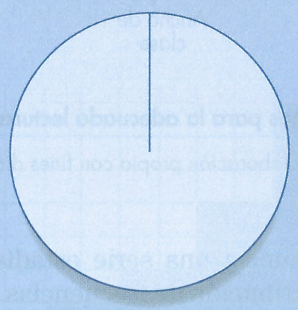


b) Basándose en los datos del cuadro siguiente, elabore un gráfico circular que informe la participación relativa de cada causa de muerte en la mortalidad infantil.

Mortalidad infantil según causa de muerte
(Costa Rica, 2010)

Causa de muerte	Muertes	Porcentaje	Grados
Afecciones en el periodo perinatal	330		
Malformaciones congénitas	234		
Enfermedades del sistema respiratorio	38		
Enfermedades del sistema circulatorio	14		
Enfermedades infecciosas o parasitarias	12		
Enfermedades del sistema nervioso	10		
Otras causas	33		
TOTAL	671		

Fuente: Ministerio de Salud, Dirección de Vigilancia de la Salud (2011)



5.3.6. EL HISTOGRAMA

El *histograma* sirve para representar una serie de variable cuantitativa continua. Este gráfico ubica en el eje X los límites reales de clase (o los puntos medios) y, para cada intervalo, en el eje Y informa la frecuencia de clase por medio del área de la columna o, si los anchos de clase fueran iguales, por su altura.

Puesto que la variable X es continua, las columnas van unidas y utilizan los códigos que se señalan en la figura siguiente:

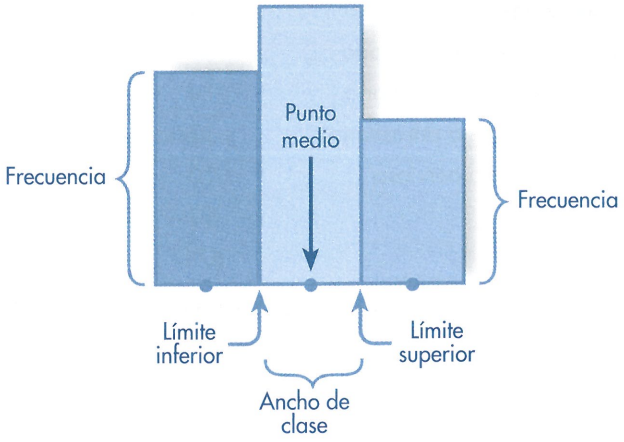


Figura 5.18. **Códigos para la adecuada lectura de un histograma**

Fuente: Elaboración propia con fines didácticos

Un histograma solamente comunica una serie estadística simple a la vez pues, en esencia, presenta una única distribución de frecuencias. Tiene la ventaja que permite la identificación de la simetría o el tipo de asimetría y del mayor o menor apuntamiento de la distribución.

Como ejemplo, tómese la siguiente información hipotética sobre las frecuencias de los pesos en gramos de 40 recién nacidos.

Cuadro 5.7

Distribución de frecuencias del peso en gramos de los recién nacidos Clínica de Turrucalito, agosto 2013

CLASES (peso en gramos)	FRECUENCIA
499,5 – 999,5	1
999,5 – 1 499,5	2
1 499,5 – 1 999,5	2
1 999,5 – 2 499,5	4
2 499,5 – 2 999,5	5
2 999,5 – 3 499,5	7
3 499,5 – 3 999,5	10
3 999,5 – 4 499,5	6
4 499,5 – 4 999,5	3
	40

Fuente: Datos hipotéticos

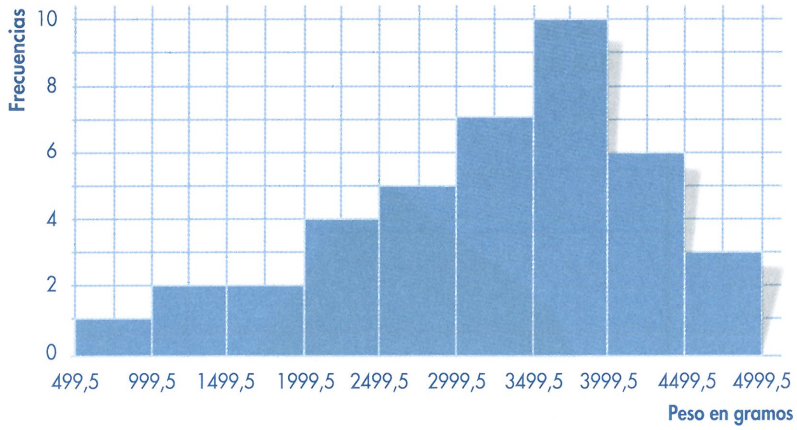


Figura 5.19. **Distribución de frecuencias del peso en gramos de recién nacidos en la Clínica de Turrucalito, agosto 2013**

Fuente: Datos hipotéticos

Nótese que en el eje X se han señalado los límites reales de las clases, pero también podrían haberse usado los puntos medios de clase.

5.3.7. EL POLÍGONO DE FRECUENCIAS

Sirve para representar variables continuas uniendo los puntos medios de cada clase con líneas rectas y cerrando en los extremos con frecuencia cero. Para poder cerrarlo, se agrega una semiclase al principio y otra al final como en la Figura 5.20.

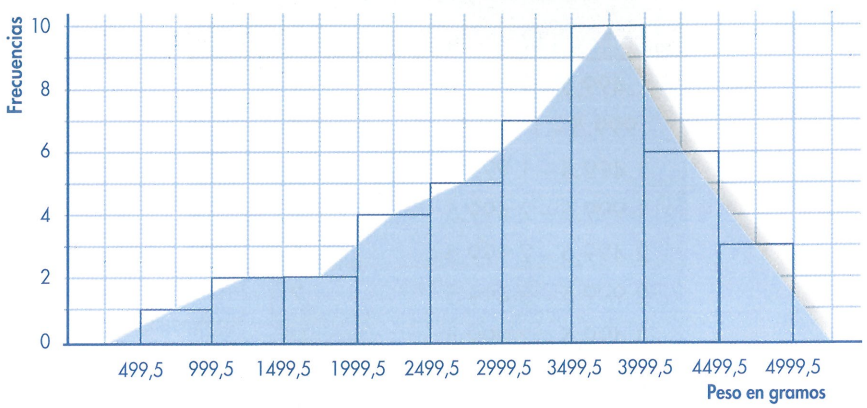


Figura 5.20. **Distribución de frecuencias del peso en gramos de recién nacidos en la Clínica de Turrucalito, agosto 2013**

Fuente: Datos hipotéticos

Una ventaja de los polígonos de frecuencia es que posibilita comparar la distribución de la misma variable para dos o más conjuntos:

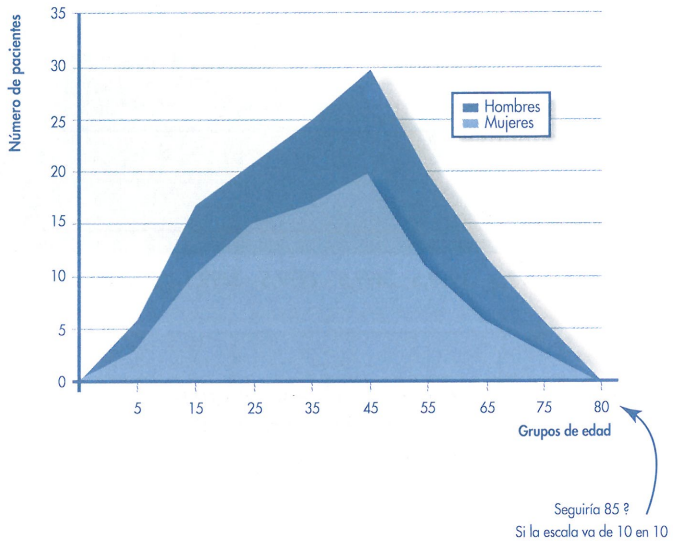


Figura 5.21. **Pacientes atendidos por fractura en extremidad superior por sexo y grupos de edad, 2012**

Fuente: Datos hipotéticos

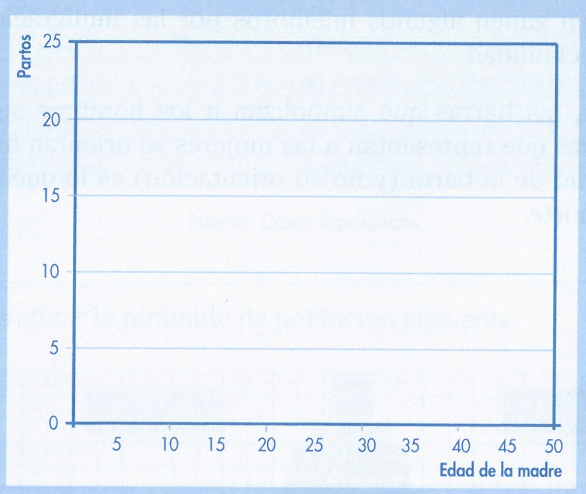
EJERCICIO 5.7

Grafique un histograma y un polígono de frecuencias a partir de los datos que se presentan seguidamente, los cuales contienen información sobre partos atendidos y registrados según edad de la madre.

Edad de la madre atendida por parto en la Clínica de la Mujer, setiembre 2016

Límite inferior	Límite superior	Punto medio	Frecuencia
9,5	14,5	12	3
14,5	19,5	17	8
19,5	24,5	22	15
24,5	29,5	27	20
29,5	34,5	32	22
34,5	39,5	37	17
39,5	44,5	42	10
44,5	49,5	47	5

Fuente: Datos hipotéticos



5.3.8. LA PIRÁMIDE DE POBLACIÓN

La pirámide de población es una variedad de histograma que expresa gráficamente la constitución por edad y sexo de los habitantes de un país.

Este tipo de histograma permite una fácil y rápida percepción de algunos fenómenos demográficos tales como:

- La composición por sexo
- El posible envejecimiento de una sociedad
- El tamaño de sus diferentes cohortes
- Los efectos de guerras, plagas y otro tipo de desastres sobre la población

Las pirámides de población más comunes organizan los segmentos de la sociedad en cohortes con ancho de clase de 5 años que, partiendo de un eje central, se representan en forma de barras horizontales. También es posible elaborar pirámides para cohortes de un año, siempre que se cuente con información anual.

En la base de la pirámide se sitúan las cohortes de menor edad y las siguientes cohortes van ascendiendo hasta llegar a las de mayor edad. Como es de esperar que por muertes o emigración cada cohorte pierda integrantes conforme el paso del tiempo, las barras superiores tenderán a ser de menor tamaño que las de la base, aunque es posible que las cohortes también ganen algunos miembros por las inmigraciones o que se den reducciones en la fecundidad.

Convencionalmente, las barras que simbolizan a los hombres se orientan hacia la izquierda, y las barras que representan a las mujeres se orientan hacia la derecha. En todo caso, la longitud de la barra (y no su orientación) es lo que informa acerca del tamaño de cada cohorte.

Analice el siguiente ejemplo. Un país vecino cuenta con los siguientes datos de población:

Cuadro 5.8

Población del país X, ordenada según sexo y grupo de edad, 2010

Grupos de edad	Hombres	Mujeres
00-04 años	699 345	685 256
05-09	540 876	534 124
10-14	339 452	342 876
15-19	299 739	302 351
20-24	234 986	284 743
25-29	256 892	267 493
30-34	203 451	209 564
35-39	199 789	201 458
40-44	175 321	182 453
45-49	153 475	155 532
50-54	141 987	143 562
55-59	123 451	126 632
60-64	109 675	113 453
65-69	81 523	82 345
70-74	45 234	49 321
75 y más	44 867	49 314

Fuente: Datos hipotéticos

Estos permiten graficar la pirámide de población siguiente:

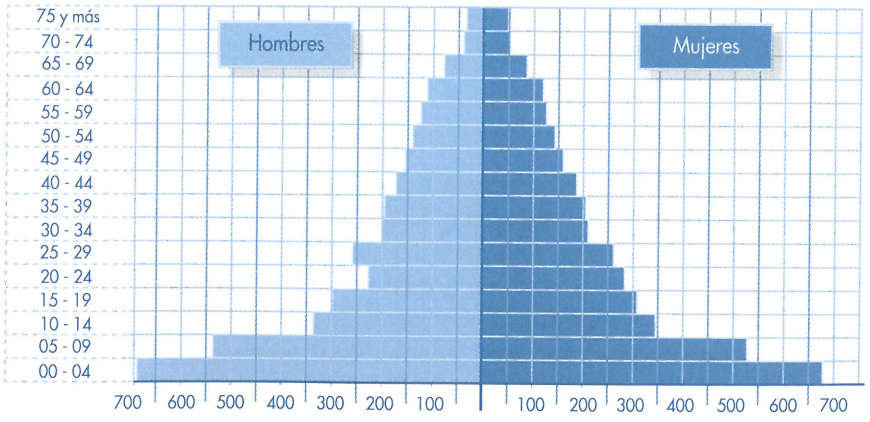


Figura 5.22. **Pirámide de población del país X para el año 2010**

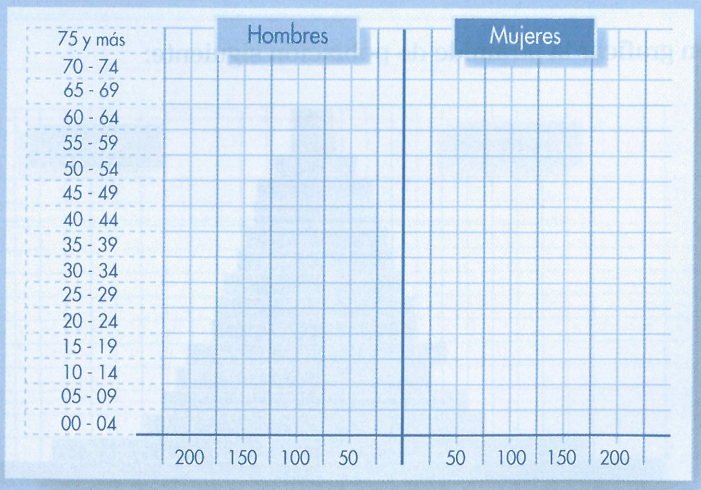
EJERCICIO 5.8

El país Y cuenta con los siguientes datos de población para el año 2010.

Grupos de edad	Hombres	Mujeres
00-04 años	99 453	96 754
05-09	102 987	100 876
10-14	129 235	129 743
15-19	196 732	198 987
20-24	192 341	194 234
25-29	191 387	193 429
30-34	193 421	195 432
35-39	191 789	193 458
40-44	176 754	181 453
45-49	173 475	175 532
50-54	151 987	153 562
55-59	133 451	136 632
60-64	129 675	133 453
65-69	121 765	124 897
70-74	113 786	117 342
75 y más	96 543	99 876

Fuente: Datos hipotéticos

Apoyado en esos datos, dibuje la pirámide de población para esa sociedad.



En forma similar a las pirámides de población, pueden elaborarse otros histogramas para informar fenómenos de cohortes humanas. Por ejemplo, la pirámide invertida de tasas de mortalidad específica por sexo y rango de edad elaborada por investigadores de la Caja Costarricense del Seguro Social, que se reproduce seguidamente:

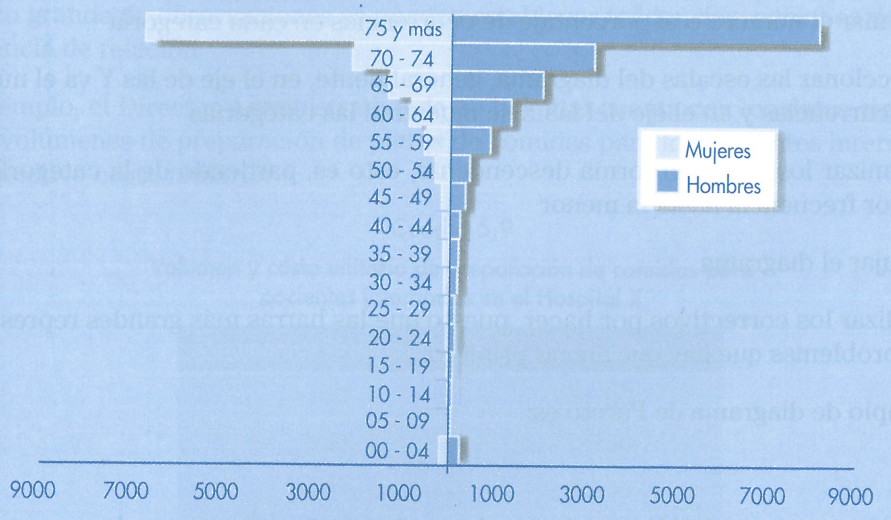


Figura 5.23. **Costa Rica: Defunciones por Cien Mil Habitantes. Según Grupos de Edad y Sexo, 2000 Tasas por 100,000 habitantes**

Fuente: CCSS (2003)

Obsérvese cómo la tasa de mortalidad crece conforme se aumenta en edad, excepto en los primeros años de vida cuando se es más vulnerable, razón por la cual las tasas de mortalidad infantil se calculan separadamente y con mayor detalle.

5.3.9. DIAGRAMA DE PARETO

Es una variación del histograma y se utiliza para clasificar los eventos (enfermedades, problemas, defectos) comenzando por el de mayor frecuencia y, en forma descendente, hasta los eventos menos significativos. El diagrama toma su nombre del economista Wilfredo Pareto quien al estudiar la situación de la economía italiana encontró que el 80 % de la riqueza estaba en manos del 20 % de la población, de allí que generalizó la regla del "80-20". En la década de 1950, Joseph Juran ratificó, en las fábricas japonesas, que el 80 % de las pérdidas se originaban en el 20 % de los problemas, y externó que se debía de centrar los esfuerzos correctivos, primero, en este 20 %. Luego la regla se amplió a muchas situaciones de control de calidad, errores de facturación, saturación hospitalaria, insatisfacción de los clientes y otras aplicaciones.

El diagrama de Pareto se elabora siguiendo estos pasos:

1. Seleccionar el objeto de estudio, por ejemplo, la insatisfacción de los pacientes con el servicio externo de un hospital
2. Recolectar los datos de acuerdo con las disconformidades detectadas (categorías)
3. Calcular el número o el porcentaje de ocurrencias en cada categoría
4. Seleccionar las escalas del diagrama, generalmente, en el eje de las Y va el número de ocurrencias y en el eje del las X se muestran las categorías
5. Organizar los datos en forma descendente, esto es, partiendo de la categoría con mayor frecuencia hasta la menor
6. Dibujar el diagrama
7. Analizar los correctivos por hacer, puesto que las barras más grandes representan los problemas que hay que atacar primero.

Un ejemplo de diagrama de Pareto es:

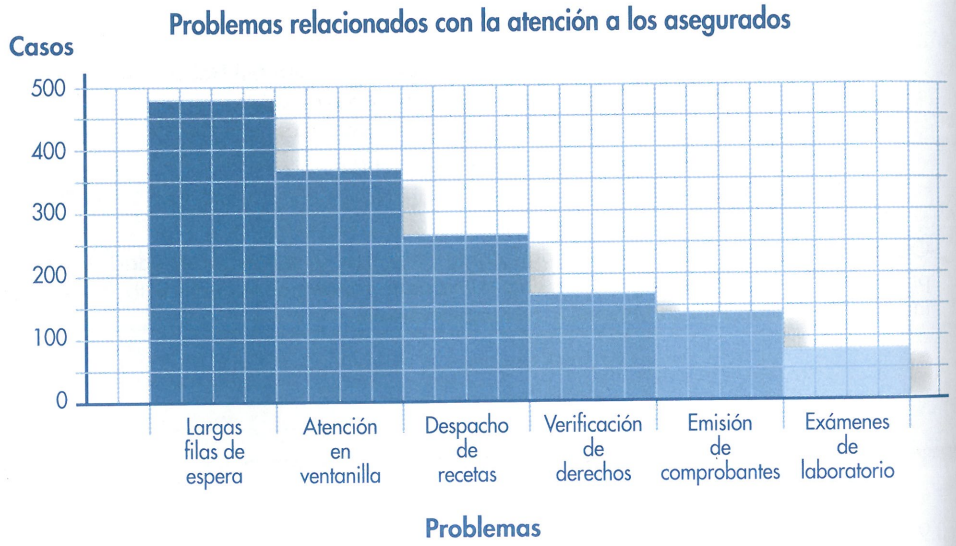


Figura 5.24. **Problemas relacionados con la atención a los asegurados**

Fuente: Datos hipotéticos

5.3.10. DIAGRAMA DE DISPERSIÓN

Es un gráfico que permite visualizar las posibles relaciones entre dos o más variables. A cada eje del plano cartesiano se le asigna una variable (estatura en X, peso en Y por ejemplo) y las observaciones se representan como puntos del plano conteniendo su información particular asociada con las variables (1,72 metros, 70 kilos). Al tener un número grande de observaciones se pueden establecer tendencias, patrones definidos o ausencia de relación.

Por ejemplo, el Director Administrativo de un hospital cuenta con los datos siguientes sobre volúmenes de preparación de platos de comidas para los pacientes internados y su respectivo costo unitario:

Cuadro 5.9

Volumen y costo unitario de preparación de comidas para pacientes internados en el Hospital X

Volumen de preparación	Costo por plato
300	3 100
350	2 900
400	2 800
450	1 900
500	1 800
550	1 700
600	1 600
650	1 500
700	1 800
750	1 700
800	1 400
850	1 500

Datos hipotéticos

Cada par ordenado lleva la información de volumen y costo generando el siguiente diagrama de dispersión:

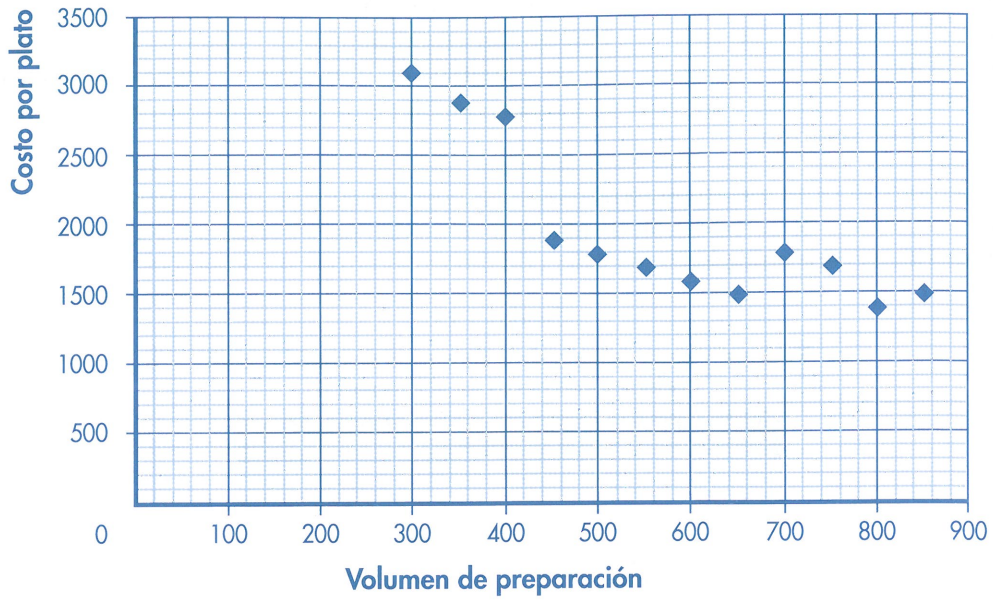


Figura 5.25. **Volumen y costo unitario de preparación de comidas para pacientes internados en el Hospital X**

La tendencia que se revela es que a mayor volumen de preparación, el costo unitario tiende a ser menor.

5.3.11. OJIVAS DE FRECUENCIAS ACUMULADAS

En el Capítulo 2 se estudiaron las frecuencias acumuladas y su forma de presentarlas en forma de cuadro, pues bien, las frecuencias acumuladas también pueden ser graficadas. La representación típica de las frecuencias acumuladas se llama *ojiva* y permite visualizar el número de observaciones que se encuentran por encima o por debajo de un valor determinado.

Las frecuencias de variable discreta pueden acumularse en los sentidos "X o menos" y "X o más". Si la variable es continua, se pueden acumular "al límite superior" o "al límite inferior".

Ejemplo, se tiene la siguiente distribución de frecuencias del número de hijos de un grupo de estudiantes que ya han sido acumuladas en el sentido "X o menos":

Cuadro 5.10.

Frecuencias acumuladas "X o menos" del número de hijos de los estudiantes del Posgrado en Registros Médicos UNED, Segundo Ciclo 2015

CLASES N. ° de hijos	FRECUENCIAS ACUMULADAS
0	1
1 o menos	3
2 o menos	10
3 o menos	18
4 o menos	22
5 o menos	24
6 o menos	25

Fuente: Datos hipotéticos

Con esos datos de variable discreta (puesto que el número de hijos es entero: 0, 1, 2, etc.) se procede a confeccionar el gráfico de la Figura 5.20, utilizando bastones del alto de la frecuencia acumulada:

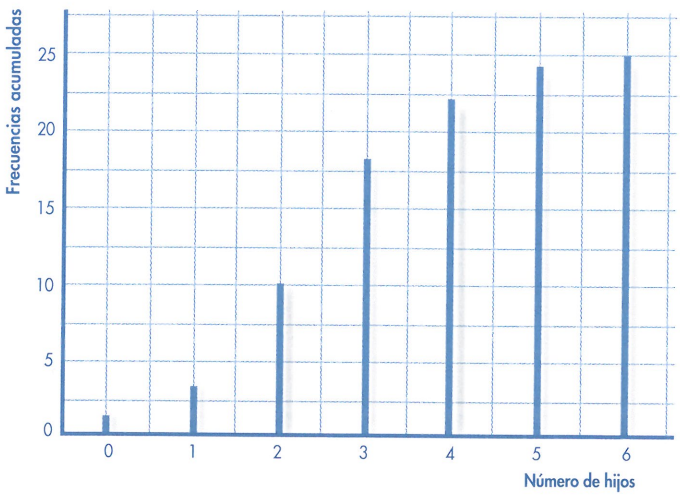


Figura 5.26. **Frecuencias acumuladas "X o menos" del número de hijos de los estudiantes del Posgrado en Registros Médicos UNED, Segundo Ciclo 2015**

La utilidad de la ojiva estriba en tener una información con solo un vistazo al gráfico, así, si interesa saber el número de participantes que tiene 2 hijos o menos, será sufuciente observar la altura del bastón que parte del 2, esto informa que 10 estudiantes tienen 2 o menos hijos.

Una variable continua como el peso en gramos, se acumula con refrencia a los límites de clase, así, las frecuencias del peso de los recién nacidos del Cuadro 5.10 han sido acumuladas al límite inferior:

Cuadro 5.11

Frecuencias absolutas y acumuladas al límite inferior del peso en gramos de los recién nacidos, Clínica de Turrucalito, agosto 2013

CLASES	FRECUENCIAS ABSOLUTAS	
(Peso en gramos)	Simples	Acumuladas
499,5 – 999,5	1	40
999,5 – 1 499,5	2	39
1 499,5 – 1 999,5	2	37
1 999,5 – 2 499,5	4	35
2 499,5 – 2 999,5	5	31
2 999,5 – 3 499,5	7	26
3 499,5 – 3 999,5	10	19
3 999,5 – 4 499,5	6	9
4 499,5 – 4 999,5	3	3
	40	

Fuente: Datos hipotéticos

Esa información genera la ojiva siguiente:

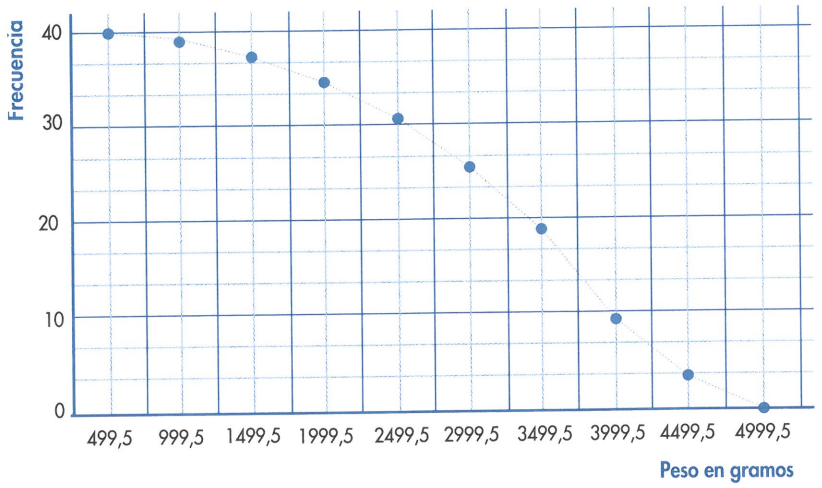


Figura 5.27. **Frecuencias acumuladas al límite inferior de clase del peso en gramos de los recién nacidos, Clínica de Turrucalito, agosto 2013**

Este gráfico permite leer que 31 recién nacidos pesan más de 2 499,5 gramos, donde 2,499,5 es el límite inferior a partir del cual se informa.

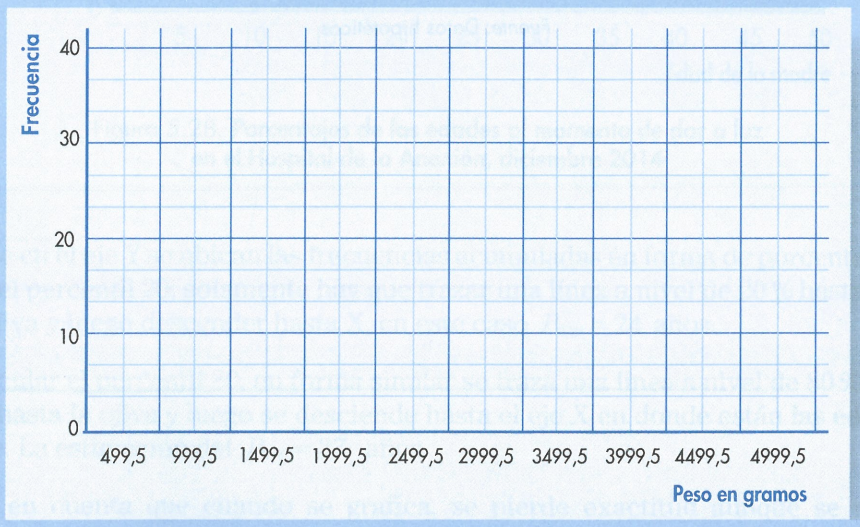
EJERCICIO 5.9

Con los datos de la tabla siguiente, calcule las frecuencias acumuladas "al límite superior" y dibuje la ojiva que se genera.

Frecuencias absolutas y acumuladas al límite superior del peso en gramos de los recién nacidos, Clínica de Turrucalito, agosto 2013

CLASES	FRECUENCIAS ABSOLUTAS	
(Peso en gramos)	Simples	Acumuladas
499,5 – 999,5	1	
999,5 – 1 499,5	2	
1 499,5 – 1 999,5	2	
1 999,5 – 2 499,5	4	
2 499,5 – 2 999,5	5	
2 999,5 – 3 499,5	7	
3 499,5 – 3 999,5	10	
3 999,5 – 4 499,5	6	
4 499,5 – 4 999,5	3	
	40	

Fuente: Datos hipotéticos



Estimación de percentiles a partir de ojivas

Para estimar percentiles en forma gráfica, es necesario transformar las frecuencias absolutas en frecuencias relativas; esto se obtiene con solo dividir las recuncias absolutas acumuladas entre el total.

Tómese como ejemplo el Cuadro 5.12 acerca de las edades de 50 mujeres al momento del parto.

Cuadro 5.12.

Edad en años cumplidos al momento de dar a luz en el Hospital de la Anexión, diciembre 2014

CLASES	FRECUENCIA ABSOLUTA	FRECUENCIA ACUMULADA	FRECUENCIA RELATIVA
(Edades)	fi	Fa	ACUMULADA
10 a menos de15	2	2	4 %
15 a menos de 20	4	6	12 %
20 a menos de 25	5	11	22 %
25 a menos de 30	12	23	46 %
30 a menos de 35	14	37	74 %
35 a menos de 40	9	46	92 %
40 a menos de 45	3	49	98 %
45 a menos de 50	1	50	100 %
	50		

Fuente: Datos hipotéticos

Observe que las frecuencias relativas se han acumulado "al límite superior" de cada clase, de modo que, a menos de 15 años está el 4 %, a menos de 20 años el 12 %, y así sucesivamente. Los pares ordenados del gráfico son: en el eje de las X se toma el límite superior y en el eje de las Y el porcentaje acumulado, resultando una ojiva como la siguiente:

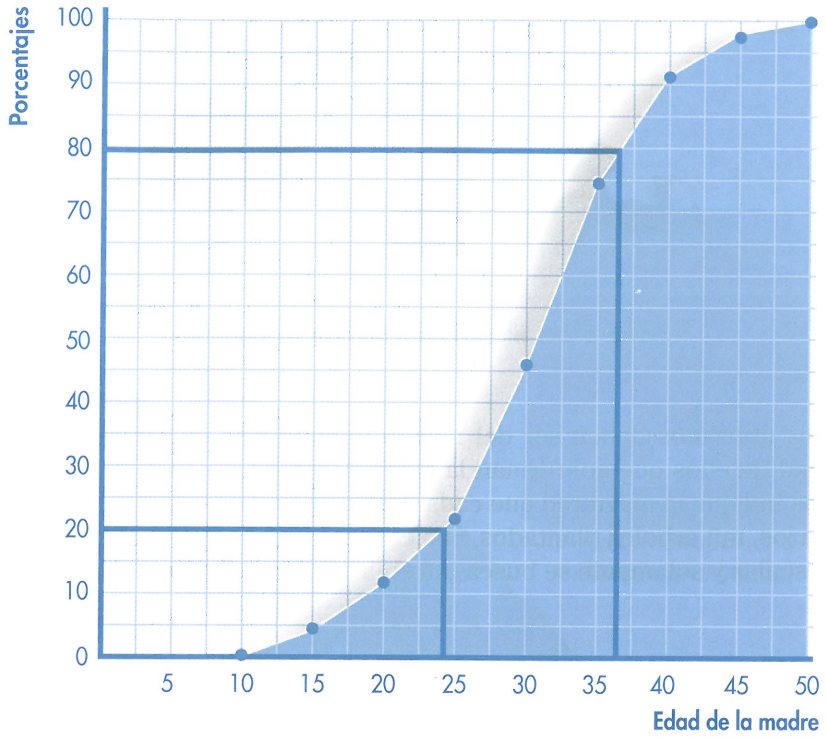


Figura 5.28. **Porcentajes de las edades al momento de dar a luz en el Hospital de la Anexión, diciembre 2014**

Dado que en el eje Y se ubican las frecuencias acumuladas en forma de porcentaje, para estimar el percentil 20, solamente hay que trazar una línea a nivel de 20 % hasta chocar con la ojiva y luego descender hasta X, en este caso $P_{0,20} = 24$ años.

Para calcular el percentil 80, en forma similar se traza una línea a nivel de 80 %, la cual se lleva hasta la ojiva y luego se desciende hasta el eje X en donde están las edades de la madre. La estimación del $P_{0,80} = 37$ años.

Tómese en cuenta que cuando se grafica, se pierde exactitud aunque se gana en comunicación. Ya en el capítulo IV se habían calculado estos percentiles y los resultados fueron 24 y 36,67 años respectivamente.

En todo caso, la aproximación es buena para tener una noción general de que, entre 24 y 37 años de edad de la madre, se ubica aproximadamente el 60 % de los nacimientos; 20 % de madres menores de 24 y 20 % de madres mayores de 37 años.

5.3.12. PICTOGRAMAS

En términos generales, los pictogramas siguen las reglas de los gráficos de barras, con la particularidad de que en lugar de un frío rectángulo, utilizan figuras alegóricas a lo que se está representando. Por ejemplo, si se informa de nacimientos, pueden usarse siluetas de niños o de mujeres embarazadas; si se habla de reforestación, puede usarse un árbol; si se trata de construcción, una casa, etc.



Figura 5.29. **Siluetas para decorar pictogramas**

Los símbolos utilizados deben ser del mismo tamaño y lo que informa su magnitud es la cantidad de ellos que conforman la barra. Debe acompañarse de una leyenda o recuadro que indique la magnitud que cada dibujo representa, por ejemplo: cien niños, cien embarazos, mil árboles plantados, 500 cabezas de ganado. En los pictogramas no se utilizan escalas y solamente se busca informar datos aproximados, por ejemplo:



Figura 5.30. **Divorcios en Costa Rica según datos de la Dirección del Registro Civil**
Fuente: Golcher, R. (2005)

5.3.13. MAPAS ESTADÍSTICOS

La base de la información la constituye un plano o un mapa, sobre el cual se sobrepone la información, por medio de puntos, sombras, colores, rayas, dibujos y hasta objetos como alfileres, botones y otros dispositivos.

Un ejemplo de mapa estadístico es la distribución geográfica del Índice de Desarrollo Humano en Costa Rica, en donde el color blanco señala los cantones con menor desarrollo humano y las más oscuras indican zonas de alto IDH.

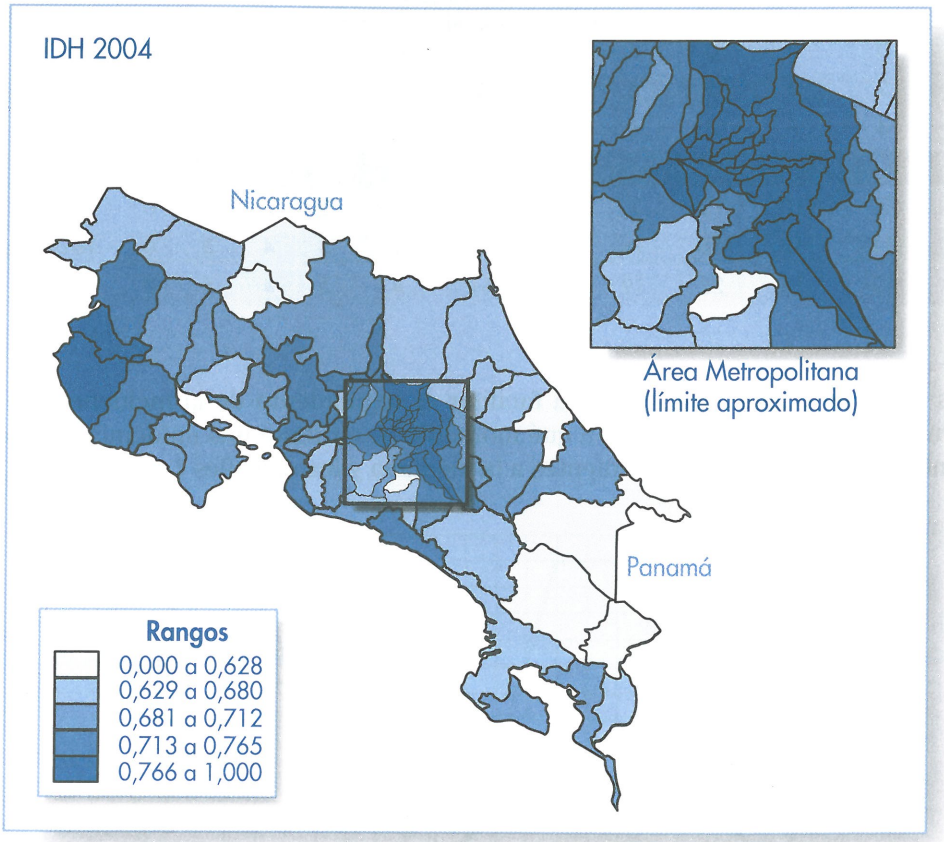


Figura 5.31. **Costa Rica: Índice de desarrollo humano 2004**

Fuente: Omodeo (2006)

5.3.14. EL GRÁFICO DE ÍNDICE ENDÉMICO

En el Capítulo IV se construyó un cuadro para dar seguimiento a los casos mensuales de rubeola en el cantón de Turrucalito entre los años 2002 y 2009. Los datos se ordenaron de menor a mayor número de casos en el mes y luego se definieron los cuartiles primero y tercero así como la mediana, obteniendo:

Cuadro 5.13

Índice endémico para los casos mensuales de rubeola en el cantón de Turrucalito (entre los años 2002 y 2009)

Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Set	Oct	Nov	Dic	
0	0	0	0	4	0	0	0	0	0	0	0	
2	**2**	**2**	**2**	**5**	**1**	**0**	**1**	**1**	**0**	**0**	**0**	**Primer Cuartil**
3	4	7	7	6	4	1	2	1	1	1	0	
3	**6**	**8**	**8**	**6**	**7**	**4**	**2**	**2**	**2**	**2**	**0**	**Mediana**
3	7	10	12	7	7	4	2	2	2	3	0	
5	**11**	**12**	**16**	**16**	**8**	**7**	**4**	**5**	**4**	**3**	**0**	**Tercer cuartil**
7	12	18	22	19	10	11	4	5	4	5	0	

Estos datos se pueden graficar para facilitar el seguimiento de la incidencia de una enfermedad y poder determinar en qué momento se enfrenta una epidemia. Las series que interesan son las correspondientes a la mediana y los cuartiles 1 y 3, así:

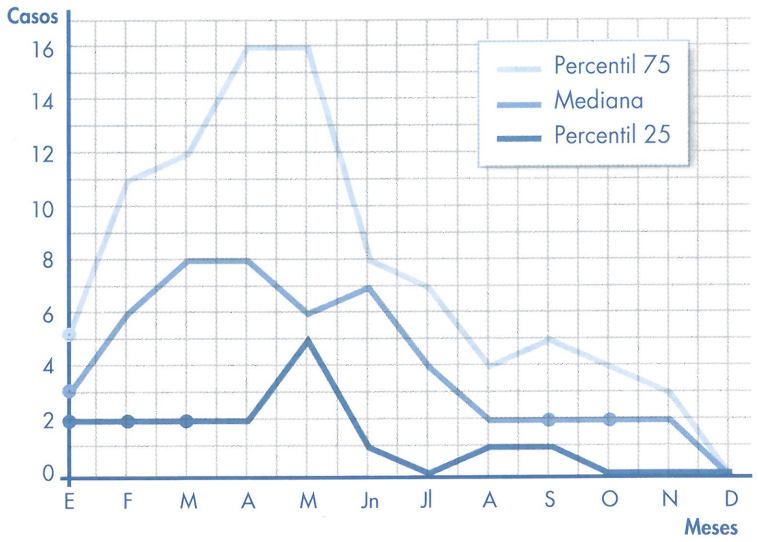


Figura 5.32. **Índice endémico para los casos mensuales de rubeola en el cantón de Turrucalito (entre los años 2002 y 2009)**

Fuente: Datos hipotéticos con propósitos didácticos

Usted recordará que el índice endémico es la mediana de casos de cada mes y que se tienen como variaciones toleradas las que se dan al interior del intervalo intercuartil del mes observado.

Las alarmas por epidemia se encenderán cuando los casos del mes sean iguales o superen al tercer cuartil.



EJERCICIOS DE REPASO Y AUTOEVALUACIÓN

Esta sección pretende que haga un repaso general de los contenidos del Capítulo.

EJERCICIO 5.10

De entre los gráficos estudiados en este Capítulo, indique qué tipo es más adecuado para representar las siguientes distribuciones:

a) El peso en kilogramos de niños recién nacidos.

b) Las consultas atendidas por especialidades médicas.

c) Los casos de malaria por distritos de una comunidad.

d) Las defunciones hospitalarias en los diferentes meses de un año, de forma que se puedan comparar las defunciones de hombres y las de mujeres.

EJERCICIO 5.11

Encierre en un círculo la opción que se ajusta al enunciado, Falsa (F) o Verdadera (V).

F V 1. La mayor ventaja de la utilización de gráficos estadísticos en los informes de investigación es la relativa facilidad para ser interpretados por el gran público.

F V 2. Un gráfico estadístico bien confeccionado no requiere de título para comunicar la información estadística relevante.

F V 3. Los gráficos tienen la ventaja de presentar los datos con elevada precisión, permitiendo la inclusión de muchas variables.

F V 4. Utilizar buenos gráficos en un informe de investigación permite prescindir de los cuadros estadísticos.

F V 5. Los gráficos lineales son muy útiles para la representación de series cronológicas (de tiempo).

F V 6. Puesto que la inclinación de las líneas que unen los diferentes puntos ayudan a la interpretación, es indispensable que el área que ocupa el cuerpo del gráfico no sobrepase una relación base por altura de 1 a 1,5.

F V 7. El gráfico de barras horizontales es muy útil para representar la participación relativa o porcentajes de las diferentes categorías.

F V 8. Cuando las categorías son muchas, en un gráfico de barras, se pueden presentar solamente las de mayor frecuencia.

F V 9. Las barras compuestas tienen dos o más componentes que dan el largo total a la barra.

F V 10. En el gráfico de barras comparativas se separan los subgrupos de cada categoría. Esto permite al lector hacer las comparaciones en forma rápida, aunque se limita la apreciación del total por categoría.

F V 11. Los gráficos de barras compuestas y los de barras comparativas compiten entre sí, por lo que se recomienda no utilizarlos conjuntamente.

F V 12. Tanto para los gráficos de barras compuestas como para los gráficos de barras comparativas, entre mayor es el número de componentes, es mejor.

F V 13. El uso de barras horizontales es apropiado tanto para series cronológicas como para variables cuantitativas discretas o categorizadas por intervalos.

F V 14. Al igual que los gráficos de barras horizontales, en el gráfico de columnas lo que informa es el ancho de las barras y no su longitud.

F V 15. Cuando lo que interesa representar es la participación porcentual de las categorías con respecto del total, se pueden utilizar pictogramas o mapas estadísticos.

F V 16. En la barra 100 % lo que se destaca es la importancia relativa de cada una de las categorías y no el valor absoluto de cada una de ellas.

F V 17. El histograma sirve para representar una distribución de frecuencias de variable cuantitativa continua.

F V 18. La pirámide de población es una variedad de histograma que expresa gráficamente la constitución por edad y sexo de los habitantes de un país.

F V 19. La representación típica de las frecuencias acumuladas se llama ojiva y permite visulizar el número de observaciones que se encuentran por encima o por debajo de un valor determinado.

F V 20. La mayor ventaja de las ojivas para el cálculo gráfico de los percentiles es su gran exactitud.

EJERCICIO 5.12

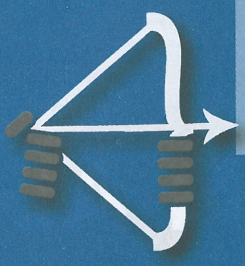
Pareo. Llene los paréntesis con el número que acompaña a cada definición de la izquierda, de forma que cada definición coincida con su concepto. Quedan dos paréntesis en blanco.

1.	Gráfico estadístico que se compone de un rectángulo suficientemente ancho cuya altura constituye el 100 %, luego se marcan las diferentes áreas para indicar los porcentajes que cada clase representa del total.	( )	Barras comparativas
2.	Tipo especial de gráfico de barras que contiene dos o más categorías que se pretenden comparar.	( )	Mapas estadísticos
3.	Representación de datos estadísticos en un plano cartesiano utilizando líneas que indican distancia, tendencia, pendiente o cambio de rumbo. Se utilizan para representar series cronológicas.	( )	Pictograma
4.	Diagrama usado para representar distribuciones de frecuencia de variable continua. Se conforma de columnas sin separación, cuya base indica el ancho de clase y el área del rectángulo la frecuencia de clase.	( )	Barra 100 %
5.	Forma de presentación de datos estadísticos sobre la base de un plano o un mapa, sobe el cual se sobrepone la información por medio de puntos, sombras, colores, rayas, dibujos y hasta objetos como alfileres, botones y otros dispositivos.	( )	Gráfico de columnas
6.	Gráfico estadístico de frecuencias acumuladas que permite visualizar el número de observaciones que se encuentran por encima o por debajo de un valor determinado.	( )	Barras de dos direcciones
7.	Representación estadística que utiliza figuras alegóricas a lo que se está informando. Siguen las reglas de los gráficos de barras y se informa tanto por el número de ilustraciones como por el largo del conjunto de ilustraciones.	( )	Diagrama de índice endémico
8.	Es una variedad de histograma que expresa gráficamente la conformación por edad y sexo de los habitantes de un país.	( )	Histograma
9.	Gráfico adecuado para representar series cronológicas e información sobre variables cuantitativas discretas. La información se transmite por el largo de la barra y no por su ancho.	( )	Pirámide de población
10.	Gráfico que permite dar seguimiento a la incidencia de una enfermedad para alertar la eventual llegada de brotes epidémicos.	( )	Polígono de frecuencias
		( )	Ojivas
		( )	Gráfico lineal

CAPÍTULO 6

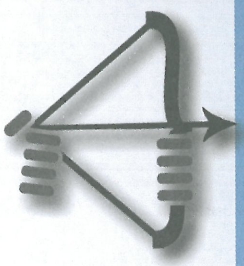
INTRODUCCIÓN AL ESTUDIO DE LAS PROBABILIDADES

OBJETIVO GENERAL



Aplicar los enfoques sobre probabilidades, cálculo del riesgo y manejo de la incertidumbre en las ciencias de la salud.

OBJETIVOS ESPECÍFICOS



Al finalizar el estudio de este capítulo usted estará en la capacidad de:

1. Enumerar los diferentes eventos que conforman el espacio muestral de un experimento.
2. Identificar, en pares de eventos, aquellos que son mutuamente excluyentes o colectivamente exhaustivos.
3. Diferenciar los enfoques de probabilidades y las situaciones concretas en que pueden ser aplicados.
4. Calcular probabilidades para eventos mutuamente excluyentes y no excluyentes.
5. Resolver problemas de probabilidades con eventos independientes y dependientes.
6. Distinguir las probabilidades conjuntas de las marginales y de las condicionales.
7. Realizar cálculos de probabilidades conjuntas y condicionales aplicables a pruebas de tamizaje.

Aunque no se puede tener plena certeza del rumbo de las actividades vitales como el nacimiento, la enfermedad o la muerte, sí se puede calcular el riesgo asociado a sus ocurrencias. Para ello, se trabajan los enfoques de probabilidad y la resolución de problemas con eventos individuales, luego con pares de eventos mutuamente excluyentes o que pueden ocurrir conjuntamente, para finalmente trabajar la forma en que ocurrencias previas condicionan la probabilidad de eventos sucesivos.

Una vez descartado lo imposible,
lo que queda, por improbable que parezca,
debe ser la verdad.

SIR ARTHUR CONAN DOYLE

La incertidumbre es parte de la vida. Se quiera o no, la humanidad está expuesta a:

- Reproducirse, enfermar, sanar o morir.
- Enfrentar eventos naturales, como los terremotos, deslizamientos, tornados, inundaciones o huracanes.
- Situaciones de juego como el lanzamiento de una moneda para escoger el lado de la cancha que un equipo defenderá, una partida de naipes, una rifa o la lotería.
- Problemas de decisión para asegurar o no una casa, un automóvil, mercadería, la salud y hasta la vida.
- Alzas y bajas en las tendencias económicas o de los negocios.
- Incluso, está influenciada por opiniones humanas sobre temas de interés colectivo.

Cuando un paciente va ser intervenido quirúrgicamente, generalmente, se pone nervioso. En el fondo, sabe que existe riesgo de complicaciones, de no sanar, incluso de morir. Por su parte, el personal médico ha estudiado las alternativas de tratamiento, los antecedentes, los riesgos involucrados y ha trabajado para minimizar la incertidumbre. Intuitivamente, tanto pacientes como médicos, saben que existe una probabilidad de error, efectos secundarios y variabilidad.

Entender las probabilidades proporciona una valiosa herramienta para medir los riesgos y ayuda a una toma de decisiones documentada; sin embargo, conocer de probabilidades no elimina completamente el riesgo. Del rumbo de las actividades vitales no se puede tener plena certeza, pero sí conocer su mayor o menor probabilidad de ocurrencia.

EJERCICIO 6.1

Explique por qué las afirmaciones siguientes están basadas en situaciones probabilísticas:

a) La tasa de sobrevivencia de los pacientes que se someten a esta cirugía es 90%.

b) Las primas que el instituto asegurador cobra por los seguros de vida son más altas para personas mayores de 55 años.

c) Fumar produce cáncer.

d) La esperanza de vida al nacer es de 77 años.

6.1. ESPACIO MUESTRAL Y EVENTOS DE UN EXPERIMENTO

La palabra *probabilidad* es un término comúnmente usado en relación con las oportunidades de que un evento particular ocurra al efectuar un experimento.

Un *experimento* es un proceso controlado que produce observaciones, por ejemplo:

- Tirar un dado
- Lanzar una moneda
- Pesar a un paciente
- Efectuar un examen

Los *eventos* son los posibles resultados del experimento:

- Que el dado caiga en seis
- Que la moneda muestre escudo
- Más de 60 a 70 kilogramos
- Paciente sano

La probabilidad de ocurrencia del evento A, se representa como $P(A)$ y es un número que va del 0 al 1,00 e indica cuan previsible es la ocurrencia del evento "A".

$$0 \leq P(A) \leq 1{,}00$$

En los extremos, cuando la probabilidad es igual a cero, se está en presencia de un *evento imposible*; en tanto que, si la probabilidad es igual a uno, se dice que es un *evento seguro* pues se tendría certeza absoluta de su ocurrencia.

Las probabilidades también pueden expresarse en forma de porcentajes, decimales o fracciones. De esta forma, la probabilidad de que a la sala de emergencias llegue un paciente infartado puede ser del 25 %, de 25 veces de entre 100, de 0,25 o de ¼, todas estas notaciones informan lo mismo.

Frecuentemente se plantean experimentos cuyo resultado no es predecible, pero se conocen de antemano todos los posibles eventos o resultados. Al conjunto de todos los eventos de un experimento se llama *espacio muestral.*

Algunos experimentos y su espacio muestral son como los siguientes:

i. Si un experimento es la concepción de un bebé, entonces el espacio muestral sobre su género es:

$$Espacio\ Muestral = \{M, F\}$$

Donde el evento "M" significa masculino y el evento "F", femenino.

ii. Si el experimento consiste en lanzar dos monedas y los eventos el caer escudo (E) o corona (C), entonces:

$$Espacio\ Muestral = \{(E, E),\ (E, C),\ (C, E),\ (C, C),\ \}$$

Este es el espacio muestral, donde el evento (E, E) ocurre si ambas monedas caen escudo, (E, C) si la primera moneda cae escudo y la segunda cae corona, (C, E) si la primera moneda cae corona y la segunda escudo, y (C, C) si ambas monedas caen corona.

iii. Si un experimento consiste en determinar el tipo de sangre de los pacientes, entonces el espacio muestral es:

$$Espacio\ Muestral = \{O,\ A,\ B,\ AB\}$$

Esto es, el espacio muestral lo conforman todos los grupos sanguíneos que se puedan documentar.

La teoría de la probabilidad toma en cuenta las relaciones que se dan entre el conjunto de eventos. Así, se tiene que el *complemento* de un evento consta de todos los resultados del espacio muestral que no forman parte del evento. Por ejemplo:

i. El complemento del evento femenino es masculino

$$\bar{F} = M$$

ii. El complemento de obtener EE al lanzar dos monedas lo forman el resto de los eventos que completan el espacio muestral:

$$\bar{E},\bar{E} = \{(E, C),\ (C, E),\ (C, C)\}$$

iii. Estadísticamente, el complemento del tipo de sangre O es

$$\bar{O} = \{A,\ B,\ AB\}$$

El complemento se denota con una línea sobre el evento, de forma que el complemento del evento A es $\bar{A}$ y la probabilidad complementaria de A será:

$$P(\bar{A}) = 1 - P(A)$$

Se dice que los eventos son *mutuamente excluyentes* si uno y solo uno de ellos puede tener lugar a la vez, es decir, que no presentan elementos en común. En el lanzamiento de una moneda, cualquiera de los dos resultados es posible, pero no ambos a la vez; por lo tanto, se dice que los eventos escudo y corona en un solo lanzamiento son mutuamente excluyentes.

Cuando en el listado de los posibles eventos resultantes de un experimento se incluyen todos los resultados posibles, se está en presencia de eventos *colectivamente exhaustivos*. En el ejemplo de la moneda, los eventos escudo y corona son colectivamente exhaustivos puesto que se supone que la moneda al caer no puede permanecer perpendicular.

Es importante observar que la probabilidad representada por todo el espacio muestral es de de 1 o, amplificado, del 100 %:

$$P(\text{espacio muestral}) = 1{,}00$$

Otra forma de decirlo es: si los eventos son mutuamente excluyentes, la suma de las probabilidades de ocurrencia de todos los eventos es igual a 1.

La probabilidad de que el evento A no ocurra es 1 menos la probabilidad de que sí lo haga:

$$1 - P(A) = P(\bar{A})$$

Esto hace que un evento y su complemento sean a la vez mutuamente excluyentes y colectivamente exhaustivos:

$$P(A) + P(\bar{A}) = 1{,}00$$

EJERCICIO 6.2

a) Indique el experimento y el espacio muestral de cada uno de los siguientes ejemplos.

   i. Tomar muestras de sangre en ayunas y clasificar a los pacientes en diabéticos y no diabéticos.

   ii. Pesar recién nacidos y anotar sus pesos. Estudios previos indican que los pesos pueden ir de 500 a 5000 gramos.

   iii. Presentar un examen de estadística y anotar las calificaciones que pueden variar de 0 a 10.

b) Elabore una breve definición de cada uno de los siguientes conceptos:

   i. Experimento

   ii. Espacio muestral

   iii. Evento

   iv. Complemento de un evento

   v. Eventos mutuamente excluyentes

   vi. Eventos colectivamente exhaustivos

c) Señale si los siguientes pares de eventos son mutuamente excluyentes.

i.	Viudo	Casado	
ii.	Desempleado	Diabético	
iii.	Esterilizado	Sucio	
iv.	Ser alérgico	Tener gripe	
v.	Menor de edad	Mayor de edad	
vi.	Enfermo	Asegurado	
vii.	Contratado	No contratado	

d) Anote si los siguientes eventos son colectivamente exhaustivos.

i.	Nacional o extranjero	
ii.	Soltero o casado	
iii.	Esterilizado o no esterilizado	
iv.	Vacunado o no vacunado	
v.	De Guanacaste o Limón	
vi.	Ganar o perder un partido de fútbol	
vii.	Tipos de sangre A o B	

e) Determine el complemento de cada uno de los siguientes eventos.

i.	Empatar un partido de futbol	
ii.	Tipos de sangre A y B	
iii.	No vacunados	
iv.	Menos de tres días de incapacidad	
v.	Tres o menos días de incapacidad	
vi.	Guanacaste, Puntarenas, Cartago	
vii.	Sacar 6 al tirar un dado	

6.2. TRES ENFOQUES SOBRE PROBABILIDADES

Existen tres formas conceptuales de abordar el estudio de las probabilidades, cada una con sus ventajas y desventajas. Estas pueden ser clasificadas en:

- Enfoque clásico
- Enfoque empírico
- Enfoque subjetivo

6.2.1. ENFOQUE CLÁSICO

El *enfoque clásico* considera que los eventos tienen una probabilidad de ocurrencia que se obtiene al dividir un evento favorable entre todos los eventos del espacio muestral:

$$P(A) = \frac{Evento\ favorable}{Número\ total\ de\ resultados\ posibles}$$

Ejemplos:

i. Al lanzar una moneda, solamente existen 2 resultados posibles, así que la probabilidad de que caiga corona es:

$$P(C) = \frac{1}{2}$$

ii. Desde esta perspectiva, la probabilidad de que el bebé por nacer sea de sexo femenino sería:

$$P(\text{femenino}) = \frac{1}{2}$$

iii. Al lanzar un dado, existen 6 resultados posibles, así que la probabilidad de obtener un 4 es:

$$P(4) = \frac{1}{6}$$

El enfoque clásico se conoce también como probabilidad *a priori*, debido a que se establece una respuesta de antemano sin necesidad de tirar monedas, lanzar dados o registrar los nacimientos. No necesariamente se deben efectuar los experimentos para llegar a conclusiones sobre distribuciones de probabilidades puesto que las estimaciones se basan en el razonamiento lógico.

Sin embargo, existen muchas situaciones en donde no se puede conocer el número total de resultados posibles y, si se asumiera un determinado escenario, no necesariamente se debe suponer que cada evento tenga la misma probabilidad de ocurrencia.

En el caso de los seguros de vida, no sería lógico establecer dos eventos: que el asegurado viva o que muera, ni tampoco asignarle a cada evento una probabilidad de ocurrencia de ½. Por eso, las empresas de seguros procesan datos sobre defunciones por rango de edad, riesgos laborales, enfermedades profesionales y esperanza de vida, todo ello para establecer modelos probabilísticos.

6.2.2. ENFOQUE EMPÍRICO

El *enfoque empírico* basa la asignación de probabilidades en registros históricos, es decir, en las frecuencias relativas en que se ha presentado un evento.

Por lo tanto, según este enfoque, la definición de probabilidad es el porcentaje de veces que un evento se ha presentado históricamente bajo las mismas condiciones. En forma de ecuación, la probabilidad es igual al número de veces que ocurre el evento A, dividido entre el número total de observaciones.

$$P(A) = \frac{\textit{Número de veces que ocurre } A}{\textit{Número total de observaciones}}$$

De esa manera, se determina con qué frecuencia ha sucedido algo en el pasado y, si las condiciones no se han alterado drásticamente, se utiliza ese dato para predecir el riesgo de que ocurra nuevamente.

Por ejemplo:

i. El Instituto Nacional de Seguros tiene registros históricos de que cada año mueren alrededor de 13 de cada 100 hombres de 65 años o más, por eso estima que la probabilidad de muerte de ese grupo de edad es:

$$P(M) = \frac{13}{100} = 0,13$$

ii. Los exámenes de sangre de 1 000 pacientes informan que 600 de ellos pertenecen al grupo sanguíneo *O*. Por lo que la probabilidad de atender a una víctima de un accidente de tránsito con sangre *O* es:

$$P(O) = \frac{600}{1000} = 0,6$$

iii. Los registros del servicio de emergencias en el Hospital Nacional de Niños indican que 32 de cada 100 niños atendidos por las noches ingresan con problemas respiratorios, por tanto el Director de este servicio calcula:

$$P(R) = \frac{32}{100} = 0,32$$

En 32 % el porcentaje de pacientes que acudirá con problemas respiratorios.

Para utilizar adecuadamente el enfoque empírico o frecuencial es necesario tomar en cuenta las limitaciones siguientes:

- La probabilidad obtenida es solamente una estimación del valor real y nunca una garantía de tener la probabilidad exacta.
- Cuanto mayor sea el tamaño de la muestra, mejor será la estimación de probabilidad.
- Cuanto más alta sea la fiabilidad de los registros, mejor será la estimación de probabilidad.
- La probabilidad estimada será apropiada siempre que se mantengan las condiciones bajo las cuales se obtuvieron los datos.

6.2.3. ENFOQUE SUBJETIVO

El *enfoque subjetivo* de la probabilidad está basado en la experiencia de las personas que efectúan la estimación. Así, un individuo puede basarse en la evidencia que tenga disponible, en experiencias previas o en una creencia meditada. Decisiones tales como el lanzamiento al mercado de una droga novedosa o la dedicación para encontrar el tratamiento efectivo de una enfermedad incurable están basadas muchas veces en el empeño del investigador o en una corazonada del dirigente.

Ejemplos:

i. Usted es el Director de un Hospital y tiene que reclutar a un nuevo trabajador social de entre tres candidatos con similar formación académica, experiencia laboral, aspiraciones, apariencia física, autoconfianza, buenas referencias, excelente hoja de servicios, disponibilidad inmediata y que manifestaron en la entrevista tener el ánimo para enfrentar nuevos retos. ¿Cuál es la probabilidad de que cada candidato se acople a la institución y se relacione exitosamente con los pacientes? Para responder a este interrogante, se tendrá que asignar probabilidades subjetivas al potencial de cada aspirante por el puesto.

ii. El Rector de una Universidad tiene que decidir sobre la apertura de una especialidad novedosa que actualmente no cuenta con un mercado laboral delimitado para sus graduados. ¿Qué probabilidad le asigna al evento de demanda alta dentro de cinco años cuando estén listos los primeros graduados?

iii. El instituto de seguridad social quiere participar en la investigación para la cura del HV/SIDA, trabajo que requiere una buena dotación de recursos humanos y financieros, justo en un periodo en que se enfrentan restricciones presupuestarias. Si se descubre un tratamiento efectivo, se patentará y generará altos ingresos, en el caso contrario los recursos asignados se considerarán desperdiciados.

Como se puede apreciar en los ejemplos, la gran mayoría de las decisiones sociales, estratégicas y políticas tratan con situaciones únicas y muy pocas son cuestiones rutinarias o idénticas, por lo que con mucha frecuencia se tendrá que hacer uso de la probabilidad subjetiva.

Es necesario tener en cuenta que dos personas razonables, enfrentadas al mismo caso, normalmente asignarán probabilidades subjetivas diferentes para el mismo evento, puesto que existen algunas personas propensas al riesgo y otras adversas al riesgo.

Al igual que los otros enfoques de probabilidad, la estimación subjetiva presenta sus desventajas propias, entre ellas está el hecho de que el porcentaje asignado es difícil de justificar si es cuestionado; las asignaciones están afectadas por los prejuicios, por ideas preconcebidas, por temores conscientes o inconscientes y por los sentimientos acerca de lo que se quiere que suceda. En todo caso, la actitud, la experiencia, la visión de conjunto y la capacidad técnica ayudarán a minimizar esas dificultades.

EJERCICIO 6.3

a) ¿Cuál es la probabilidad de obtener el 6 en un lanzamiento de un dado? ¿Cuál es la probabilidad de su complemento?

b) Una urna contiene 50 bolas plásticas: 20 amarillas, 15 rojas, 10 azules y 5 blancas. Si se saca una bola en forma aleatoria, calcule la probabilidad de que esta bola sea:

i. Amarilla
ii. Roja
iii. Roja o azul
iv. Diferente de azul
v. Negra
vi. Diferente de negra

c) Los registros del servicio de salas de emergencia de un hospital indican que, históricamente, se han atendido los siguientes casos:

Ataque cardíaco	24 %
Enfermedades respiratorias	22 %
Accidentes de tránsito	20 %
Quemados	10 %
Envenenamientos	8 %
Otras emergencias	16 %
TOTAL	100 %

i. ¿Qué supuestos debe adoptar para considerar esos datos como insumos para generar probabilidades?

ii. En caso de validar sus supuestos, ¿cuál es la probabilidad de tener que atender a un paciente víctima de un accidente de tránsito?

iii. ¿Cuál es la probabilidad de que el paciente por atender no sea víctima de un accidente de tránsito?

d) Schapiro (1972) realizó exámenes a 10,530 personas para determinar su grupo sanguíneo, con los resultados siguientes:

TIPO	FRECUENCIA	PROBABILIDAD
O	6 332	
A	3 027	
B	774	
AB	397	
	10 530	

i. Calcule la distribución de probabilidades de los diferentes grupos sanguíneos.

ii. ¿Cuál es la probabilidad de que una persona escogida al azar sea del grupo B?

iii. ¿Cuál es la probabilidad de que no sea del grupo B?

6.3. REGLAS DE LAS PROBABILIDADES

Al utilizar la teoría de las probabilidades, con frecuencia interesan dos situaciones:

- Que se presente uno u otro evento: P (A o B)
- Que dos o más eventos se presenten al mismo tiempo: P (A y B)

Si lo que interesa es determinar la probabilidad de que por lo menos un evento ocurra, se está en situaciones de P (A o B), y el cálculo será distinto si se trata de eventos mutuamente excluyentes o no.

6.3.1. EVENTOS MUTUAMENTE EXCLUYENTES

En caso de que los eventos sean *mutuamente excluyentes*, la probabilidad de A o B es la suma de cada una de las probabilidades individuales.

$$P(A \text{ o } B) = P(A) + P(B)$$

Por ejemplo:

i. La probabilidad de que al tirar un dado se obtenga cinco o seis es:

$$P(5 \text{ ó } 6) = P(5) + P(6) = \frac{1}{6} + \frac{1}{6} = \frac{2}{6}$$

ii. Schapiro (1972) realizó exámenes a 10,530 personas para determinar su grupo sanguíneo, con los resultados siguientes:

TIPO	FRECUENCIA
O	6 332
A	3 027
B	774
AB	397
	10 530

¿Cuál es la probabilidad de atender una emergencia de un paciente del grupo sanguíneo A o del B?

$$P(A \text{ o } B) = P(A) + P(B) = \frac{3\,027}{10\,530} + \frac{774}{10\,530} = \frac{3\,801}{10\,530}$$

iii. De una urna que contiene 50 bolas plásticas: 20 amarillas, 15 rojas, 10 azules y 5 blancas. La probabilidad de sacar una bola amarilla o roja es:

$$P(A \text{ o } R) = P(A) + P(R) = \frac{20}{50} + \frac{15}{50} = \frac{35}{50}$$

EJERCICIO 6.4

a) Al sacar una bola de la urna del ejemplo anterior, ¿cuál es la probabilidad que la bola sea roja o blanca?

b) Al sacar una carta de un mazo de 52 naipes, la probabilidad de que esta sea una reina (Q) o un rey (K) es:

c) El once titular de la selección masculina de futbol estará compuesto por 3 jugadores de Saprissa, 3 de Herediano, 3 de Alajuelense, 1 de Puntarenense y 1 de Cartago. Al finalizar el partido, se efectuará un control antidoping a un jugador escogido al azar, ¿cuál es la probabilidad de que el escogido sea de Saprissa o de Puntarenense?

d) ¿Cuál es la probabilidad de que el escogido para el control antidoping sea de Herediano o de Alajuelense?

6.3.2. EVENTOS QUE NO SON MUTUAMENTE EXCLUYENTES

Si dos eventos *no son mutuamente excluyentes*, ambos pueden suceder al mismo tiempo, el cálculo deberá entonces tomar en cuenta el hecho de que se presente uno de ellos o ambos, la regla de la suma se ajusta de la forma siguiente:

$$P(A \text{ o } B) = P(A) + P(B) - P(A \text{ y } B)$$

Ejemplos:

i. La probabilidad de sacar un as o una carta de diamantes en un mazo de 52 cartas es:

$$P(A \text{ o } D) = P(A) + P(D) - P(A \text{ y } D) = \frac{4}{52} + \frac{13}{52} - \frac{1}{52} = \frac{16}{52}$$

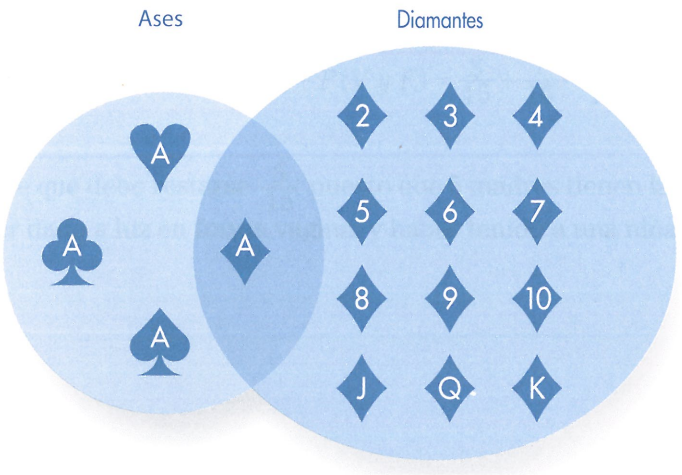


Figura 6.1. **Probabilidad de sacar un as o una carta de diamantes**

Se sabe que el mazo de 52 naipes tiene 4 ases y 13 cartas de diamantes. En caso de simplemente efectuar la suma de 4+13 se estaría considerando al as de diamantes dos veces; por ello, para evitar el conteo doble se tiene que restar la probabilidad de obtenerlo, esto es restar $\frac{1}{52}$.

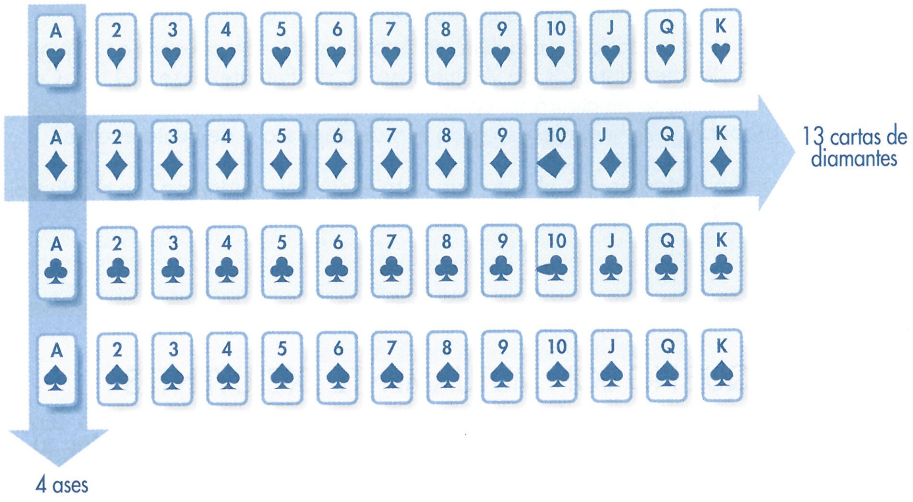


Figura 6.2. **Mazo de cartas con 52 naipes**

ii. Un grupo de trabajo formado por 3 hombres y 7 mujeres. De ellos, 5 personas (1 hombre y 4 mujeres) se manifiestan en contra del cambio de horario. La probabilidad de seleccionar a una persona que sea hombre o se manifieste en contra es:

$$P(H\ o\ C) = P(H) + P(C) - P(H\ y\ C) = \frac{3}{10} + \frac{5}{10} - \frac{1}{10} = \frac{7}{10}$$

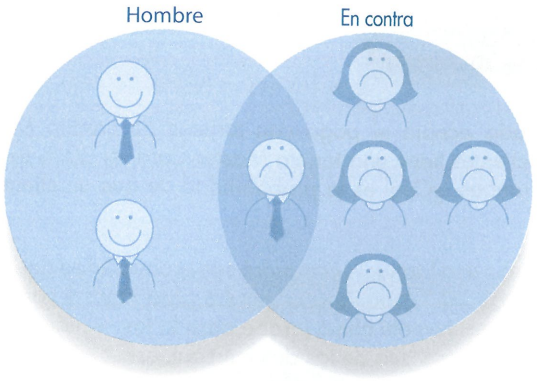


Figura 6.3. **Personal masculino o en contra del cambio de horario**

Hay 5 personas en contra del cambio de horario y 3 hombres, pero si simplemente se sumara 5+3 se estaría contando dos veces al hombre que está en contra.

iii. En una sala de obstetricia se atendieron 15 partos justo el Día de la madre. El informe médico ordenado por tipo de parto y sexo del recién nacido se presenta luego de este párrafo. Un periodista de un canal de televisión entrevistará a una madre escogida al azar ¿Cuál es la probabilidad que esta madre haya dado a luz en forma vaginal o su bebé sea de sexo femenino?:

	SEXO DEL RECIÉN NACIDO				
TIPO DE PARTO	FEMENINO		MASCULINO		TOTAL
	Palotes	F	Palotes	F	
Cesaria	/	1	////	4	5
Episiotomía	/	1	/	1	2
Vaginal	/////	5	///	3	8
TOTAL		7		8	**15**

Figura 6.4. **Informe de Sala de Obstetricia: ordenado por tipo de parto y sexo del recién nacido.**

$$P(V\ o\ F) = P(V) + P(F) - P(V\ y\ F) = \frac{8}{15} + \frac{7}{15} - \frac{5}{15} = \frac{10}{15}$$

No olvide que debe restarse $\frac{5}{15}$ puesto que 5 madres tienen la doble condición de haber dado a luz en forma vaginal y haber tenido a una niña.

EJERCICIO 6.5

a) Una clínica privada acepta el pago con tarjetas de crédito Master Card o Popular. Un 25 % de los clientes tiene una tarjeta Master Card, el 30 % tiene una tarjeta Popular y el 10 % tiene ambas. ¿Cuál es la probabilidad de que un cliente tenga al menos una de esas tarjetas?

b) Un cliente que entra a una farmacia comprará con probabilidad del 30 % un reconstituyente. Con probabilidad de 20 % adquirirá un jarabe para la tos y con probabilidad del 10 % comprará tanto reconstituyente como jarabe para la tos. ¿Cuál es la probabilidad de que el cliente compre al menos una de las dos mercancías?

c) Una empresa fabricante de termómetros debe revisar cuidadosamente cada aparato antes de entregarlo al Departamento de Compras Locales de la Seguridad Social. El jefe de control de calidad se ha dado cuenta que de cada 10,000 termómetros sometidos a revisión, 10 tienen defectos en la cantidad de mercurio, 8 tienen defectos en la impresión de la escala y 5 presentan ambos defectos ¿cuál es la probabilidad que, al inspeccionar un termómetro, este tenga alguno de los 2 defectos?

d) Se estima que cuando usted se gradúe tendrá el 70 % de probabilidades de ser llamado a trabajar para la seguridad social, también tendrá un 40 % de probabilidades de ser convocado por una empresa privada y el 30 % de ser requerido por ambas. ¿Cuál es la probabilidad de que sea llamado por una u otra institución?

6.3.3. PROBABILIDAD COMPLEMENTARIA

Ya se ha visto que el complemento de un evento lo forman todos los eventos que no le son favorables, de allí que la probabilidad complementaria sea igual a 1 menos la probabilidad del evento de interés:

$$\text{Probabilidad complementaria} = 1 - P(E)$$

Por ello, la probabilidad complementaria se puede expresar como un no evento:

$$P(\bar{E}) = 1 - P(E)$$

Ejemplos:

i. De una urna que contiene 50 bolas plásticas: 20 amarillas, 15 rojas, 10 azules y 5 blancas. La probabilidad de sacar una bola amarilla es:

$$P(A) = \frac{20}{50}$$

Y la probabilidad de que la bola no sea amarilla es:

$$P(\bar{A}) = 1 - \frac{20}{50} = \frac{30}{50}$$

ii. La probabilidad de atender una emergencia de un paciente del grupo sanguíneo A o del B es:

TIPO	FRECUENCIA	PROBABILIDAD
O	6 332	0,6013
A	3 027	0,2875
B	774	0,0735
AB	397	0,0377
	10 530	1,0000

Fuente: Schapiro (1972)

$$P(A \text{ o } B) = P(A) + P(B) = \frac{3\,027}{10\,530} + \frac{774}{10\,530} = \frac{3\,801}{10\,530}$$

Y la probabilidad de que un paciente no sea de ninguno de esos grupos sanguíneos es:

$$1 - P(A \text{ o } B) = 1 - \frac{3\,801}{10\,530} = \frac{6\,729}{10\,530}$$

EJERCICIO 6.6

a) Una clínica privada acepta el pago con tarjetas de crédito Visa o Popular. Un 25 % de los clientes tiene una tarjeta Visa, el 30 % tiene una tarjeta Popular y el 10 % tiene ambas. ¿Cuál es la probabilidad de que el cliente no tenga ninguna de esas tarjetas?

b) Un cliente que entra a una farmacia comprará con probabilidad del 30 % un reconstituyente. Con probabilidad de 20 % adquirirá un jarabe para la tos y con probabilidad del 10 % comprará tanto reconstituyente como jarabe para la tos. ¿Cuál es la probabilidad de que el cliente no compre ninguna de esas mercancías?

c) Una empresa fabricante de termómetros debe revisar cuidadosamente cada aparato antes de entregarlo al Departamento de Compras Locales de la Seguridad Social. El jefe de control de calidad se ha dado cuenta que de cada 10,000 termómetros sometidos a revisión, 10 tienen defectos en la cantidad de mercurio, 8 tienen defectos en la impresión de la escala y 5 presentan ambos defectos. ¿Cuál es la probabilidad de que no tengan ninguno de esos 2 defectos?

d) Se estima que cuando usted se gradúe tendrá el 70 % de probabilidades de ser llamado a trabajar para la seguridad social, también tendrá un 40 % de probabilidades de ser convocado por una empresa privada y el 30 % de ser requerido por ambas. ¿Cuál es la probabilidad de que no sea llamado por ninguna de estas dos instituciones?

6.3.4. PROBABILIDAD CONJUNTA

La probabilidad P (A y B), de que dos eventos ocurran sucesivamente o al mismo tiempo, se denomina *probabilidad conjunta*. El cálculo de la probabilidad conjunta varía si los eventos son independientes o dependientes uno del otro.

Dos eventos son *independientes* cuando la ocurrencia de uno no afecta la ocurrencia del otro. En el caso del lanzamiento de dos monedas, si la primera cae escudo no afecta para nada la probabilidad de la cara que presentará la otra moneda. En cambio, cuando los eventos son *dependientes*, conocer el resultado del primero es muy útil para predecir en mejor forma la probabilidad de ocurrencia del otro.

Cuando dos eventos son *independientes*, entonces la probabilidad de que ambos ocurran es igual a la multiplicación de sus probabilidades individuales o marginales.

Ejemplos:

i. Al lanzar una moneda, la probabilidad marginal de que caiga escudo es:

$$P(E) = \frac{1}{2} = 0,5$$

Ahora bien, si se lanzan dos monedas, la probabilidad conjunta de que ambas caigan escudo es igual al producto de la probabilidad marginal de que la primera caiga escudo por la probabilidad marginal de que la segunda también caiga escudo.

$$P(E\ y\ E) = \frac{1}{2} \times \frac{1}{2} = 0,5 \times 0,5 = 0,25$$

ii. Al lanzar dos dados no cargados la probabilidad de que ambos caigan en seis es:

$$P(6\ y\ 6) = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$$

iii. Las fallas de tres máquinas para efectuar hemodiálisis son independientes entre sí. Si sus probabilidades de avería en un día cualquiera son 3 %, 5 % y 8 % respectivamente. ¿Cuál es la probabilidad de que todas las máquinas se descompongan el mismo día?

$$0,03 \times 0,05 \times 0,08 = 0,00012$$

¿Cuál es la probabilidad de que todas funcionen adecuadamente?

$$0,97 \times 0,95 \times 0,92 = 0,84778$$

EJERCICIO 6.7

a) Según el resultado de la investigación sobre "Prevalencia del Asma en Costa Rica", cuyo cuadro resumen se reproduce seguidamente:

Prevalencia de asma (%) en diferentes países de latinoamérica de acuerdo con el repporte de ISAAC en dos grupos etareos

	Sibilancias en últimos 12 meses		Diagnóstico médico de asma	
Edad	6-7	13-14	6-7	13-14
Brasil	22,4	24	11,3	16,1
Chile	18,2	9,7	12,5	10,1
Costa Rica	32,1	23,7	26,9	18,5
Panamá	23,5	17,6	19,3	16,9
Perú	-	26,0	-	28

Fuente: Soto-Martínez y Soto-Quirós (2004)

De los niños de 6 a 7 años que vivían en Costa Rica en el año 2004, el 26,9 % tenía diagnóstico médico de asma.

Si se escoge aleatoriamente a 2 niños de esta edad, ¿cuál es la probabilidad de que ambos sean asmáticos?

b) Si se escoge aleatoriamente a 3 niños de 6 a 7 años, ¿cuál es la probabilidad de que ninguno sea asmático?

c) Las probabilidades de pertenecer a un grupo sanguíneo se presentan en la tabla siguiente.

TIPO	FRECUENCIA	PROBABILIDAD
O	6 332	0,6013
A	3 027	0,2875
B	774	0,0735
AB	397	0,0377
	10 530	1,0000

Fuente: Schapiro (1972)

¿Cuál es la probabilidad de atender en una emergencia a dos pacientes del grupo sanguíneo B?

d) Según el estudio sobre la prevalencia de caries en niños y niñas escolares, llevada a cabo por Montero y otros (2007), "84,4 escolares de 12 años de cada 100 están con la enfermedad o con secuelas de la misma", ¿cuál es la probabilidad de que al examinar 3 niños escogidos al azar, todos tengan caries o secuelas de la misma?

6.4. TABLAS DE CONTINGENCIA

Retómense las pruebas de tamizaje analizadas en el Capítulo III, en particular el siguiente ejemplo:

A 79 pacientes se les efectúa una ecografía para diagnosticar cáncer gástrico, de las cuales se obtienen 8 ecografías positivas y 71 negativas. Gastroscopías y biopsias posteriores proporcionan datos que permiten construir el siguiente cuadro:

Cuadro 6.1

Prueba de tamizaje (Ecografía gástrica)

	Con cáncer gástrico	Pacientes sanos	TOTAL
Ecografía positiva	6	2	8
Ecografía negativa	1	70	71
	7	72	79

Fuente: Adaptado con propósitos didácticos de Díaz Vélez (2010).

Si el dato de cada casilla se divide entre los 79 pacientes analizados, se genera la siguiente *tabla de contingencia* o tabla de distribución de probabilidades:

Cuadro 6.2

Prueba de tamizaje: Tabla de contingencia o distribución de probabilidades

	Con **cáncer gástrico**	Pacientes **sanos**	TOTAL
Ecografía **positiva**	0,0759	0,0253	0,1013
Ecografía **negativa**	0,0127	0,8861	0,8987
	0,0886	0,9114	1,0000

Con la ayuda de esta tabla de contingencia se facilita responder a las siguientes preguntas sobre probabilidades de ocurrencia:

i. ¿Cuál es la probabilidad de que la ecografía resulte positiva?

$$P(E^{+}) = 0,1013$$

Esta es una *probabilidad marginal* o simple porque solamente toma en cuenta la prueba de tamizaje (la ecografía) y no el diagnóstico final.

ii. ¿Cuál es la probabilidad de paciente sano luego del examen final?

$$P(S) = 0,9114$$

Esta también es una *probabilidad marginal* puesto que no toma en cuenta el resultado de la prueba de tamizaje (ecografía).

iii. ¿Cuál es la probabilidad de que la ecografía sea negativa y en el diagnóstico final resulte paciente sano?

$$P(E^{-} \text{ y } S) = 0{,}8861$$

Esta es una *probabilidad conjunta* puesto que considera la doble situación de dar negativo en la prueba de tamizaje y sano en la gastroscopía o la biopsia posterior. En este caso, los eventos no son excluyentes pues ambos pueden ocurrir en el mismo paciente.

iv. ¿Cuál es la probabilidad de que la ecografía resulte positiva o el diagnóstico final sea cáncer?

$$P(E^{+} \text{ o } C) = P(E^{+}) + P(C) - P(E^{+} \text{ y } C)$$

$$P(E^{+} \text{ o } C) = 0{,}1013 + 0{,}0886 - 0{,}0759$$

$$P(E^{+} \text{ o } C) = 0{,}114$$

Observe que cuando los eventos no son mutuamente excluyentes, si solamente se sumaran las probabilidades marginales, se estaría duplicando la participación de la probabilidad conjunta, esto es, de los casos simultáneamente favorables de dar positivo en la ecografía y tener cáncer. Para evitar esta duplicación, se resta la probabilidad conjunta.

Recuérdese que la probabilidad de ocurrencia de los eventos individuales se denomina *probabilidad marginal* y se llama *probabilidad conjunta* a la probabilidad de ocurrencia simultánea de dos o más eventos no excluyentes.

Si los eventos son excluyentes, la probabilidad conjunta será igual a cero debido a que no se tienen elementos en común.

EJERCICIO 6.8

a) 1400 varones mayores de 40 años son examinados por medio del Tacto Rectal como prueba de tamizaje para detectar cáncer de próstata, de los cuales resultan 540 con agrandamiento sospechoso y 860 sin agrandamiento. Puesto que la enfermedad es de alta prevalencia y gran letalidad, posteriormente se les realizan biopsias prostáticas, lo cual genera la siguiente tabla:

	Resultado de la biopsia prostática		
Resultado del tacto rectal	Cáncer	Benigno	Total
Agrandamiento	460	80	540
Sin agrandamiento	200	660	860
	660	740	1 400

Calcule la tabla de contingencia de probabilidades y conteste las preguntas que se formulan más adelante:

	Resultado de la biopsia prostática		
Resultado del tacto rectal	Cáncer	Benigno	Total
Agrandamiento			
Sin agrandamiento			

i. ¿Cuál es la probabilidad de que el Tacto Rectal detecte agrandamiento?

ii. ¿Cuál es la probabilidad de situación benigna luego de la biopsia?

iii. ¿Cuál es la probabilidad de que el Tacto Rectal resulte sin agrandamiento y de que la biopsia informe benigno?

iv. ¿Cuál es la probabilidad de que el Tacto Rectal detecte agrandamiento o de que el diagnóstico final sea cáncer?

b) A 700 pacientes se les practica una baciloscopía como prueba de tamizaje para la detección de la tuberculosis, de los cuales resultan 20 positivos y 680 negativos. Luego se efectúan cultivos de muestras no pulmonares, generando la siguiente tabla:

	Cultivo de muestras no pulmonares		
Baciloscopía	Enfermo	Sano	Total
Positiva	19	1	20
Negativa	41	639	680
	60	640	700

Calcule la tabla de contingencia de probabilidades y conteste las preguntas que se formulan seguidamente:

	Cultivo de muestras no pulmonares		
Baciloscopía	Enfermo	Sano	Total
Positiva			
Negativa			

i. ¿Cuál es la probabilidad de que la baciloscopía resulte negativa?

ii. ¿Cuál es la probabilidad de que el paciente esté enfermo de tuberculosis?

iii. ¿Cuál es la probabilidad de que el paciente en la baciloscopía salga positivo y de que en los cultivos de muestras no pulmonares resulte sano?

iv. ¿Cuál es la probabilidad de que la baciloscopía salga positiva o de que en el diagnóstico final esté enfermo de tuberculosis?

c) 290 mujeres con Síndrome Hipertensivo por Embarazo enfrentan el riesgo de que el feto sea pequeño para la edad de gestación. Para diagnosticarlo, son sometidas a ultrasonografía como prueba de tamizaje, de las cuales se detectan 10 con retraso del crecimiento fetal y 280 sin retraso del crecimiento fetal. Al nacer, los bebés son pesados y se obtienen los datos siguientes:

	Peso del recién nacido		
Ultrasonografía	Bajo Peso	Sanos	Total
Retraso del Crecimiento Fetal	8	2	10
Sin Retraso en el Crecimiento Fetal	40	240	280
	48	242	290

Calcule la tabla de contingencia de probabilidades y conteste las preguntas que se formulan a continuación:

	Peso del recién nacido		
Ultrasonografía	Bajo Peso	Sanos	Total
Retraso del Crecimiento Fetal			
Sin Retraso en el Crecimiento Fetal			

i. ¿Cuál es la probabilidad de que la ultrasonografía indique retraso del crecimiento fetal?

ii. ¿Cuál es la probabilidad de que el bebé efectivamente nazca con bajo peso?

iii. ¿Cuál es la probabilidad de que la ultrasonografía indique sin retraso en el crecimiento fetal y de que al nacer el bebé resulte con bajo peso?

iv. ¿Cuál es la probabilidad de que la ultrasonografía indique sin retraso en el crecimiento fetal o de que al nacer el bebé resulte con bajo peso?

6.5. INDEPENDENCIA Y DEPENDENCIA

Ya se ha dicho que la probabilidad P (A y B), de que dos eventos ocurran sucesivamente o a la vez, se conoce como *probabilidad conjunta* y su cálculo varía si los eventos son dependientes o independientes uno del otro.

Los eventos son *independientes* cuando la ocurrencia de uno no afecta la ocurrencia del otro. En el caso del lanzamiento de dos monedas, si la primera cae escudo no afecta para nada la probabilidad del evento que presentará la otra moneda. En cambio, cuando los eventos son *dependientes*, conocer la ocurrencia de uno es muy útil para predecir mejor la ocurrencia del otro.

Cuando dos o más eventos son *independientes*, la probabilidad conjunta de ocurrencia es igual a la multiplicación de sus probabilidades individuales o marginales.

$$P\,(A \text{ y } B) = P\,(A) * P\,(B)$$

Si los eventos son *dependientes*, la ocurrencia de uno afecta la del otro, entonces, saber que un evento ha ocurrido puede ser útil para predecir la ocurrencia del otro. Como regla general, se puede decir que la probabilidad *conjunta* de dos eventos dependientes es igual a la probabilidad de un evento multiplicada por la probabilidad condicional del otro.

$$P\,(A \text{ y } B) = P\,(A) * P\,(B|A)$$

La probabilidad condicional que se escribe P (B|A), se lee: "la probabilidad de que se presente el evento B, dado que el evento A ya se ha presentado".

La probabilidad condicional es la probabilidad de que un segundo evento (B) se presente si un primer evento (A) ya ha sucedido.

También podría decirse:

$$P\,(A \text{ y } B) = P\,(B) * P\,(A|B)$$

P (A|B) se lee: "la probabilidad de que se presente el evento A, dado que el evento B ya se ha presentado".

Para eventos estadísticamente independientes, la probabilidad condicional de que suceda el evento B dado que el evento A ya se ha presentado, es simplemente la probabilidad del evento B.

$$P\,(B|A) = P\,(B)$$

Observe que, por definición, eventos independientes son aquellos cuyas probabilidades no se ven afectadas de ninguna forma por la ocurrencia del otro.

Ejemplos:

i. ¿Cuál es la probabilidad de que, en el segundo lanzamiento de una moneda, esta caiga corona dado que en el primero cayó escudo?

Simbólicamente esto es:

$$P(C_2|E_1) = P(C)$$

Puesto que para eventos independientes el resultado del primer lanzamiento no tiene absolutamente ningún efecto sobre el resultado del segundo y como las probabilidades de obtener escudo o corona en cada lanzamiento son exactamente iguales, la probabilidad de obtener corona en el segundo lanzamiento es de $\frac{1}{2}$, por lo que se dice:

$$P(C_2|E_1) = P(C) = \frac{1}{2}$$

ii. Suponga que usted obtiene un 3 en el primer lanzamiento de un dado no cargado. ¿Cuál es la probabilidad de obtener un 5 en el siguiente lanzamiento?

Conocer que se ha obtenido un 3 en el lanzamiento anterior no ayuda en nada para saber qué ocurrirá en el segundo lanzamiento, por lo tanto:

$$P(5_2|3_1) = P(5) = \frac{1}{6}$$

iii. Todas las semanas el Comité Local de Voluntarios vende cien acciones para la rifa de ¢ 50 000, los cuales usted necesita con urgencia. Como usted es muy ordenado, ha llevado el registro de los números favorecidos cada semana y sabe que en la semana anterior el número premiado fue el 69. Usted decide jugar el 70. ¿Cuál es la probabilidad de "pegar" con el 70, dado que la semana recién pasada "cayó" el 69?

Nuevamente se está frente a un caso de independencia estadística, por lo cual la(s) ocurrencia(s) anterior(es) no afecta(n) a la(s) siguiente(s), por ello:

$$P(70_2|69_1) = P(70) = \frac{1}{100}$$

EJERCICIO 6.9

a) ¿Cuál es la probabilidad de que el segundo hijo de una pareja sea varón dado que primero tuvieron una niña? (asuma probabilidades bajo el enfoque clásico)

b) Sabiendo que "84,4 escolares de 12 años de cada 100 presenta caries o secuelas de la misma ¿Cuál es la probabilidad de que el segundo paciente tenga caries o secuelas dado que el primero no tenía este padecimiento?

c) Las probabilidades de pertenecer a un grupo sanguíneo se presentan en la tabla siguiente.

TIPO	FRECUENCIA	PROBABILIDAD
O	6 332	0,6013
A	3 027	0,2875
B	774	0,0735
AB	397	0,0377
	10 530	1,0000

Fuente: Schapiro (1972)

¿Cuál es la probabilidad de que el segundo paciente tenga sangre tipo O, dado que el primero tenía sangre tipo AB?

6.6. LA PROBABILIDAD CONDICIONAL

Cuando los eventos son *estadísticamente dependientes*, la probabilidad *condicional* de obtener el evento B, dado que el evento A ya se ha presentado, es igual a la probabilidad conjunta de los eventos A y B dividida entre la probabilidad marginal de ocurrencia del evento A:

$$P(B|A) = \frac{P(A \text{ y } B)}{P(A)}$$

Y la probabilidad condicional de obtener el evento A, dado que B ya se presentó, es igual a la probabilidad conjunta dividida entre la probabilidad marginal de B:

$$P(A|B) = \frac{P(A \text{ y } B)}{P(B)}$$

Retomando la siguiente *tabla de contingencia* o tabla de distribución de probabilidades, donde la ecografía se utilizó como prueba de tamizaje y las biopsias discriminaron con mayor precisión a los pacientes con cáncer gástrico de los sanos:

Cuadro 6.3

Prueba de tamizaje: Tabla de contingencia o distribución de probabilidades

	Con cáncer gástrico	Pacientes sanos	TOTAL
Ecografía positiva	0,0759	0,0253	0,1013
Ecografía negativa	0,0127	0,8861	0,8987
	0,0886	0,9114	1,0000

Esta tabla de contingencia permite calcular las siguientes probabilidades condicionales:

i. ¿Cuál es la probabilidad de que la ecografía resulte positiva dado que el paciente realmente tiene cáncer gástrico?

$$P(E^+|C) = \frac{P(E^+ y\ C)}{P(C)}$$

$$P(E^+|C) = \frac{0,0759}{0,0886} = 0,857$$

Esto indica la *sensibilidad* de la prueba de tamizaje y señala la proporción de pacientes con cáncer gástrico que se identifica por medio de una ecografía positiva.

ii. ¿Cuál es la probabilidad de que la ecografía indique negativo dado que el paciente está sano?

$$P(E^-|S) = \frac{P(E^- y\ S)}{P(S)}$$

$$P(E^-|S) = \frac{0,8861}{0,9114} = 0,972$$

Esto indica la *especificidad* de la prueba, que informa la proporción de pacientes sanos que resulta negativo en la prueba de tamizaje.

iii. ¿Cuál es la probabilidad de que el paciente tenga cáncer gástrico dado que la ecografía informó positivo?

$$P(C|E^+) = \frac{P(C\ y\ E^+)}{P(E^+)}$$

$$P(C|E^+) = \frac{0,0759}{0,1013} = 0,75$$

Esto cuantifica el *valor predictivo de una prueba positiva* (VPPP) e informa del porcentaje de pacientes que está realmente enfermo y fue diagnosticado positivo en la prueba de tamizaje.

iv. ¿Cuál es la probabilidad de que el paciente esté sano dado que la ecografía informó negativo?

$$P\,(S|E^-) = \frac{P(S\ y\ E^-)}{P(E^-)}$$

$$P\,(S|E^-) = \frac{0{,}8861}{0{,}8987} = 0{,}986$$

Esto cuantifica el *valor predictivo de una prueba negativa* (VPPN) e informa del porcentaje de pacientes que está sano y fue diagnosticado negativo en la prueba de tamizaje.

EJERCICIO 6.10

a) 1 400 varones mayores de 40 años son examinados por medio del Tacto Rectal como prueba de tamizaje para detectar cáncer de próstata, de los cuales resultan 540 con agrandamiento sospechoso y 860 sin agrandamiento. Puesto que la enfermedad es de alta prevalencia y gran letalidad, posteriormente se les realizan biopsias prostáticas, lo cual genera la siguiente tabla de contingencia de probabilidades:

	Resultado de la biopsia prostática		
Resultado del tacto rectal	Cáncer	Benigno	Total
Agrandamiento	0,3286	0,0571	0,3857
Sin agrandamiento	0,1429	0,4714	0,6143
	0,4714	0,5286	1,0000

Usando la tabla de contingencia, calcule las siguientes probabilidades condicionales:

i. ¿Cuál es la probabilidad de que el Tacto Rectal informe agrandamiento dado que el paciente realmente tiene cáncer prostático?

ii. ¿Cuál es la probabilidad de que el Tacto Rectal indique "sin agrandamiento" dado que es benigno?

iii. ¿Cuál es la probabilidad de que el paciente tenga cáncer de próstata dado que el Tacto Rectal informó agrandamiento?

iv. ¿Cuál es la probabilidad de que el paciente esté sano (benigno) dado que el Tacto Rectal informó sin agrandamiento?

b) A 700 pacientes se les practica una baciloscopía como prueba de tamizaje para la detección de la tuberculosis, de los cuales resultan 20 positivos y 680 negativos. Luego se efectúan cultivos de muestras no pulmonares, generando la siguiente tabla de contingencia de probabilidades:

	Cultivo de muestras no pulmonares		
Baciloscopía	Enfermo	Sano	Total
Positiva	0,0271	0,0014	0,0286
Negativa	0,0586	0,9129	0,9714
	0,0857	0,9143	1,0000

Usando la tabla de contingencia calcule las siguientes probabilidades condicionales:

i. ¿Cuál es la probabilidad de que la baciloscopía resulte positiva dado que el paciente está enfermo de tuberculosis?

ii. ¿Cuál es la probabilidad de que la baciloscopía indique negativo dado que el paciente está sano?

iii. ¿Cuál es la probabilidad de que el paciente esté enfermo de tuberculosis dado que la baciloscopía salió positiva?

iv. ¿Cuál es la probabilidad de que el paciente esté sano dado que la baciloscopía salió negativa?

c) 290 mujeres con Síndrome Hipertensivo por Embarazo enfrentan el riesgo de que el feto sea pequeño para la edad de gestación. Para diagnosticarlo, son sometidas a ultrasonografía como prueba de tamizaje, de las cuales se detectan 10 con retraso del crecimiento fetal y 280 sin retraso del crecimiento fetal. Al nacer, los bebés son pesados y se obtiene la siguiente tabla de contingencia de probabilidades:

	Peso del recién nacido		
Ultrasonografía	Bajo Peso	Sanos	Total
Retraso del Crecimiento Fetal	0,0276	0,0069	0,0345
Sin Retraso en el Crecimiento Fetal	0,1379	0,8276	0,9655
	0,1655	0,8345	1,0000

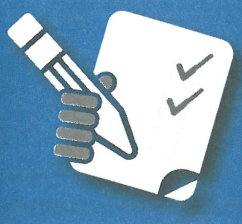
Usando la tabla de contingencia calcule las siguientes probabilidades condicionales:

i. ¿Cuál es la probabilidad de que la ultrasonografía informe retraso en el crecimiento fetal dado que el bebé resultará con bajo peso?

ii. ¿Cuál es la probabilidad de que la ultrasonografía indique sin retraso en el crecimiento fetal dado que el bebé está sano?

iii. ¿Cuál es la probabilidad de que el bebé nazca con bajo peso dado que la ultrasonografía informó retraso en el crecimiento fetal?

iv. ¿Cuál es la probabilidad de que el bebé nazca sano dado que la ultrasonografía informó sin retraso en el crecimiento fetal?



EJERCICIOS DE REPASO Y AUTOEVALUACIÓN

Esta sección pretende que haga un repaso general de los contenidos del Capítulo.

EJERCICIO 6.11

Encierre en un círculo la opción que se ajusta al enunciado, Falsa (F) o Verdadera (V).

F V 1. Conocer de probabilidades elimina completamente el riesgo y la incertidumbre.

F V 2. La probabilidad de ocurrencia del evento "*A*" se representa como *P*(*A*) y es un número que va del 0 al 1,00 e indica cuán previsible es la ocurrencia del evento "*A*".

F V 3. Los eventos son mutuamente excluyentes cuando pueden tener muchos elementos en común.

F V 4. Los eventos colectivamente exhaustivos siempre agotan todo el espacio muestral.

F V 5. El espacio muestral es igual al 100 % de toda la distribución.

F V 6. Un evento y su complemento son a la vez mutuamente excluyentes y colectivamente exhaustivos.

F V 7. El enfoque clásico considera que los eventos tienen una probabilidad de ocurrencia que se obtiene al dividir los eventos favorables entre todos los eventos del espacio muestral.

F V 8. Afortunadamente solo existen situaciones en las que se puede conocer todo el espacio muestral.

F V 9. El enfoque empírico basa la asignación de probabilidades en registros históricos, es decir en las frecuencias relativas en que se ha presentado un evento.

F V 10. Según el enfoque empírico, la probabilidad viene dada por el porcentaje de veces que un evento se ha presentado históricamente bajo las mismas condiciones.

F V 11. Aunque se alteren drásticamente las condiciones, las probabilidades empíricas seguirán siendo válidas puesto que se basan en muchas observaciones de registros médicos.

F V 12. Según el enfoque subjetivo de probabilidades, cualquier sujeto es capaz de determinar las probabilidades de ocurrencia de un fenómeno.

F V 13. La gran mayoría de las decisiones sociales, estratégicas y políticas tratan con situaciones únicas y muy pocas son cuestiones rutinarias o idénticas, por lo que con mucha frecuencia se tendrá que hacer uso de la probabilidad subjetiva.

F V 14. La actitud, la experiencia, la visión de conjunto y la capacidad técnica ayudarán a minimizar las dificultades para la asignación de probabilidades subjetivas.

F V 15. En caso de que los eventos sean mutuamente excluyentes, la probabilidad de A o B es la suma de cada una de las probabilidades individuales.

F V 16. Si dos eventos no son mutuamente excluyentes, ambos pueden suceder al mismo tiempo, el cálculo deberá entonces tomar en cuenta el hecho de que se presente uno de ellos o ambos.

F V 17. La probabilidad complementaria es igual a 1 menos la probabilidad del evento de interés.

F V 18. Cuando dos eventos son independientes, entonces la probabilidad de que ambos ocurran es igual a la multiplicación de sus probabilidades individuales o marginales.

F V 19. Como regla general, se puede decir que la probabilidad conjunta de dos eventos dependientes es igual a la probabilidad de un evento multiplicada por la probabilidad condicional del otro.

F V 20. Cuando los eventos son estadísticamente dependientes, la probabilidad condicional de obtener el evento B, dado que el evento A ya se ha presentado, es igual a la probabilidad conjunta de los eventos A y B dividida entre la probabilidad marginal de ocurrencia del evento A.

EJERCICIO 6.12

Pareo. Llene los paréntesis con el número que acompaña a cada definición de la izquierda, de forma que cada definición coincida con su concepto. Quedan dos paréntesis en blanco.

1.	Son los eventos que no tienen elementos en común o que no pueden ocurrir al mismo tiempo.	( )	Probabilidad complementaria
2.	Informa de la probabilidad de ocurrencia de un evento, dado que otro se ha presentado previamente y lo afecta.	( )	Evento seguro
3.	Es igual a 1 menos la probabilidad del evento de interés.	( )	Probabilidad conjunta
4.	Es aquel evento cuya probabilidad de ocurrencia es cero.	( )	Enfoque clásico
5.	Es aquel evento cuya probabilidad de ocurrencia es igual a uno.	( )	Enfoque empírico
6.	Cuando la ocurrencia de un evento no afecta en nada a la ocurrencia de otro.	( )	Mutuamente excluyentes
7.	Es la probabilidad de que dos eventos ocurran sucesivamente o al mismo tiempo.	( )	Dependencia
8.	Enfoque donde los eventos tienen una probabilidad de ocurrencia predecible teóricamente.	( )	Enfoque subjetivo
9.	Enfoque en el cual la asignación de probabilidades está basada en la experiencia de las personas que efectúan la estimación.	( )	Experimento
10.	Enfoque que basa la asignación de probabilidades en registros históricos, es decir, en las frecuencias relativas en que se ha presentado un evento.	( )	Independencia
		( )	Evento imposible
		( )	Probabilidad condicional

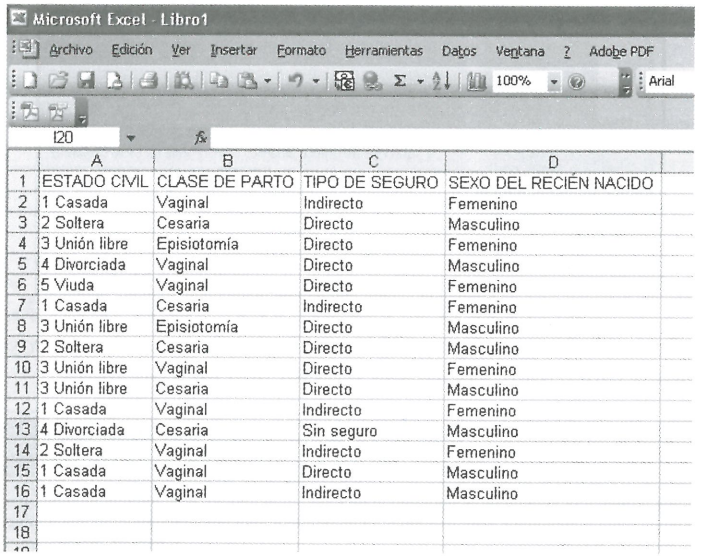
ANEXOS

ANEXO 1

Confección de tablas dinámicas con Excel

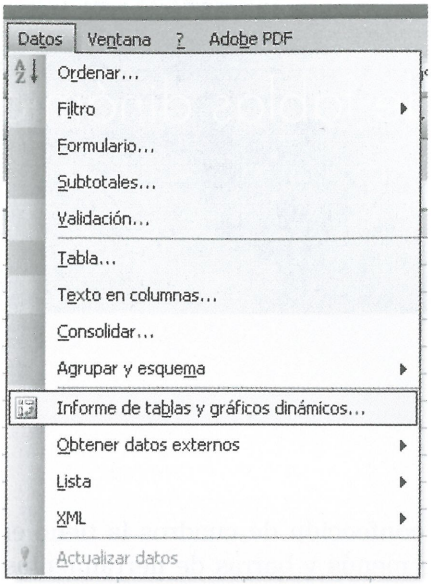
Una herramienta para la confección de cuadros la tiene en Excel de Microsoft. Los formatos de las barras de menús y barras de herramientas han sido modificados de la versión 2003 a la versión 2007 (que se conserva en la versión 2010), para lograr un mejor aprovechamiento, se presentan ambas:

Versión 2003. Teniendo digitados los datos en una hoja de trabajo, en donde cada fila corresponde al registro de un informante y cada columna contiene la característica observada:

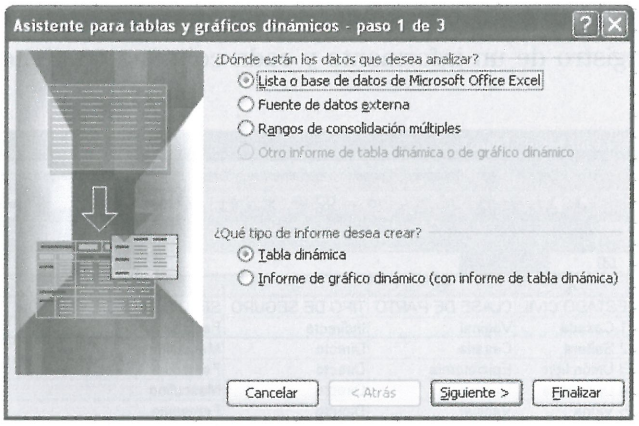


	A	B	C	D
1	ESTADO CIVIL	CLASE DE PARTO	TIPO DE SEGURO	SEXO DEL RECIÉN NACIDO
2	1 Casada	Vaginal	Indirecto	Femenino
3	2 Soltera	Cesaria	Directo	Masculino
4	3 Unión libre	Episiotomía	Directo	Femenino
5	4 Divorciada	Vaginal	Directo	Masculino
6	5 Viuda	Vaginal	Directo	Femenino
7	1 Casada	Cesaria	Indirecto	Femenino
8	3 Unión libre	Episiotomía	Directo	Masculino
9	2 Soltera	Cesaria	Directo	Masculino
10	3 Unión libre	Vaginal	Directo	Femenino
11	3 Unión libre	Cesaria	Directo	Masculino
12	1 Casada	Vaginal	Indirecto	Femenino
13	4 Divorciada	Cesaria	Sin seguro	Masculino
14	2 Soltera	Vaginal	Indirecto	Femenino
15	1 Casada	Vaginal	Directo	Masculino
16	1 Casada	Vaginal	Indirecto	Masculino
17				
18				

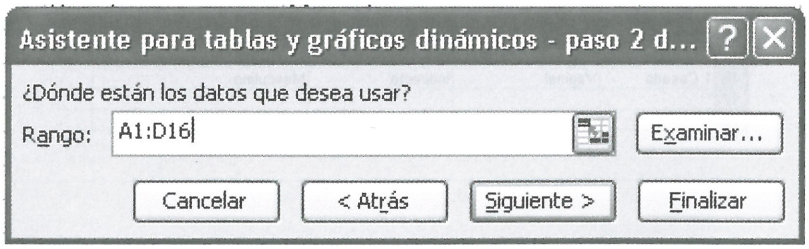
Se abre la ventana *Datos* y se da clic en *Informe de tablas y gráficos dinámicos.*



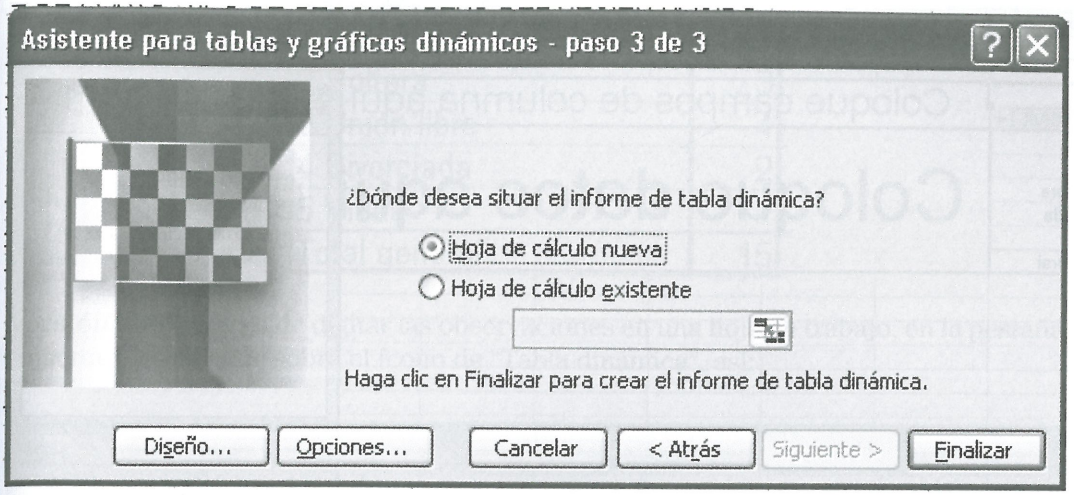
Con lo que se abre un *Asistente para tablas y gráficos dinámicos*:



En el *paso 1 de 3* se señala que hay una *Lista o base de datos de Microsoft Office Excel* y que se desea crear una *Tabla dinámica* y se da clic en *Siguiente.*

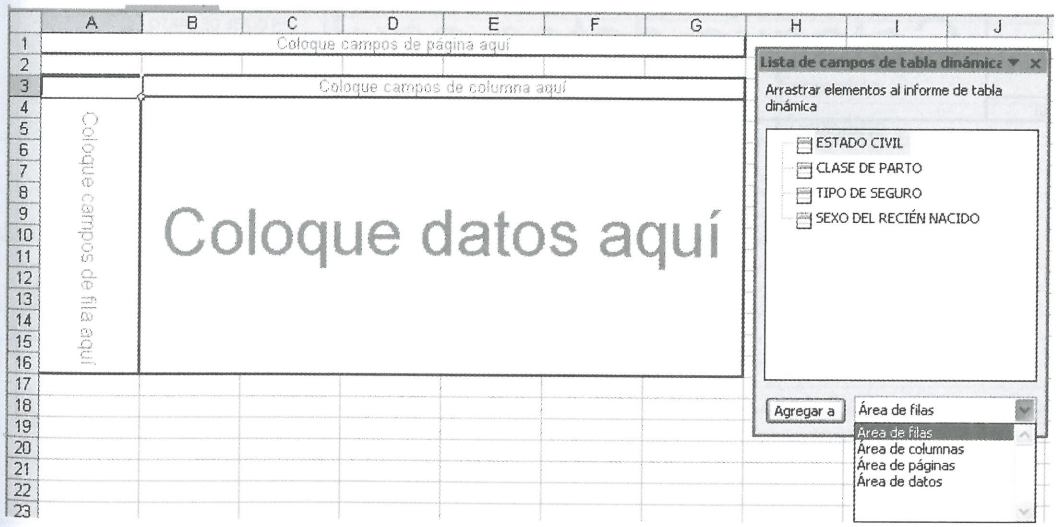


En el *paso 2 de 3* Excel pregunta: ¿Dónde están los datos que se desea usar?, para este ejemplo se digita A1:D16 puesto que en esas celdas se ha digitado previamente los registros de las parturientas y se da clic en *Siguiente.*

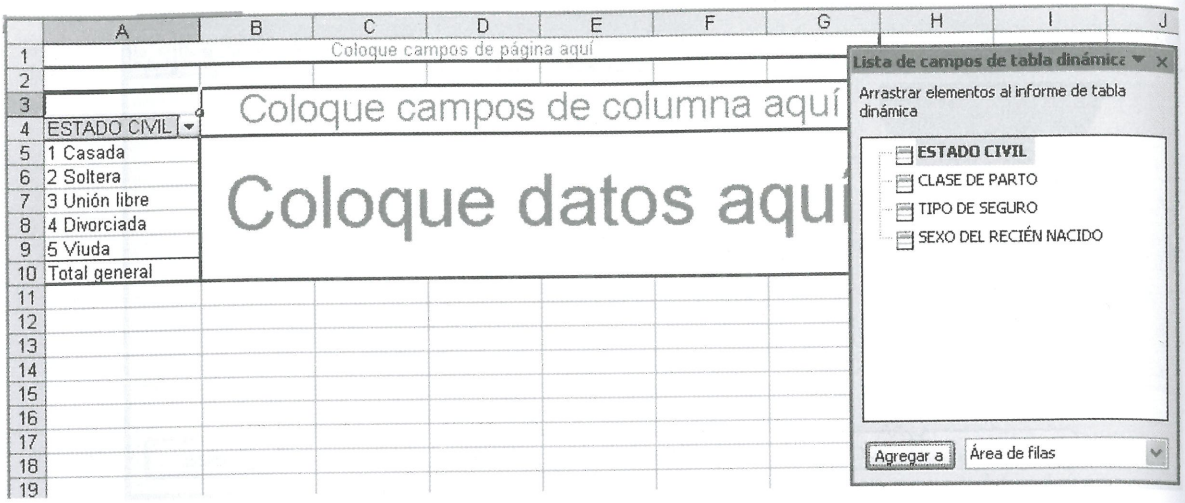


El *paso 3 de 3* pregunta ¿Dónde se desea situar el informe de tabla dinámica? que puede ser una hoja de cálculo nueva o la hoja existente. Escogeremos una *Hoja de cálculo nueva* y se hace clic en *Finalizar.*

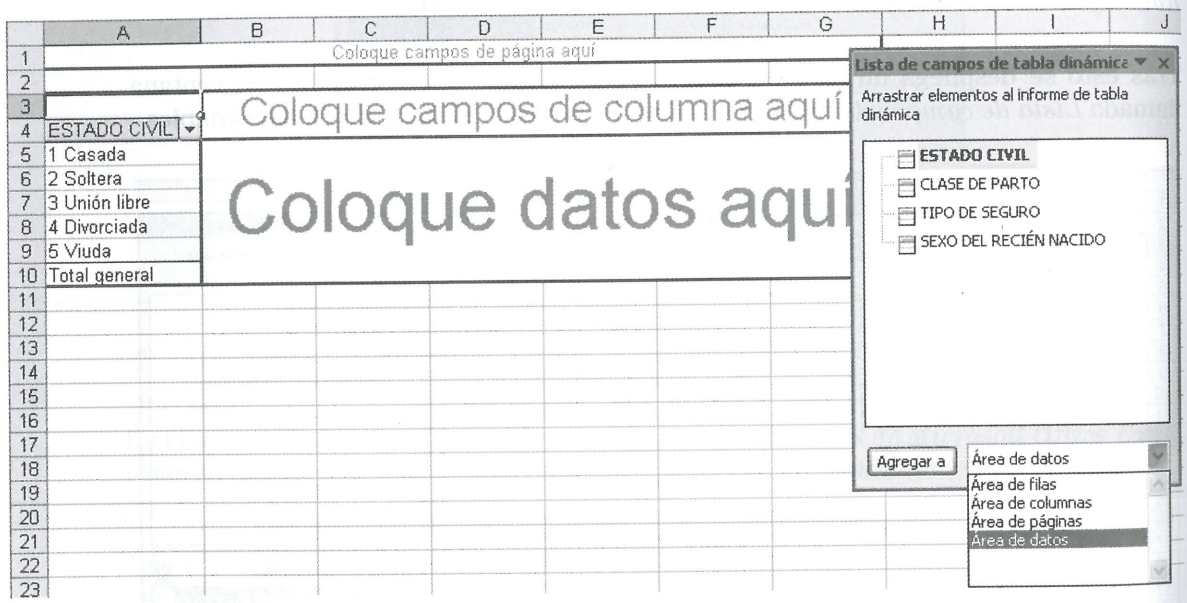
Tras esto se despliega un formato para construir la tabla, además de una ventana llamada *Lista de campos de tabla dinámica* que contiene los rótulos de las variables.



En la ventana *Lista de campos de tabla dinámica* se ilumina ESTADO CIVIL, si se quiere que los montos se indiquen en filas se escoge en el submenú el Área de filas y se da clic en el botón *Agregar a*.



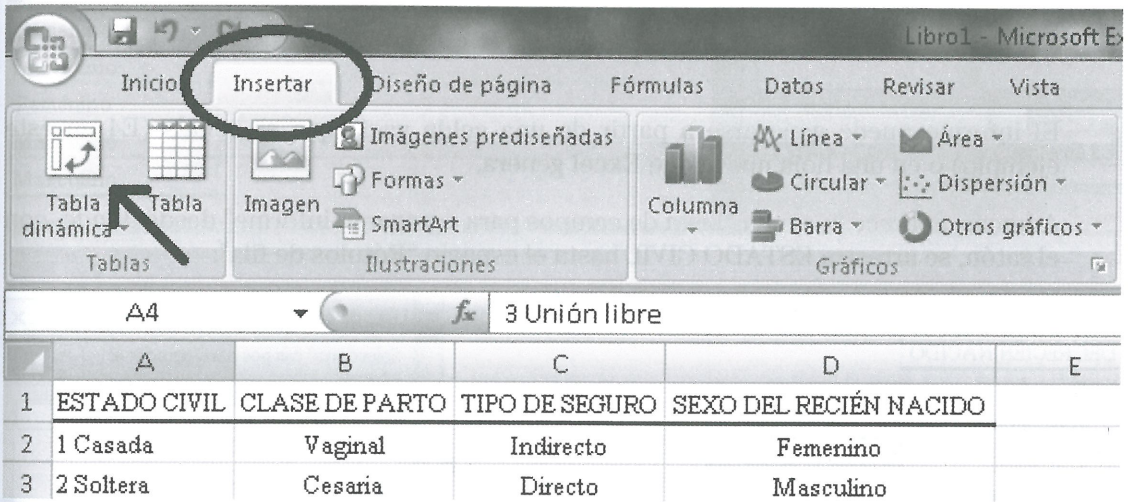
Nuevamente, en la lista de campos se ilumina ESTADO CIVIL, se escoge en el submenú el Área de datos y se da clic en el botón *Agregar a*.



Con lo cual Excel realiza el conteo e informa las frecuencias para cada ESTADO CIVIL.

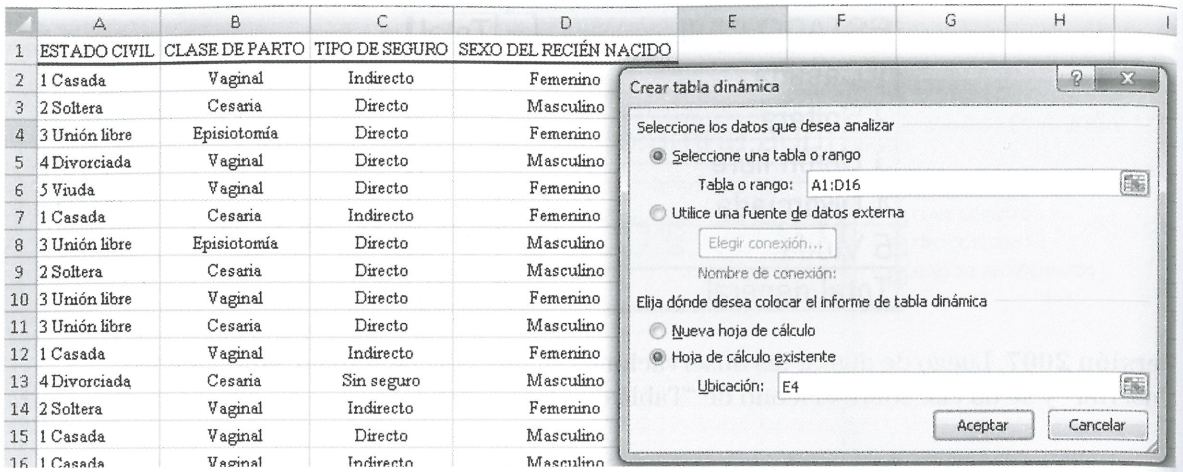
Cuenta de ESTADO CIVIL	
ESTADO CIVIL	Total
1 Casada	5
2 Soltera	3
3 Unión libre	4
4 Divorciada	2
5 Viuda	1
Total general	15

Versión 2007. Luego de digitar las observaciones en una hoja de trabajo, en la pestaña "Insertar" y se da clic sobre el ícono de "Tabla dinámica", así:



	A	B	C	D	E
1	ESTADO CIVIL	CLASE DE PARTO	TIPO DE SEGURO	SEXO DEL RECIÉN NACIDO	
2	1 Casada	Vaginal	Indirecto	Femenino	
3	2 Soltera	Cesaria	Directo	Masculino	

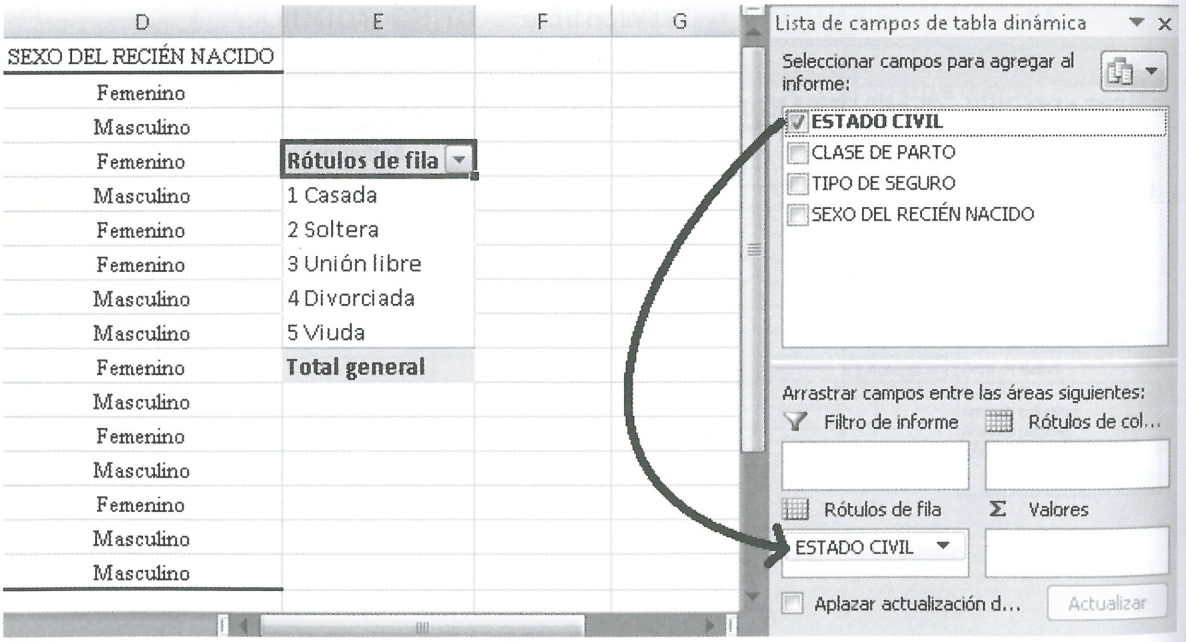
Con lo cual se despliega un diálogo que pregunta por el origen de los datos y en dónde se quiere colocar la tabla dinámica.



	A	B	C	D
1	ESTADO CIVIL	CLASE DE PARTO	TIPO DE SEGURO	SEXO DEL RECIÉN NACIDO
2	1 Casada	Vaginal	Indirecto	Femenino
3	2 Soltera	Cesaria	Directo	Masculino
4	3 Unión libre	Episiotomía	Directo	Femenino
5	4 Divorciada	Vaginal	Directo	Masculino
6	5 Viuda	Vaginal	Directo	Femenino
7	1 Casada	Cesaria	Indirecto	Femenino
8	3 Unión libre	Episiotomía	Directo	Masculino
9	2 Soltera	Cesaria	Directo	Masculino
10	3 Unión libre	Vaginal	Directo	Femenino
11	3 Unión libre	Cesaria	Directo	Masculino
12	1 Casada	Vaginal	Indirecto	Femenino
13	4 Divorciada	Cesaria	Sin seguro	Masculino
14	2 Soltera	Vaginal	Indirecto	Femenino
15	1 Casada	Vaginal	Directo	Masculino
16	1 Casada	Vaginal	Indirecto	Masculino

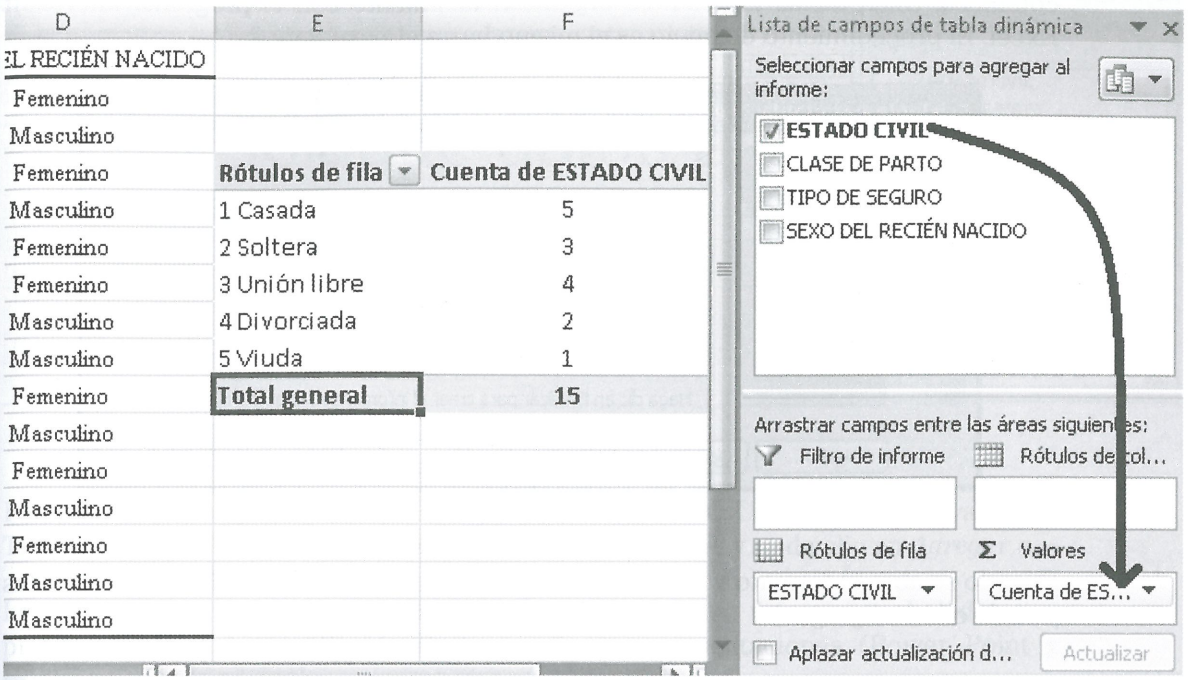
El informe puede generarse a partir de una celda vacía que se señala (E4 en este ejemplo) o en una hoja nueva que Excel genera.

Además, aparece la ayuda "Lista de campos para agregar al informe" desde donde, con el ratón, se arrastra ESTADO CIVIL hasta el espacio "Rótulos de fila":



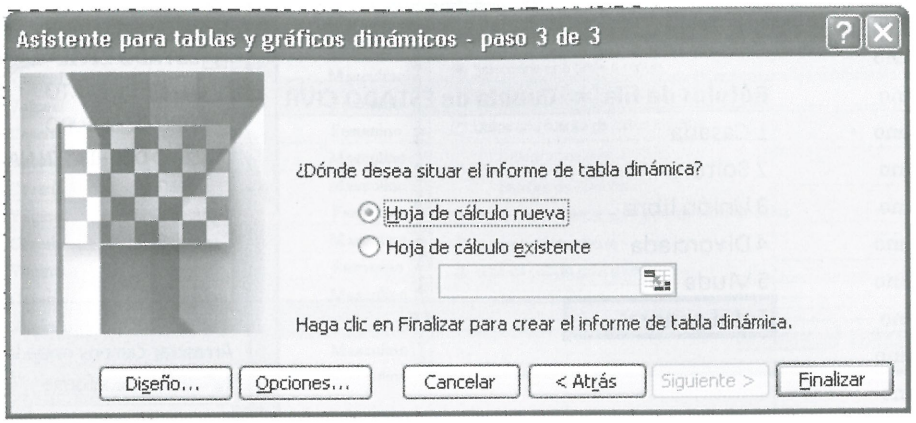
D	E
SEXO DEL RECIÉN NACIDO	
Femenino	
Masculino	
Femenino	Rótulos de fila
Masculino	1 Casada
Femenino	2 Soltera
Femenino	3 Unión libre
Masculino	4 Divorciada
Masculino	5 Viuda
Femenino	Total general
Masculino	
Femenino	
Masculino	
Femenino	
Masculino	
Masculino	

Luego se arrastra con el ratón, ESTADO CIVIL hasta el espacio "Σ Valores" que realizará el conteo generando un informe de tabla de una entrada:

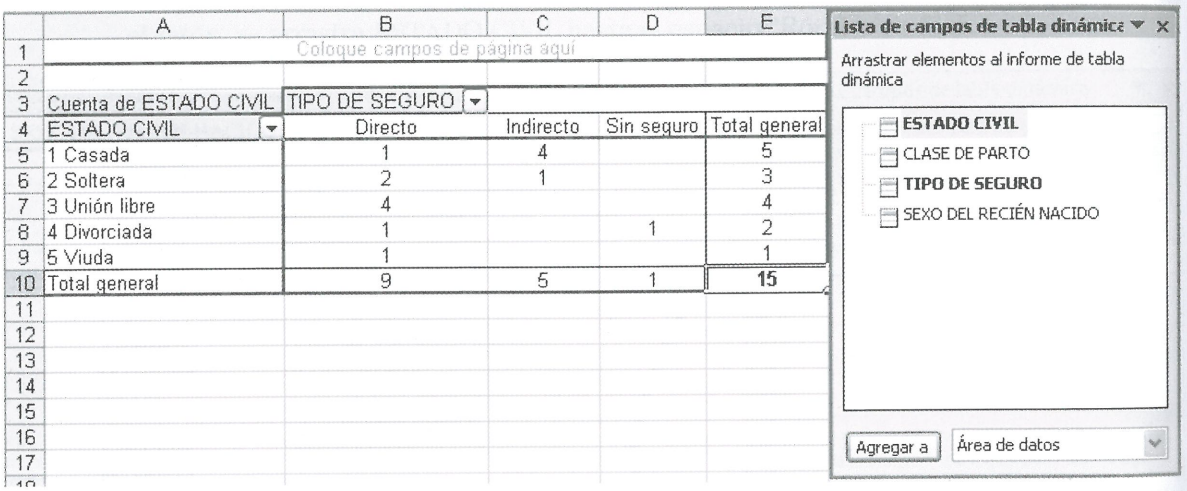


TABLAS DE DOBLE ENTRADA

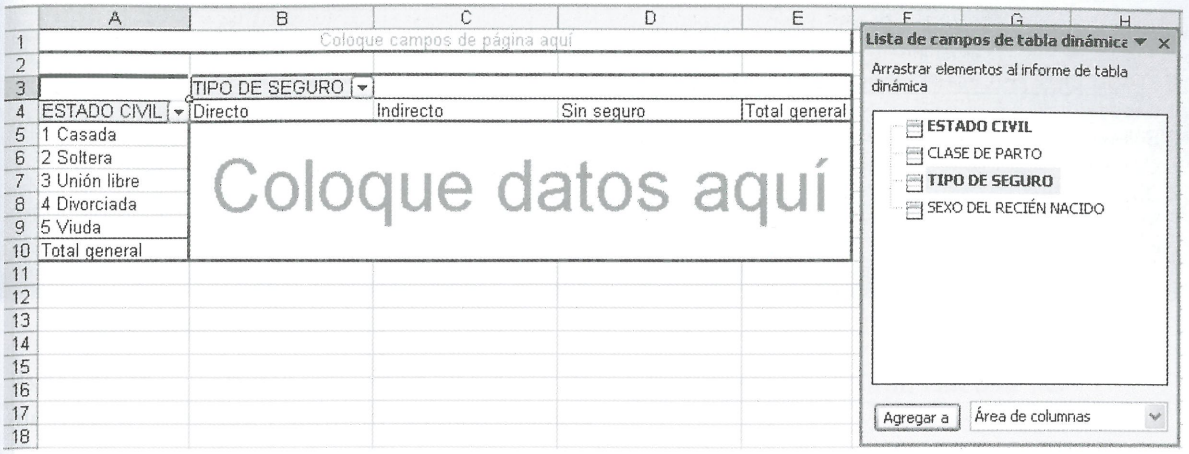
Versión 2003. Las tablas de doble entrada informan de una serie por filas y otra por columnas; estas tablas también se organizan fácilmente con el programa Microsoft Excel. El procedimiento de inicio es el mismo hasta el *paso 3 de 3*:



Cuando se abre el formato y la lista de campos, se señala ESTADO CIVIL, Área de filas y se da clic en *Agregar a*:

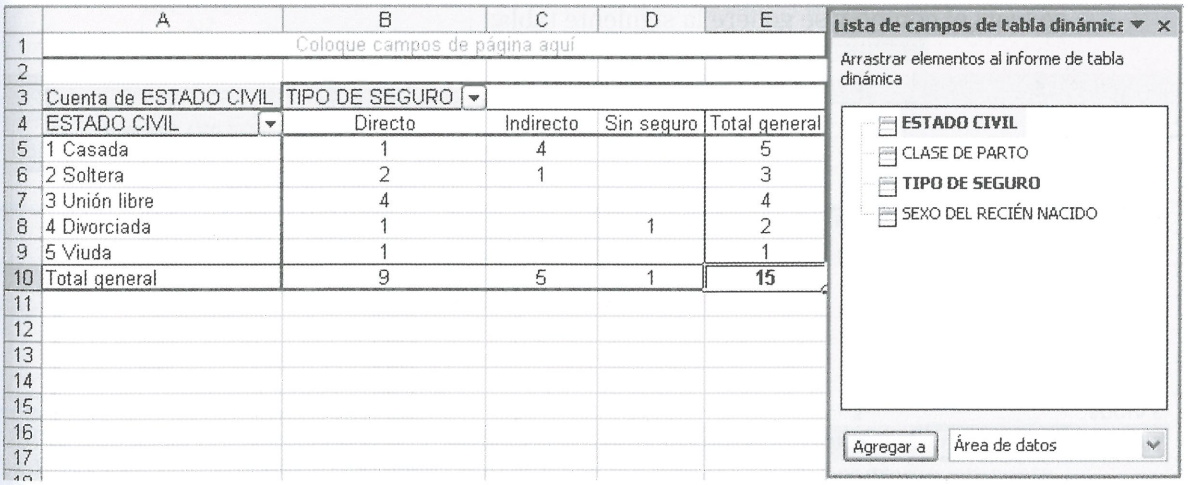


Luego se escoge TIPO DE SEGURO en el Área de columnas y se da clic en *Agregar a*:

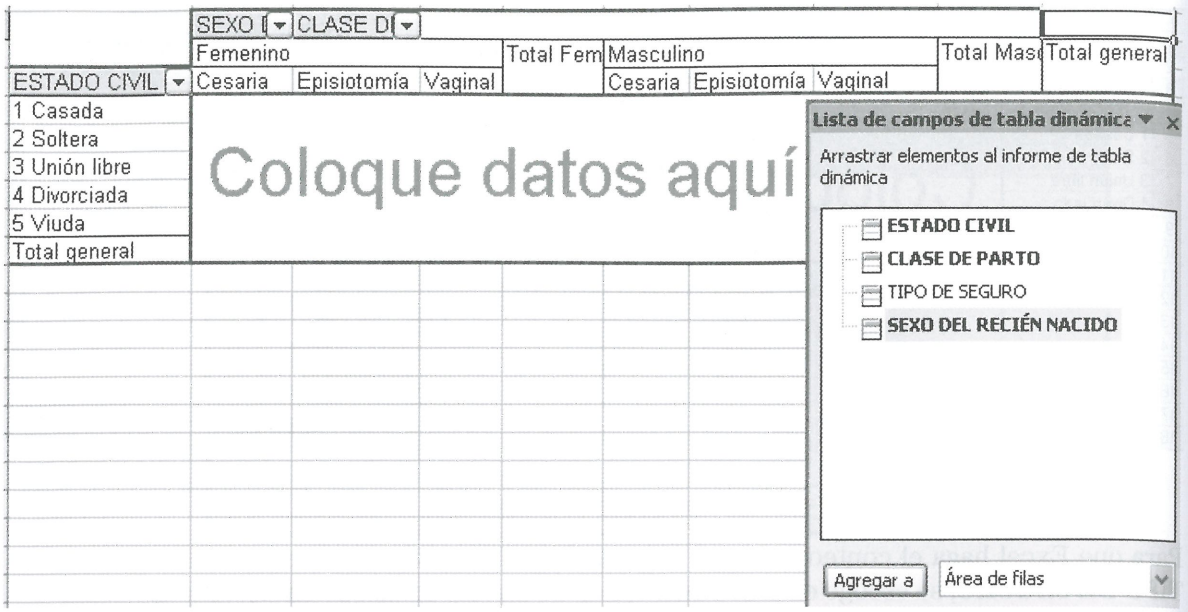


Para que Excel haga el conteo, se debe iluminar ESTADO CIVIL o, alternativamente, TIPO DE SEGURO; se escoge en el submenú Área de datos y se da clic en *Agregar a*.

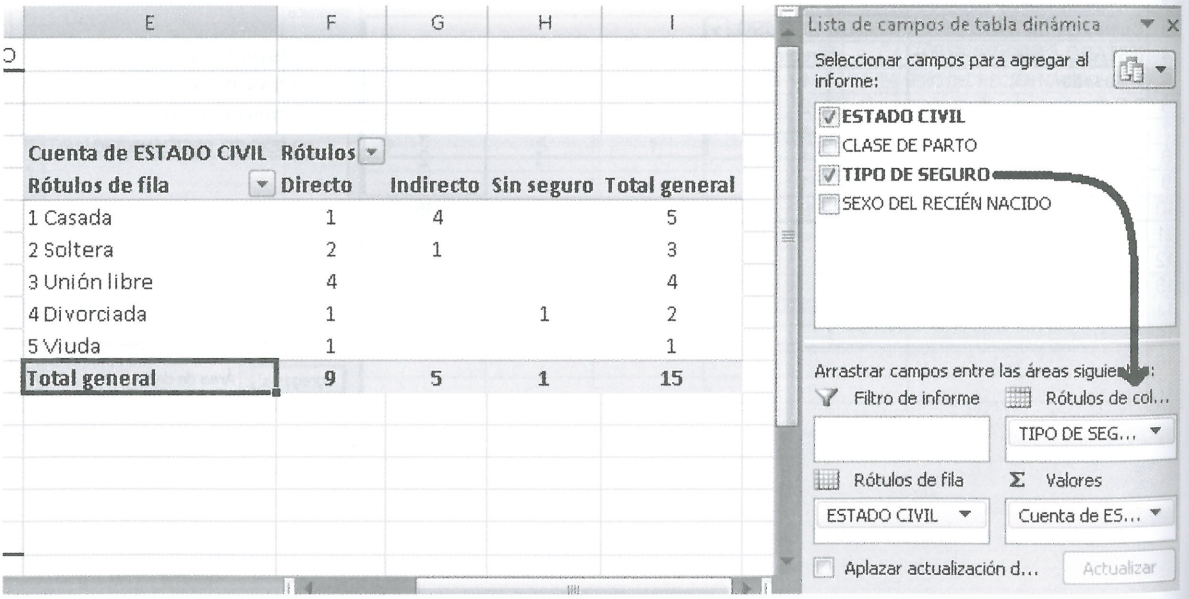
La tabla puede luego ser editada, decorada e incluida en informes trabajados con procesadores de texto (Word, Works y otros) o presentaciones (Power Point y similares).



En forma parecida puede elaborarse tablas de tres o más variables.



Versión 2007. Continuando el ejercicio, una tabla de doble entrada requiere tanto de filas como de columnas, en filas se tiene ya el ESTADO CIVIL, basta entonces con arrastrar el campo TIPO DE SEGURO hasta el espacio "Rótulos de columna", para qué se haga el conteo y se genere la siguiente tabla:



Arrastrando otros campos hacia filas o hacia columnas, puede generarse tablas de tres o más variables.

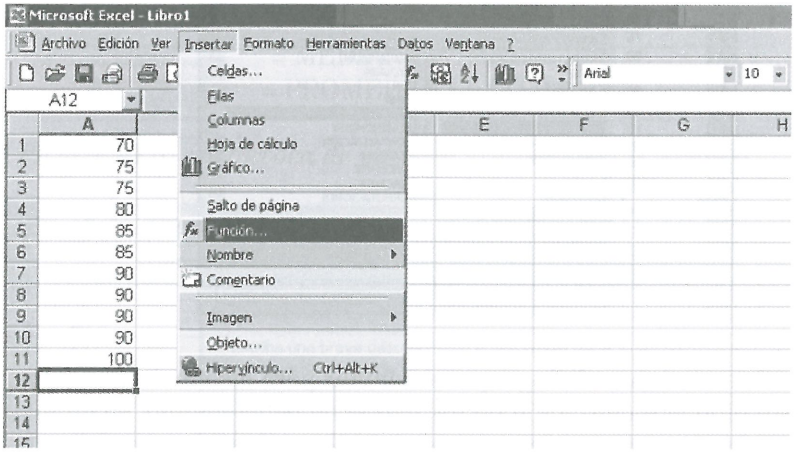
ANEXO 2

Cálculo de medidas de posición y variabilidad con Excel

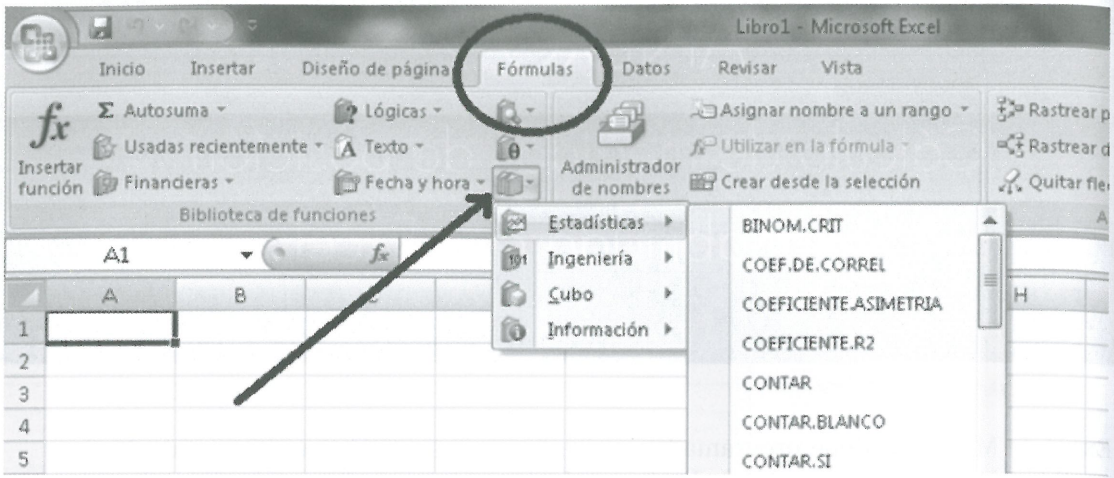
Excel de Microsoft tiene programadas varias funciones que le permitirán calcular las medidas de posición y variabilidad más usuales.

Teniendo digitados los datos en una hoja de trabajo, y estando posicionado el cursor en una celda libre, basta con punzar el ícono f_x para que se despliegue una ventana con varias funciones estadísticas, financieras, matemáticas y otras.

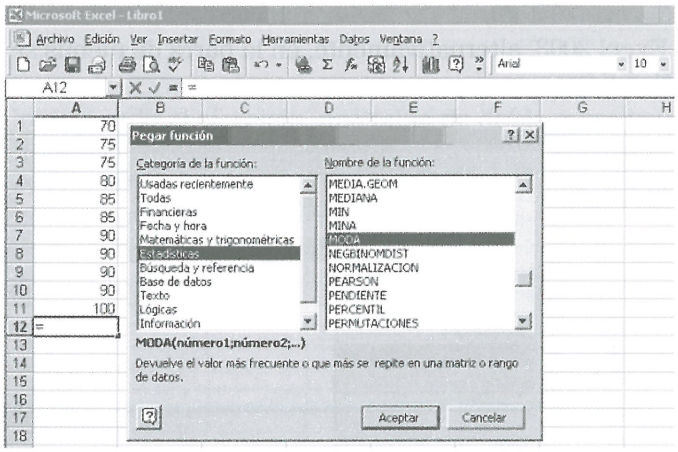
En la versión de Excel 2003, alternativamente, en la barra de menús se busca *Insertar*, donde se punza *Función*.



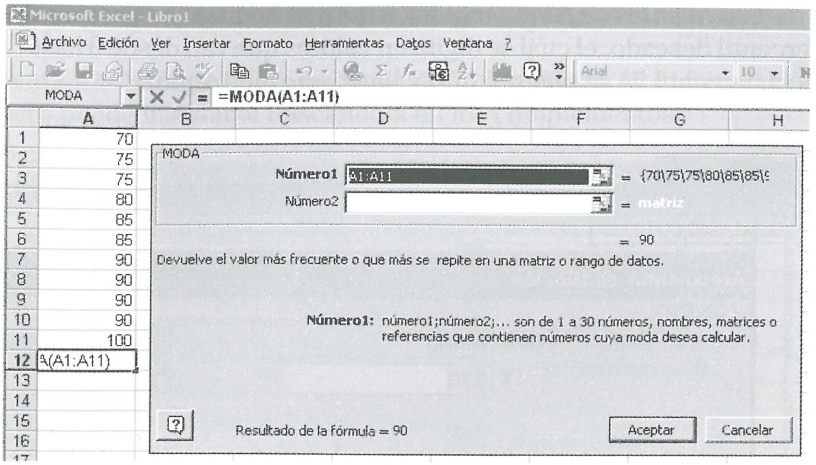
En las versiones de Excel 2007 y 2010, se puede simplemente punzar en el ícono f_x o desplazarse a la pestaña "Fórmulas" y desplegar las funciones. El programa presenta, de entrada, las funciones que sus programadores consideraron de uso más frecuente, además usted encontrará un ícono con más alternativas, entre ellas se escogerá "Estadísticas"



Con ese comando la máquina presenta un *Diálogo* que contiene varias categorías de función, se punza en *Estadísticas* para desplegar un listado de funciones en orden alfabético, para este ejemplo se punzará en *MODA*.



Esta acción abre un nuevo diálogo que es similar para las versiones 2003, 2007 y 2010 de Excel, este pide la localización de los datos; en este ejemplo, se alimenta A1:A11 y se da clic en *Aceptar*. Con esos comandos, la máquina inserta el valor de la moda en la celda en que estaba posicionado el cursor.



Otra manera de llegar al mismo resultado es, primero escoger una celda libre en la cual se insertará el dato buscado, luego se digita:

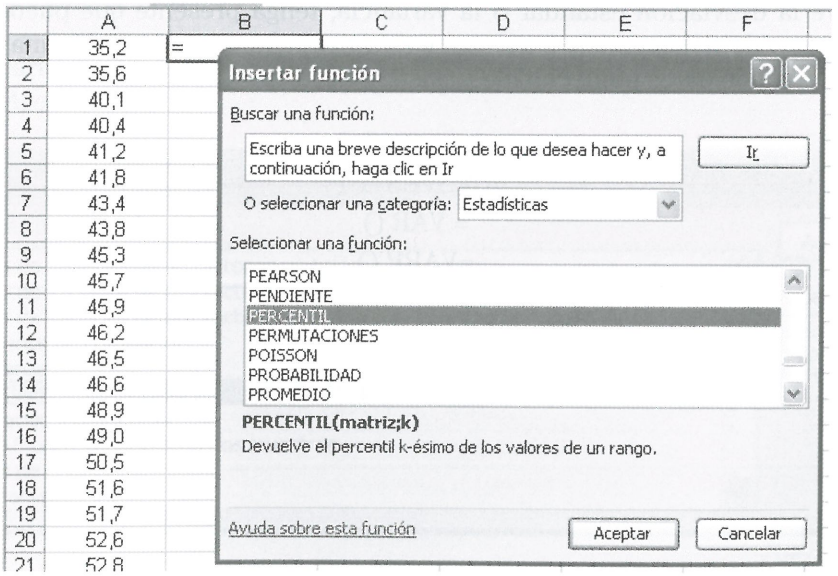
= MODA(A1:A11)

Tras dar *Enter*, el valor modal de 90 se insertará en la celda escogida.

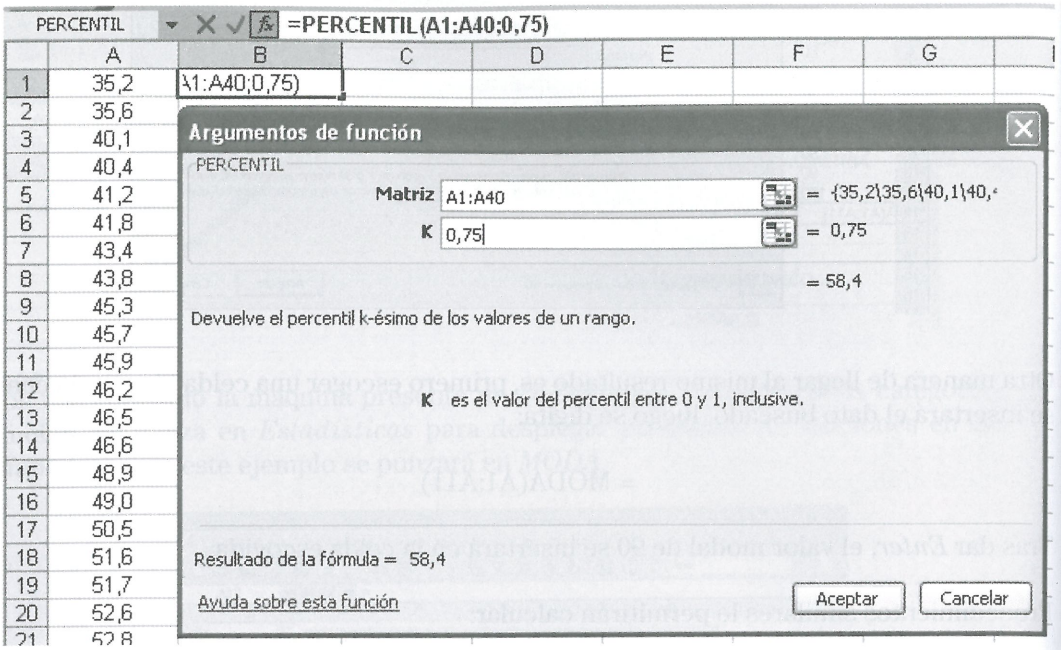
Procedimientos similares le permitirán calcular:

= MEDIANA(A1:A11)
= PROMEDIO(A1:A11)

La función PERCENTIL se convoca de manera similar, y tras Aceptar*:*



El diálogo de *Argumentos de función* pide en *Matriz* el rango en donde están los datos y en *K* el percentil deseado, el cual se digita en valores que van desde 0 hasta 1 inclusive. Si se busca el percentil 75 se digitará 0,75 y luego clic en *Aceptar*.



Al calcular las medidas de dispersión, la desviación media deberá buscarla como:

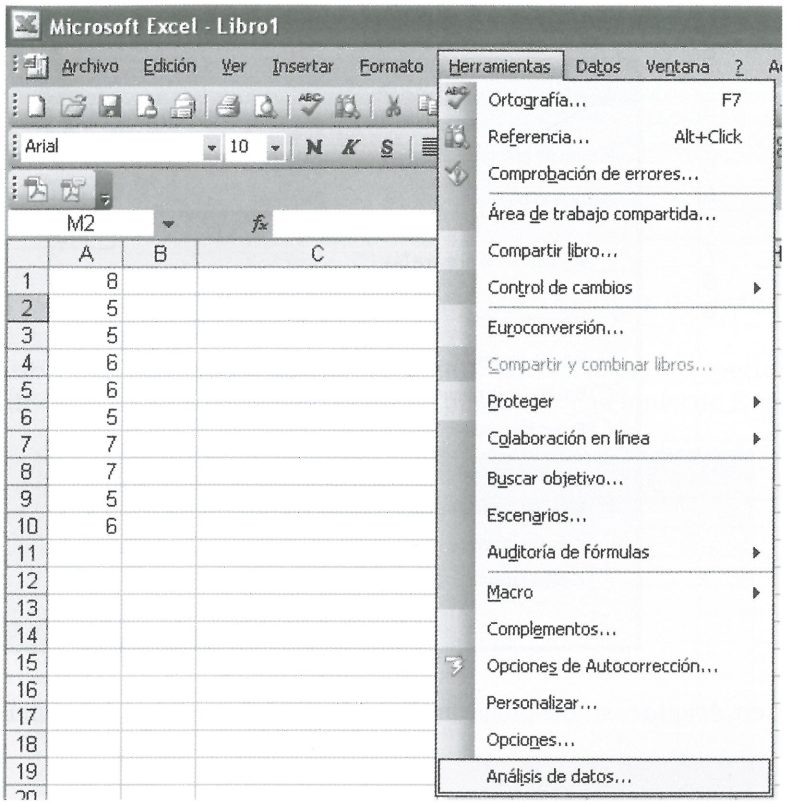
=DESVPROM ()

Si requiere la desviación estándar o la variancia, tenga presente que puede tratarse de muestras o de poblaciones; para hacer esta distinción Excel coloca una P al final cuando se trata de datos poblacionales, así:

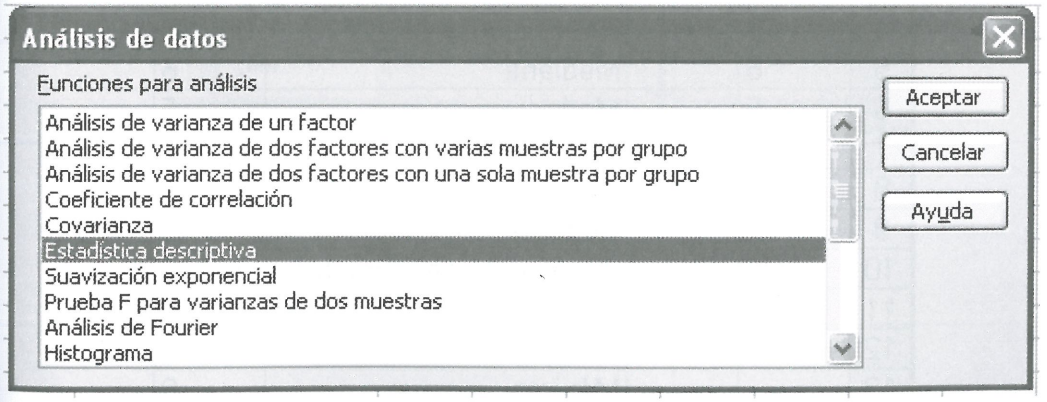
=DESVEST ()
=DESVESTP ()
= VAR ()
=VARP ()

CÁLCULO UTILIZANDO HERRAMIENTAS DE ANÁLISIS

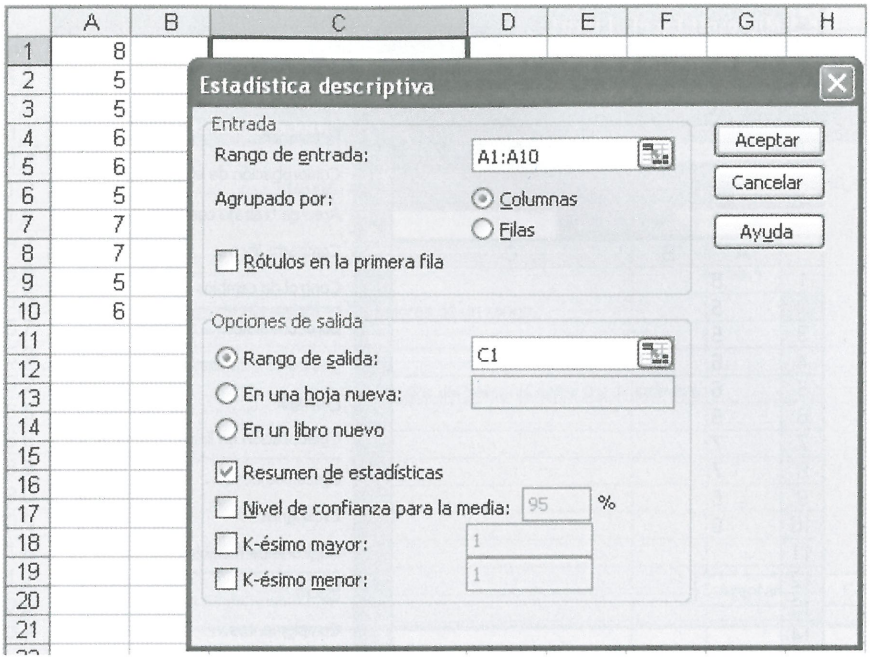
La herramienta *Análisis de datos* no viene predeterminada en la instalación estándar de Excel, pero puede instalarse buscándola en los complementos.



Al convocarla, se despliega un diálogo, en el cual se escoge *Estadística descriptiva* y se punza *Aceptar*.



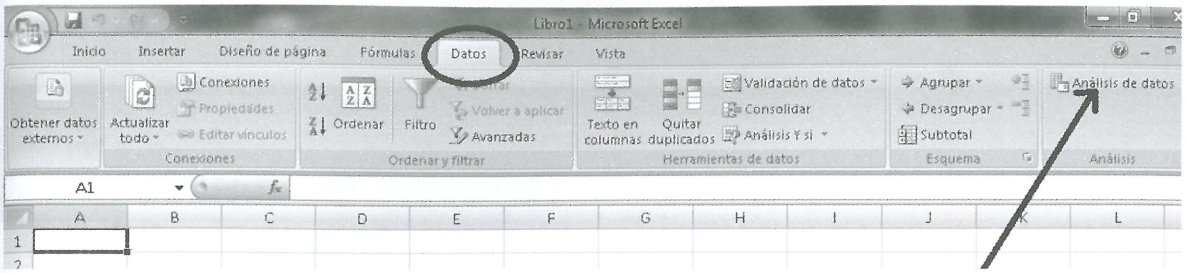
Al abrir un nuevo diálogo, en *Rango de entrada* se digitan las celdas en donde están los datos, en este ejemplo A1:A10, en la opción *Agrupado por* se escoge columnas o filas. Si existieran *Rótulos en la primera fila* se marcan. Luego se señala alguna opción de salida y se marca *Resumen de estadísticas*.



Tras punzar en *Aceptar*, se desplegarán las principales medidas de posición, forma y variabilidad:

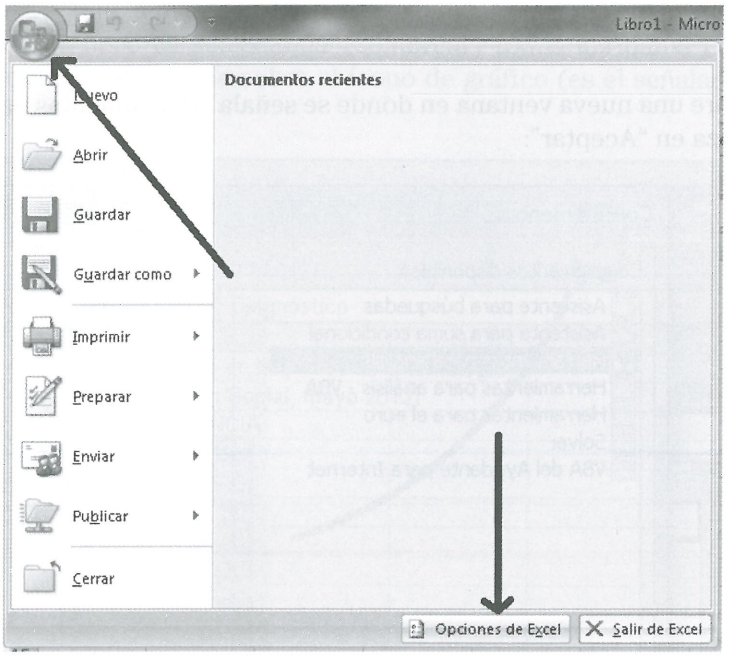
	A	B	C	D
1	8		*Columna1*	
2	5			
3	5		Media	6
4	6		Error típico	0,3333
5	6		Mediana	6
6	5		Moda	5
7	7		Desviación estándar	1,0541
8	7		Varianza de la muestra	1,1111
9	5		Curtosis	-0,45
10	6		Coeficiente de asimetría	0,7115
11			Rango	3
12			Mínimo	5
13			Máximo	8
14			Suma	60
15			Cuenta	10

Versión 2007. El cambio más significativo está en la ubicación de "Análisis de datos". Este se buscará en la pestaña "Datos" y, en la esquina superior derecha, se punza para convocar los diálogos de alimentación de datos.

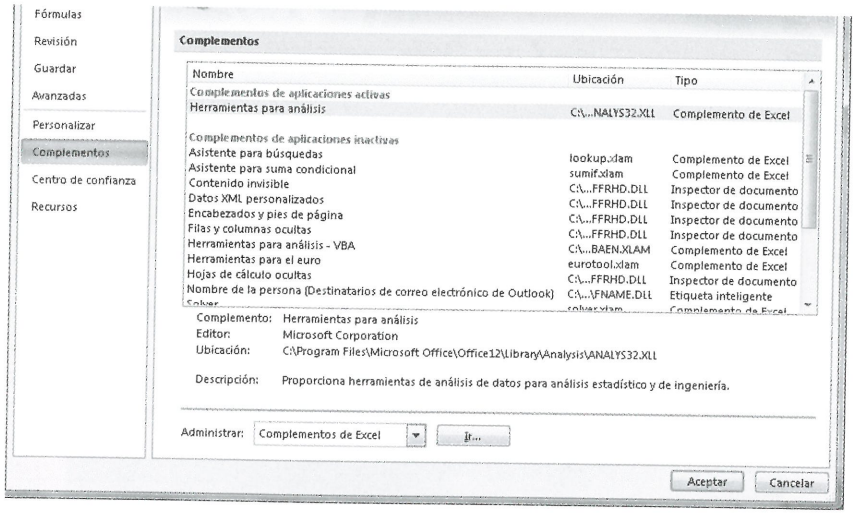


¿Qué hacer si no tiene instalado el complemento "Análisis de datos"?

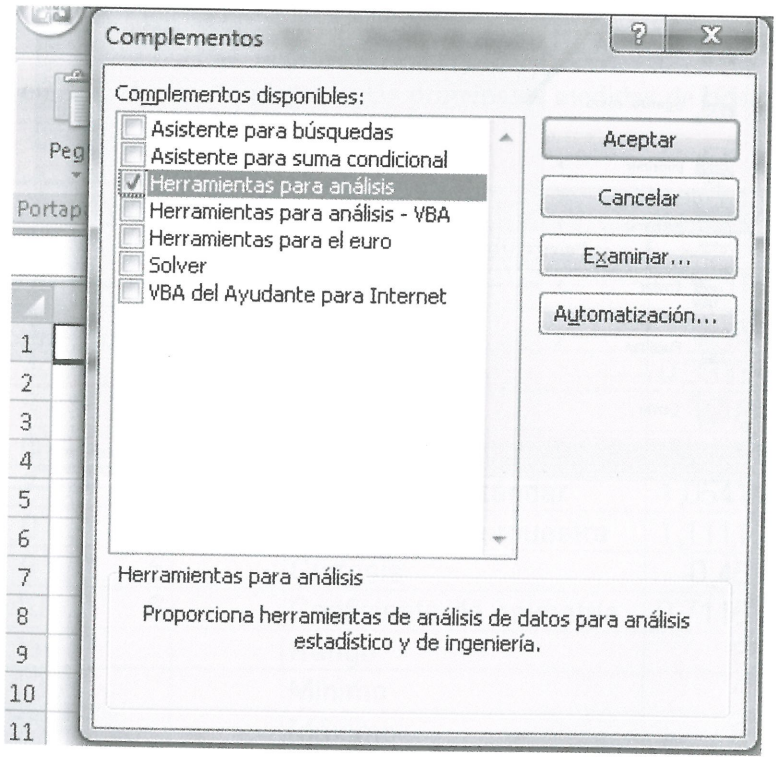
Para instalarlo, se da clic en el botón de inicio (ubicado en la parte superior izquierda). Esa acción despliega una ventana en la cual se punza en "Opciones de Excel"



Esto abre una nueva ventana en la cual se punza "Complementos" (ubicado en la columna de la izquierda), luego se ilumina "Herramientas para el análisis" y se punza en "Ir..."



Esta acción abre una nueva ventana en donde se señala "Herramientas para el análisis" y luego se punza en "Aceptar":



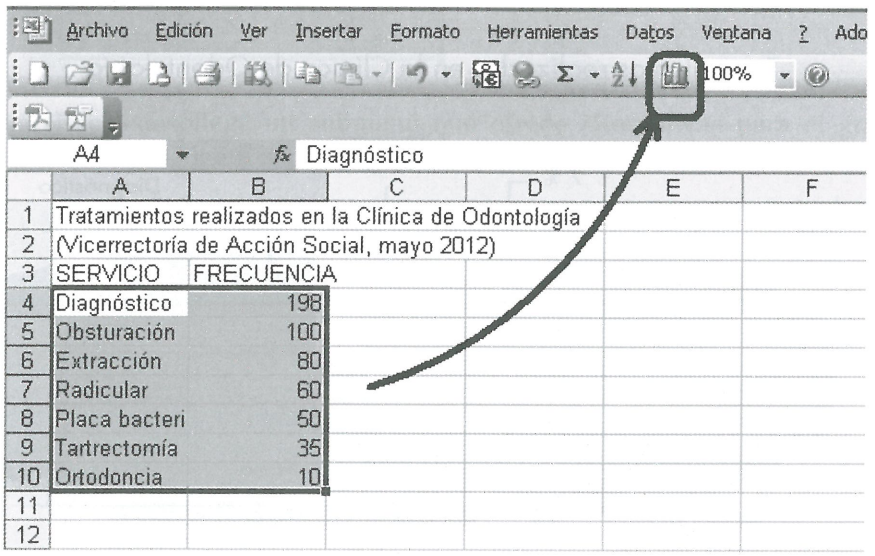
Tras esperar unos segundos por la instalación, busque el complemento en la pestaña "Datos". El resto es muy similar en las tres versiones de Excel (2003, 2007 y 2010).

ANEXO 3

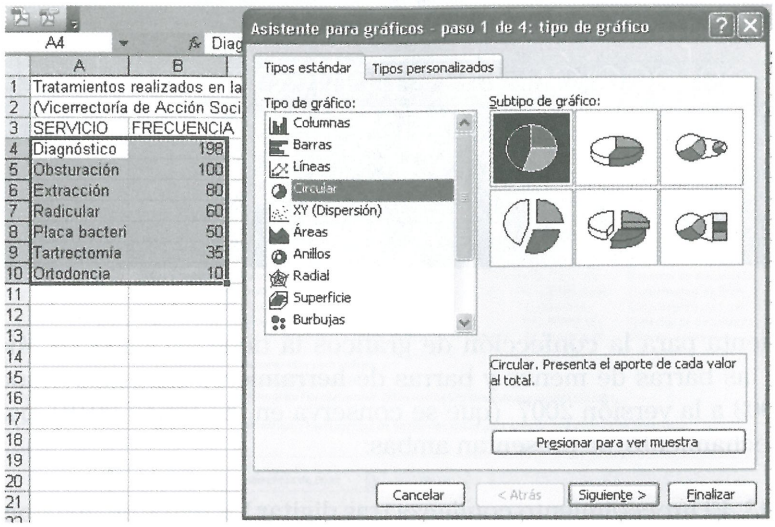
Confección de gráficos con Excel

Una herramienta para la confección de gráficos la tiene en *Excel* de Microsoft. Los formatos de las barras de menús y barras de herramienta s han sido modificadas de la versión 2003 a la versión 2007 (que se conserva en la versión 2010), para lograr un mejor aprovechamiento, se presentan ambas:

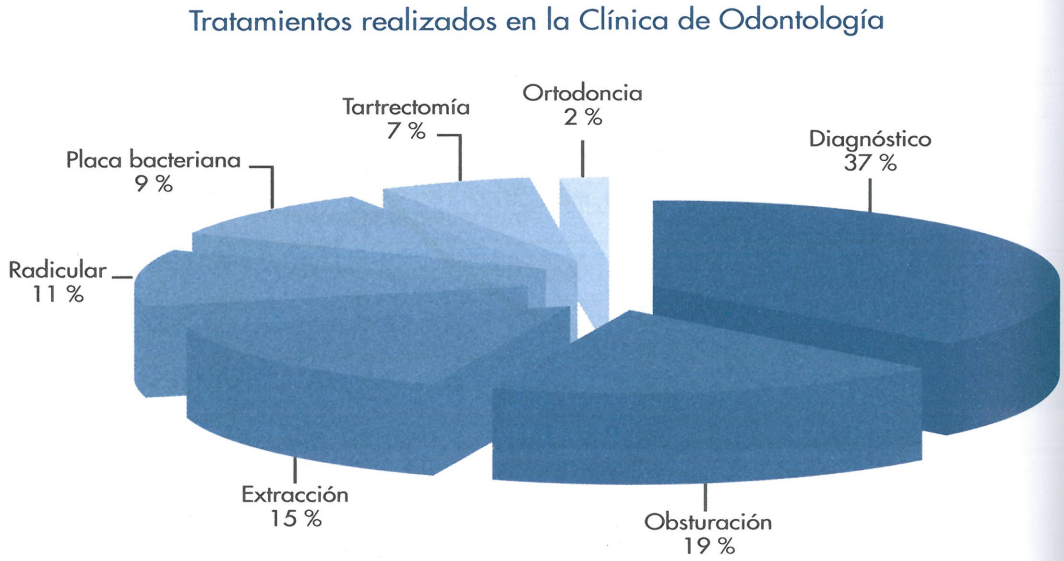
Versión 2003. El procedimiento comienza tras digitar los datos en una hoja de trabajo. Lo primero que se hace es seleccionar (sombrear) todos los datos con los que se elaborará el gráfico y se da clic sobre el ícono de gráfico (es el señalado en la figura siguientw, así:



Con ello se despliega una ventana con el *Asistente para gráficos* que contiene cuatro pasos muy amigables de seguir:

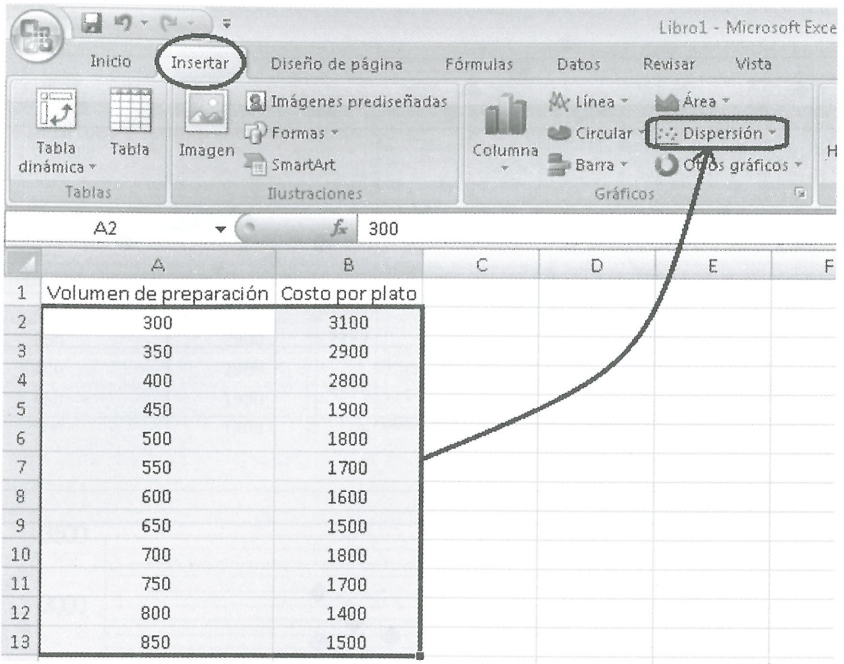


Para este ejemplo se selecciona el tipo de gráfico circular. Ayudado por el asistente se agregan detalles como rótulos, datos, títulos y otros, que permiten generar con mucha facilidad gráficos como el siguiente:



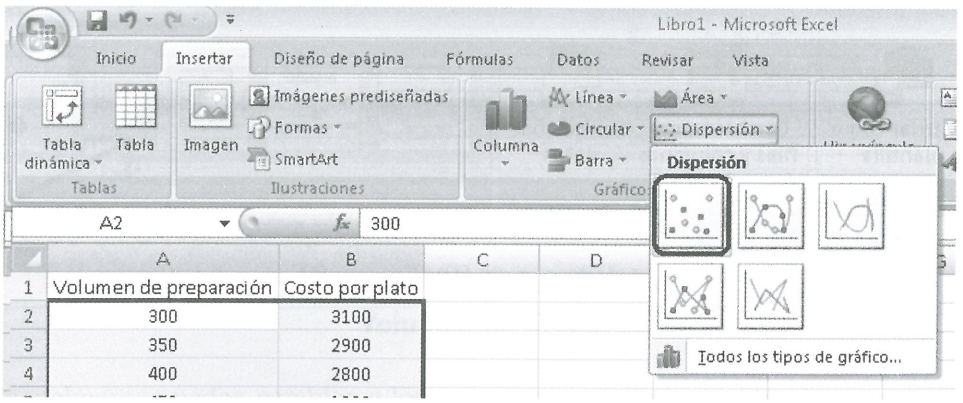
Fuente: Datos hipotéticos.

Versión 2007. Luego de digitar las observaciones en una hoja de trabajo, se seleccionan (sombrean) todos los datos con los que se elaborará el gráfico; se escoge la pestaña "Insertar" y se da clic sobre el ícono del gráfico por utilizar (para este ejemplo se elaborará un gráfico de dispersión), así:



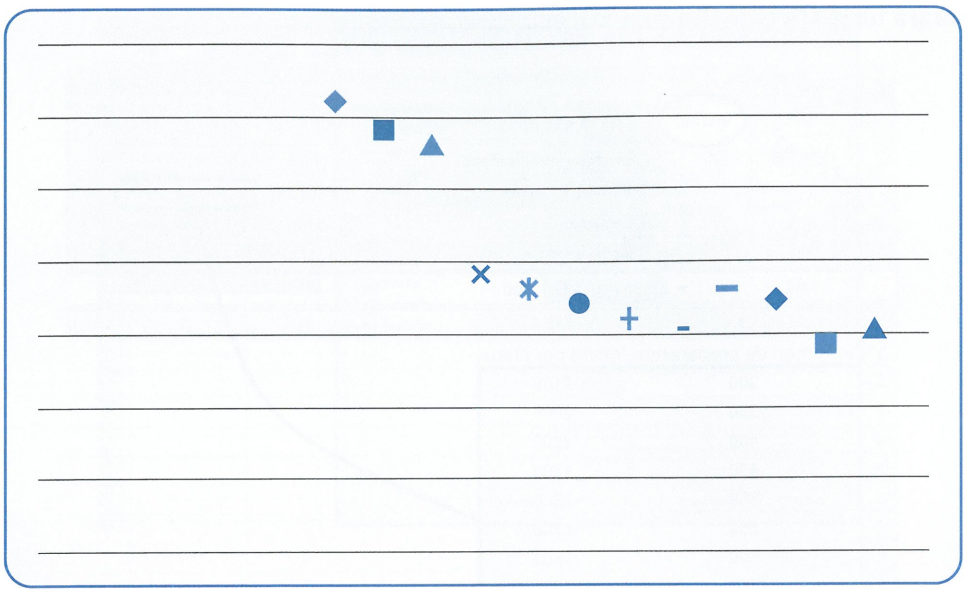
	A	B
1	Volumen de preparación	Costo por plato
2	300	3100
3	350	2900
4	400	2800
5	450	1900
6	500	1800
7	550	1700
8	600	1600
9	650	1500
10	700	1800
11	750	1700
12	800	1400
13	850	1500

Al punzar, Excel despliega un submenú que ofrece alternativas para el gráfico de dispersión, aquí se escogerá la primera opción:

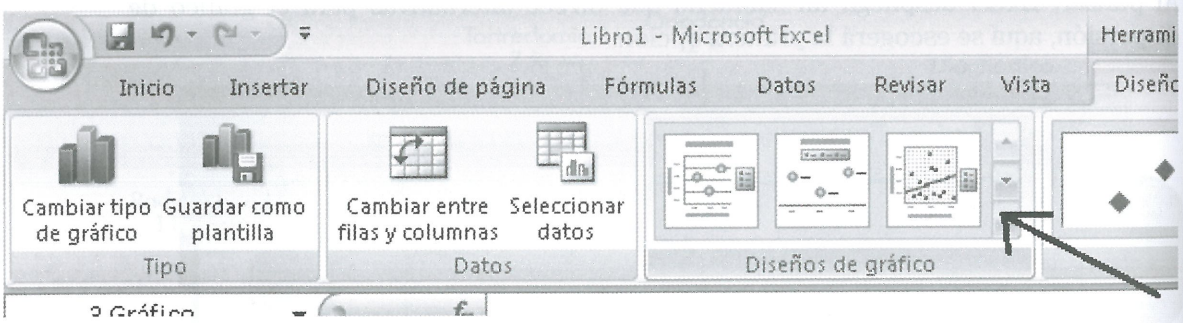


	A	B
1	Volumen de preparación	Costo por plato
2	300	3100
3	350	2900
4	400	2800

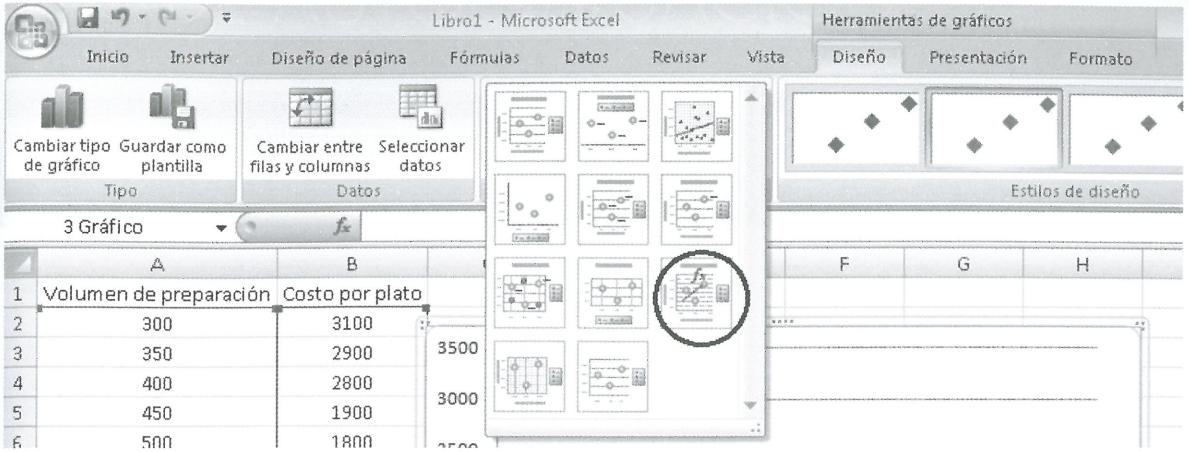
Con lo cual se presenta una propuesta de gráfico:

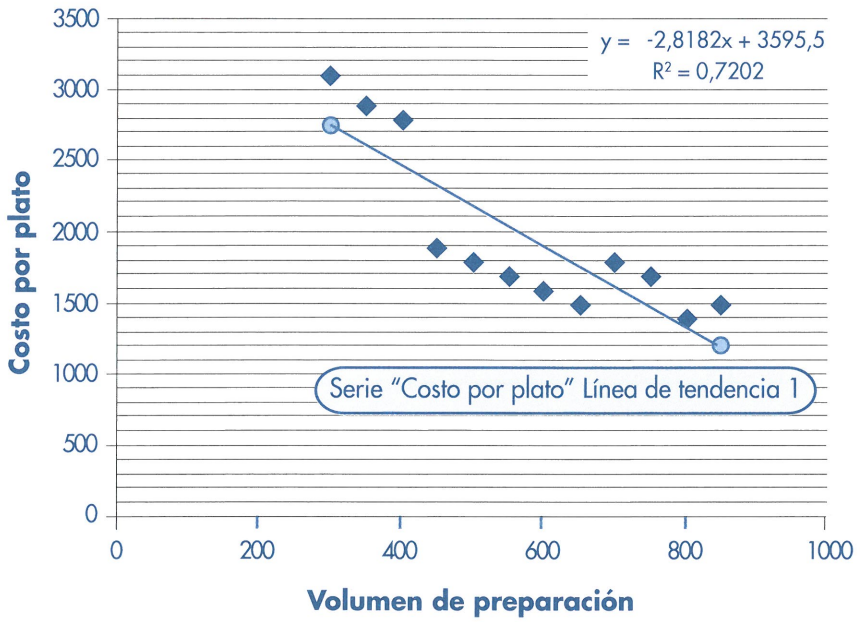


Luego, desplegando el submenú "Diseños de gráfico"



Se presentan variaciones sobre el modelo básico. Aquí se escogerá la novena opción pues, entre otras cosas, determina la ecuación de la recta explicativa de la tendencia de los datos y el coeficiente de determinación.





Adicionalmente pueden cambiarse los colores y variar las presentaciones a gusto del investigador.

RESPUESTAS A LOS EJERCICIOS DE AUTOEVALUACIÓN

CAPÍTULO 1

EJERCICIO 1.1

Referencia a la ciencia	Estadística es la ciencia
Objeto de estudio	que trata del desarrollo y aplicación de las técnicas de
Tareas que efectúa	recolección, organización, presentación, análisis e interpretación de datos numéricos
Métodos que aplica	obtenidos por observación o experimentación
Fin (¿para qué?)	para obtener conclusiones válidas y fundamentar la toma de decisiones.

EJERCICIO 1.2

	Palabras Clave	Descriptivo	Inferencial
a)	Señala una tendencia, hace un pronóstico de la esperanza de vida		X
b)	Gráfico informativo, presenta porcentajes, describe modo de transmisión	X	
c)	Usa muestra e infiere para toda la población, calcula riesgo, establece límites probables		X
d)	Calcula tasa de prevalencia de diabetes, indica porcentajes	X	
e)	Se apoya en fuentes diversas, se basa en muestras y lo afirma para toda la población		X
f)	Hace un pronóstico, establece probabilidades a partir de datos conocidos		X

EJERCICIO 1.3

	Palabras clave	Transversal	Longitudinal
a)	Embarazadas en agosto 2011	X	
b)	10 años de observación		X
c)	20 años de observación		X
d)	Muestra en un determinado momento	X	
e)	Muestra de n personas en un momento dado	X	
f)	5 años de observación		X

EJERCICIO 1.4

	Palabras clave	Descriptivo	Analítico
a.	Detalla una situación, describe	X	
b)	Establece relaciones de causa-efecto y correlaciona datos		X
c)	Conocer una situación, describirla	X	
d)	Correlaciona características, establece conexiones		X
e)	Comprueba causalidad, establece consecuencias		X
f)	Describir al paciente bajo riesgo	X	

EJERCICIO 1.5

	Palabras clave	Individuales	Ecológicos
a.	Estudio de una comunidad		X
b)	Un caso específico	X	
c)	Se analiza a la población		X
d)	Se documentan muchos casos		X
e)	Un caso específico	X	
f)	Se registra a una muestra de operados y se calculan tasas para la población		X

EJERCICIO 1.6

	Elementos clave	Obs.	Exp.	Cuasi-exp
a.	Es un experimento "natural" pues cada comunidad tiene diferente distancia y sentido de los vientos, pero las variables no son manipuladas por el estudioso			X
b)	No hay interferencia del investigador, solamente observa y mide	X		
c)	Hay una alteración intencional en un grupo y existe un grupo de control para comparar resultados		X	
d)	Es un experimento "natural" puesto que los patrones de dieta de las comunidades obedecen a criterios diferentes de la manipulación de las variables por parte de los nutricionistas			X
e)	No hay alteración de tratamientos, solamente observaciones por medio de chequeos médicos	X		

EJERCICIO 1.7

		Finita	Infinita
a.	Niños a los que se debe aplicar la vacuna de la poliomielitis		X
b)	Quemados atendidos en el Hospital de Niños en diciembre de 2012	X	
c)	Producción de ácido-acetil-salicílico en los laboratorios de la CCSS		X
d)	Pacientes con dengue hemorrágico atendidos en el Hospital de la Anexión durante el invierno de 2010	X	
e)	Especialistas en Imágenes Médicas graduados por el convenio CCSS-UCR durante la primera década del siglo XXI	X	
f)	Personal de obstetricia graduado en Costa Rica en los últimos 5 años	X	
g)	Funcionarios participantes en la última huelga médica, declarada ilegal por los tribunales de trabajo	X	
h)	Estudiantes que se matricularán en Microbiología en los próximos años		X
i.	Socorristas voluntarios que actualmente se encuentran acreditados por la Cruz Roja	X	

EJERCICIO 1.8

ATRIBUTOS

Característica de interés	Unidad de medida	Ejemplo de observación
Estado civil	- - -	Casada
Tipo de Bombero		Asalariado
Grado		Teniente

VARIABLES DISCRETAS

Característica de interés	Unidad de medida	Ejemplo de observación
Número de hijos	Hijo	1 hijo
Número de fracturas	Fractura	1
Inhalación de gases	Eventos sufridos	3
Caídas de más de 2 metros	Eventos sufridos	1
Quemaduras	Eventos sufridos	2

VARIABLES CONTINUAS

Característica de interés	Unidad de medida	Ejemplo de observación
Edad *	Años	3 años
Tiempo servido	Años	24 años
Peso	Kilogramos	57 kilogramos
Estatura	Centímetros	168 centímetros

* La naturaleza de estas variables es continua, aunque convencionalmente se reducen a años cumplidos, kilogramos y centímetros completos.

VARIABLES DICOTÓMICAS

Característica de interés	Unidad de medida	Ejemplo de observación
Sexo	---	Femenino
Hipertensión		No
Diabetes		No
Enfermedades respiratorias		No

EJERCICIO 1.9

	Claves	Intencional	Conveniencia	Aleatoria
a.	Identifica informantes que pueden aportar mejores datos por trabajar en emergencias	X		
b)	Todos los preseleccionados pueden ser escogidos en el sorteo			X
c)	Los concesionarios tienen igual probabilidad de ser escogidos en la muestra			X
d)	Se cita a los informantes por comodidad y cercanía de la Sede Central		X	
e)	Los visitantes manejan Internet y buscan alguna información o trámite en el Ministerio de Educación, además participan voluntariamente		X	
f)	La muestra se escoge al azar. Este es un tipo de muestreo que primero divide la población en dos estratos (administrativos y médicos) y luego hace una selección sistemática de los expedientes			X

EJERCICIO 1.10

Explicación sobre el posible origen del sesgo.

Para analizar la efectividad de una dieta se analizan los resultados de todas las personas que completaron el programa.	Sesgo de selección	La mayoría de los que abandonaron el programa no estaban obteniendo los resultados esperados. Los que se quedaron están motivados y se espera que tengan mejores resultados.
Se quiere comprobar la eficacia de un tratamiento para dejar de fumar y para ello se llama a voluntarios fumadores.	Sesgo de selección	Ser voluntario ya anuncia una predisposición a dejar de fumar y los voluntarios pueden ver en el tratamiento una oportunidad gratuita.
En un experimento participan dos grupos, uno expuesto a la variable estudiada y otro de control. El investigador principal hace personalmente las mediciones en el grupo experimental y manda a sus asistentes a hacer las mediciones del grupo de control.	Sesgo de medición	Los datos provenientes del grupo experimental estarán recibiendo más atención y serán medidos por un experto con mayor capacidad de observación.
Para estudiar la práctica de medidas de salud reproductiva se envía un cuestionario por la vía del correo electrónico a 2000 informantes.	Sesgo de selección	En la actualidad, solamente un porcentaje de la población tiene acceso a Internet, por lo que la muestra es sesgada. Es posible que las personas que no tienen correo electrónico difieran mucho en nivel educativo y tengan menor información y acceso a las medidas de salud reproductiva
En un estudio retrospectivo se pregunta a los participantes sobre los niveles de consumo de un fármaco en diferentes etapas de su vida.	Sesgo de medición	Aquí la medición está afectada por la memoria, la probabilidad de olvido es muy alta.
Con una encuesta se quiere medir la frecuencia de la violencia doméstica y, para ello, se pregunta sobre la cantidad de incidentes y ataques intrafamiliares, además de su periodicidad)	Sesgo de medición	Hay imprecisión sobre la unidad de medida. La violencia no solo incluye incidentes y ataques, esta puede ser crónica y habitual. ¿Qué es violencia doméstica? ¿Incluye la violencia psicológica? ¿La económica? ¿Contra el adulto mayor? ¿El abandono? ¿La no aplicación de tratamientos médicos?

EJERCICIO 1.11

OBSERVACIÓN

Ventajas	Desventajas
Los datos son recogidos por el mismo investigador con lo cual se minimizan los errores de conteo y transcripción. Los informantes que no pueden comunicarse verbalmente (niños, pacientes en coma) siempre pueden aportar datos.	Exige mucha pericia de personal recolector de los datos brutos. Absorbe muchos recursos financieros y tiempo del investigador.

ENTREVISTA

Ventajas	Desventajas
Existe retroalimentación y aclaración para las preguntas confusas o ambiguas. Se puede conocer las causas de la no respuesta (ausencia, error en la dirección, muerte. El recolector de datos puede adecuar la entrevista a la situación psicológica del informante (nerviosismo, violencia).	Normalmente es el método más caro. Exige capacitación de los entrevistadores. Requiere de ganarse la confianza entre entrevistador y entrevistado.

REGISTRO

Ventajas	Desventajas
La instancia recolectora de los datos ya cuenta con la legitimación para efectuarlos. La obligatoriedad de informar está regulada por ley y ya se tienen facilidades administrativas y campañas de concientización. Se reduce el subregistro, las omisiones, duplicaciones e incongruencias.	Los datos generados no siempre coinciden con lo buscado por el investigador, o con las necesidades especiales de un estudio particular. Los datos son muy generales, adecuados para el uso del sistema y de tipo administrativo.

CORREO

Ventajas	Desventajas
Costo muy bajo. Útil para estudios exploratorios y para gremios en los cuales se conocen las direcciones y manejan un mismo registro del lenguaje.	Alto nivel de no respuesta y respuestas incompletas. Exige alto nivel de alfabetismo, servicio postal eficiente o acceso a Internet.

EJERCICIO 1.2

1. Verdadero.

   La estadística es instrumental y tiene mayor sentido cuando interactúa con el resto de las disciplinas científicas.

2. Falso.

   La estadística inferencial hace generalizaciones a partir del estudio de muestras o de distribuciones teóricas de probabilidad)

   Es la estadística descriptiva la que se especializa en la recolección, el resumen y la presentación de los datos, destaca sus características y relaciones para hacer la información más comprensible.

3. Verdadero.

   La inferencia es precisamente generalizar a la totalidad, partiendo de lo que se sabe de la muestra o de experimentos.

4. Verdadero.

   Una hipótesis es una propuesta de explicación provisional o una proposición que, aunque no esté comprobada, es aceptable como punto de partida de investigación.

5. Falso.

   La estadística trabaja con datos numéricos, por lo que su mayor aporte es en el análisis cuantitativo. Por otra parte, la estadística inferencial trabaja con el método inductivo puesto que hace generalizaciones a partir de información muestral.

6. Falso.

   Son los estudios longitudinales los que requieren del transcurso del tiempo para identificar las causas, comportamiento y tendencia de un fenómeno. Los estudios longitudinales pueden ser retrospectivos o prospectivos. Los retrospectivos parten de una situación de hoy y "retroceden" a buscar las causas valiéndose de datos del pasado. Los estudios prospectivos comienzan hoy con datos que se recolectarán en el futuro, conforme aparezcan.

7. Falso.

   Los estudios retrospectivos parten de una situación de hoy y "retroceden" a buscar las causas valiéndose de datos del pasado. Los estudios prospectivos comienzan hoy con datos que se recolectarán en el futuro, conforme aparezcan.

8. Verdadero.

   Recuerde que ecología es un neologismo construido por las raíces griegas *oikos* = eco = casa, entorno, habitat y *logos* = logía = análisis; entonces, ecológico es un estudio que considera las interrelaciones entre los seres y su entorno. Por ello la unidad de estudio de las investigaciones ecológicas son los grupos, las comunidades o las poblaciones.

9. Falso.

   En las investigaciones cuasi-experimentales la variable en estudio NO es manipulada por el estudioso, sino que son "experimentos naturales" en los que la variación se produce exógenamente.

10. Verdadero.

    La población se compone del total de elementos sobre los cuales se realiza un estudio.

11. Falso.

    Los atributos no se pueden medir, pero sí pueden ser contados. Son las variables las que pueden ser medidas.

12. Falso.

    Son las variables continuas las que pueden asumir infinito número de valores en un intervalo.

13. Verdadero.

    Dicotomía viene de las raíces griegas *dikh* = dos y *tomía* = corte, partición, generando "división en dos". De allí que se llaman variables dicotómicas a las que solamente pueden asumir dos valores.

14. Falso.

    Un censo estudia a *toda* la población. Una muestra es un subconjunto que va a ser sometido a la investigación.

15. Verdadero.

    Sesgar es torcer hacia un lado alguna cosa, darle una tendencia, de allí que en estadística es importante que los datos y los estimadores sean insesgados.

16. Falso.

    Un sesgo de medición surge por un problema en el instrumento de medida (báscula, metro, termómetro) o por la tendencia de los informantes a declarar menos (cigarrillos fumados diariamente. o más (escolaridad). Esta tendencia no se corrige al aumentar el tamaño de la muestra.

17. Falso.

    El riesgo de errores de transcripción es mayor, muchas veces se combinan las categorías, se omite información acerca de los métodos de recolección y procesamiento, o la resumen.

18. Verdadero.

    Recuerde que no todos los datos que se recogen o publican son de calidad, por ello siempre es necesario evaluar el origen de los datos.

19. Verdadero.

    No todos los datos son comparables; para lograr la comparabilidad internacional de los informes en salud, se ha trabajado y se sigue trabajando en clasificaciones de enfermedades y causas de muerte uniformes y estandarizadas, se capacita en métodos de recolección y formas de procesamiento de la información.

20. Falso.

    La mayor desventaja de las encuestas por correo es su alto porcentaje de no respuesta y las respuestas incompletas.

EJERCICIO 1.13

1.	Es la ciencia que trata del desarrollo y aplicación de las técnicas de recolección, organización, presentación, análisis e interpretación de datos numéricos obtenidos por observación o experimentación para obtener conclusiones válidas y fundamentar la toma de decisiones.	(4)	Entrevista
2.	Estudios que pretenden establecer conexiones entre variables, medir su correlación o nivel de asociación, comprobar su causalidad, plantear escenarios futuros, medidas preventivas y las consecuencias sobre la salud)	(8)	Hipótesis
3.	Es una investigación sobre las características de toda la población.	(9)	Muestra
4.	Método de recolección de datos en el cual el informante debe contestar preguntas o llenar un cuestionario.	(3)	Censo
5.	Es el riesgo de que la muestra con que se trabaja no sea representativa y surge por la variabilidad del muestreo.	(7)	Fuentes primarias
6.	Características de las unidades de estudio que pueden ser medidas en su longitud, peso, temperatura, precio o duración en el tiempo.	( )	Fuentes secundarias
7.	Datos, documentos e informes generados por la misma institución o sus dependencias.	(6)	Variables
8.	Enunciado que propone una relación entre los datos, una explicación preliminar a un problema, una posible solución o un método de comprobación de las determinaciones entre variables.	(1)	Estadística
9.	Subconjunto de la población que se examina para observar las características de interés.	(5)	Error de muestreo
10.	Conteo de los hechos conforme se van presentando, por obligación de ley, como requisitos para tramitar indemnizaciones o por necesidades justificadas.	( )	Descriptivos
		(2)	Analíticos
		(10)	Registro

CAPÍTULO 2

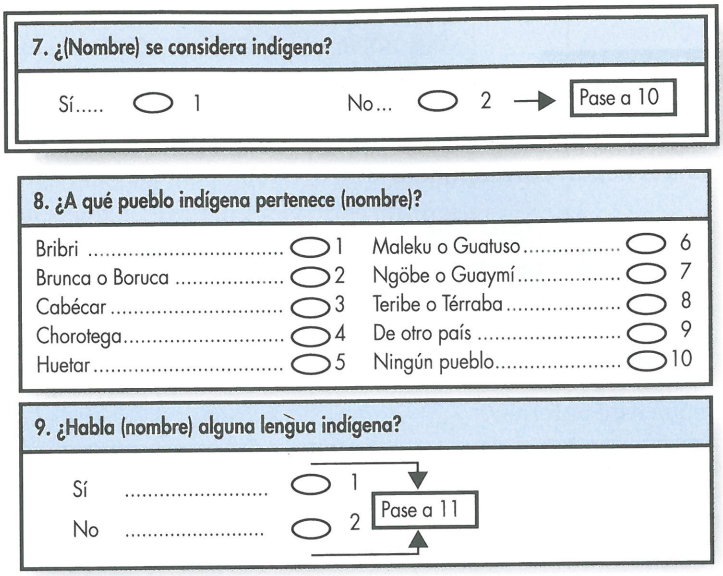
EJERCICIO 2.1

Señale si las siguientes preguntas son de hechos, comportamiento, intención u opinión, marcando con una X según corresponda.

		Hechos	Comp.	Intención	Opinión
a.	¿Comería más frutas y verduras?			X	
b)	¿A qué edad comenzó a fumar?	X			
c)	¿Fuma aunque esté enfermo?		X		
d)	¿Sus preocupaciones le han hecho perder el sueño?	X			
e)	¿Cuál candidato presidencial cree que mejorará los servicios de salud pública?				X
f)	¿Intentaría dejar de apostar?			X	
g)	¿Recientemente ha perdido más de 6 kilogramos en un periodo de seis meses?	X			
h)	¿Se siente mejor atendido en el hospital o en el EBAIS?				X
i)	¿Qué especialidad médica escogerá cuando se gradúe?			X	
j)	¿Por quién votó en las pasadas elecciones?		X		
k)	¿Cuántos miembros conforman su hogar?	X			
l)	¿Cuántos de ellos trabajan?	X			

EJERCICIO 2.2

Las preguntas filtro son la 7 y la 9.



La pregunta 7 elimina la aplicación de las preguntas 8 y 9 si el entrevistado no se considera indígena.

Si el censado se considera indígena, se le aplicarán las preguntas 8 y 9, pero no la pregunta 10, puesto que responder de cualquier forma a la pregunta 9, elimina la necesidad de plantear la pregunta 10, remitiendo a la 11.

EJERCICIO 2.3

Podrían replantearse de la siguiente forma (no son las únicas respuestas posibles).

a) ¿Edad en años cumplidos?

1. ❑ Menos de 20 años
2. ❑ De 20 a 29 años
3. ❑ De 30 a 39 años
4. ❑ De 40 a 49 años
5. ❑ De 50 a 59 años
6. ❑ De 60 a 69 años
7. ❑ De 70 a 79 años
8. ❑ 80 años o más

b) ¿Cuál es su ocupación principal?

1. ❑ Estudiante
2. ❑ Ama de casa
3. ❑ Asalariado
4. ❑ Empresario
5. ❑ Desempleado
6. ❑ Pensionado
7. ❑ Otra

Especifique: _____

c) Si es estudiante, ¿dónde estudia?

1. ❑ Escuela primaria
2. ❑ Colegio diurno
3. ❑ Colegio nocturno
4. ❑ Escuela comercial
5. ❑ INA
6. ❑ Universidad
7. ❑ Otra

Especifique: _____

d) Si desempeña alguna actividad remunerada, ¿cuál es?

1. ❑ Funcionario público
2. ❑ Trabajador por cuenta propia
3. ❑ Microempresario
4. ❑ Empleado de empresa privada
5. ❑ Cooperativista
6. ❑ Otra

Especifique: _____

e) ¿Cuál es su ingreso mensual aproximado?

1. Menos de 100 000 colones
2. De 100 000 a menos de 200 000 colones
3. De 200 000 a menos de 300 000 colones
4. De 300 000 a menos de 400 000 colones
5. De 400 000 a menos de 500 000 colones
6. 500 000 colones o más

EJERCICIO 2.4

Marque una X en la casilla correspondiente si usted está de acuerdo o en desacuerdo con las siguientes medidas:

		Acuerdo	Desacuerdo
a.	Debería venderse cigarrillos solo a mayores de edad.		
b)	Los cigarrillos únicamente deben venderse en paquetes sellados.		
c)	Hay que continuar con la campaña antitabaco.		
d)	El Estado debe contratar empleados libres del tabaquismo.		
e)	Tiene que mantenerse la información masiva sobre las enfermedades causadas por el fumado.		

EJERCICIO 2.5

Un otorrinolaringólogo entrevista a una paciente:

		Sí	No
a)	¿Tiene dificultad para respirar?		
b)	¿Ha padecido de pitos o respiración de gato por la noche?		
c)	¿Sufre de ardor en la garganta al tragar por la mañana?		
d)	¿Usted escupe flemas con frecuencia?		
e)	¿Le han salido ampollas o llagas en la boca?		

EJERCICIO 2.6

		Sí	No
a)	¿Cree que la mayoría de sus alumnos tiene sobrepeso?		
b)	Es decir, ¿piensa que están gorditos?		
c)	¿Les ha afectado la disminución en el presupuesto del comedor escolar?		
d)	¿Ya les aplicaron el recorte presupuestario?		
e)	¿El Director es indiferente ante las limitaciones y carencias?		

EJERCICIO 2.7

Señale si los siguientes pares de categorías son mutuamente excluyentes.

a)	Soltero	Casado	Mutuamente excluyentes
b)	Trabajar	Tener caries	---
c)	Contaminado	No contaminado	Mutuamente excluyentes
d)	Tener tos	Tener diarrea	---
e)	Enfermo	Sano	Mutuamente excluyentes
f)	Enfermo	Hospitalizado	---
g)	Fuma	No fuma	Mutuamente excluyentes
h)	Nacional	Extranjero	Mutuamente excluyentes

EJERCICIO 2.8

Anote si las siguientes clases son colectivamente exhaustivas

a)	De Heredia o Cartago	- - -
b)	Soltero, casado, unión libre, separado, divorciado o viudo	Colectivamente exhaustivas
c)	Contaminado o no contaminado	Colectivamente exhaustivas
d)	Dengue o paludismo	- - -
e)	Estar incapacitado o no	Colectivamente exhaustivas
f)	Ganar, perder o empatar un partido de fútbol	Colectivamente exhaustivas
g)	Obtener 7.0, 8.0 o 9.0 en el final de Estadística	- - -

EJERCICIO 2.9

a) Informe de la distribución de frecuencias para la variable "Estado civil".

ESTADO CIVIL	PALOTES	FRECUENCIA ABSOLUTA
Casada	/////	5
Soltera	///	3
Unión libre	////	4
Divorciada	//	2
Viuda	/	1
TOTAL		**15**

b) Informe de la distribución de frecuencias para la variable "Clase de Parto".

CLASE DE PARTO	PALOTES	FRECUENCIA ABSOLUTA
Cesaria	/////	5
Episiotomía	//	2
Vaginal	///// ///	8
TOTAL		**15**

c) Informe de la distribución de frecuencias para la variable "Sexo del Recién Nacido".

SEXO DEL RECIÉN NACIDO	PALOTES	FRECUENCIA ABSOLUTA
Femenino	///// //	7
Masculino	///// ///	8
TOTAL		**15**

EJERCICIO 2.10

a) Cuadro de doble entrada que relacione el estado civil de la parturienta con la clase de parto.

	CLASE DE PARTO						
ESTADO CIVIL	CESARIA		EPISIOTOMÍA		VAGINAL		TOTAL
	Palotes	F	Palotes	F	Palotes	F	
Casada	/	1			////	4	5
Soltera	//	2			/	1	3
Unión libre	/	1	//	2	/	1	4
Divorciada	/	1			/	1	2
Viuda					/	1	1
TOTAL		5				8	**15**

b) Cuadro de doble entrada que relacione tipo de parto y sexo del recién nacido.

	SEXO DEL RECIÉN NACIDO				
TIPO DE PARTO	FEMENINO		MASCULINO		TOTAL
	Palotes	F	Palotes	F	
Cesaria	/	1	////	4	5
Episiotomía	/	1	/	1	2
Vaginal	/////	5	///	3	8
TOTAL		7		8	**15**

EJERCICIO 2.11

Complete el cuadro siguiente que informa del número de terapias trimestrales a las que acudió un grupo de 80 pacientes diagnosticados con artrosis.

CLASES	FRECUENCIA ABSOLUTA			FRECUENCIA RELATIVA		
N.° de terapias en el trimestre	Simple	Acumulada "X o menos"	Acumulada "X o menos"	Simple	Acumulada "X o menos"	Acumulada "X o menos"
0	8	8	80	0,100	0,100	1,000
1	2	10	72	0,025	0,125	0,900
2	8	18	70	0,100	0,225	0,875
3	16	34	62	0,200	0,425	0,775
4	24	58	46	0,300	0,725	0,575
5	8	66	22	0,100	0,825	0,275
6	8	74	14	0,100	0,925	0,175
7	4	78	6	0,050	0,975	0,075
8 o más	2	80	2	0,025	1,000	0,025
	80			1,000		

Una vez confeccionado el cuadro, conteste:

a) 24 pacientes.

b) 58 pacientes → (8+2+8+16+24).

c) 46 pacientes → (2+4+8+8+24).

d) 20 %

e) 42,5 % → (0,10+0,025+0,10+0,20).

f) 77,5 % → (0,025+0,05+0,10+0,10+0,30+0,20).

EJERCICIO 2.12

a) Distribución de frecuencias absolutas y relativas:

CLASES	Palotes	FRECUENCIA ABSOLUTA	FRECUENCIA RELATIVA
10 a menos de15	//	2	0,04
15 a menos de 20	////	4	0,08
20 a menos de 25	/////	5	0,10
25 a menos de 30	///// ///// //	12	0,24
30 a menos de 35	///// ///// ////	14	0,28
35 a menos de 40	///// ////	9	0,18
40 a menos de 45	///	3	0,06
45 a menos de 50	/	1	0,02
		50	1,00

b) Frecuencias acumuladas "al límite superior" y "al límite inferior":

	Frecuencias acumuladas al límite superior		Frecuencias acumuladas al límite inferior	
CLASES	Absoluta	Relativa	Absoluta	Relativa
10 a menos de15	2	0,04	50	1
15 a menos de 20	6	0,12	48	0,96
20 a menos de 25	11	0,22	44	0,88
25 a menos de 30	23	0,46	39	0,78
30 a menos de 35	37	0,74	27	0,54
35 a menos de 40	46	0,92	13	0,26
40 a menos de 45	49	0,98	4	0,08
45 a menos de 50	50	1,00	1	0,02

c) Conteste las siguientes preguntas:

i. 12 mujeres tenían de 25 a menos de 30 años al momento de dar a luz.

ii. El 74 % de las mujeres tenía menos de 35 años para la fecha del alumbramiento.

iii. 8 % de las mujeres tenía 40 años o más.

iv. 2 mujeres tenían menos de 15 años en la fecha del parto.

v. 12 % de mujeres tenía menos de 20 años.

EJERCICIO 2.13

Encierre en un círculo la opción que se ajusta al enunciado, Falsa (F) o Verdadera (V).

1. Verdadero

   Las investigaciones buscan aclarar aquello que no conocemos o responder a lo que ignoramos.

2. Verdadero

   La identificación del problema es central. Cuando se tiene una molestia pero no se puede identificar el problema se corre el riesgo de disparar para todos lados, malgastando los recursos.

3. Falso

   Un investigador serio revisa todas las publicaciones posibles sobre el tema investigado, en algunos casos se apoyará en lo que ya se conoce del tema, en otros refutará los hallazgos previos.

4. Verdadero

   Los objetivos determinan hasta dónde llegan los alcances de una investigación.

5. Falso

   Dato no es sinónimo de información, pero puede llegar a serlo, toda vez que: esté claramente definido o se pueda saber qué es, pueda ser observado en sus manifestaciones físicas, sea medible cuantitativamente, se pueda categorizar cualitativamente y tenga atributos o sea variable

6. Verdadero

   La observación, los experimentos y los registros son instrumentos directos de captura de datos.

7. Falso

   Las entrevistas, las encuestas y los grupos focales se consideran instrumentos indirectos de recolección de información, puesto que son los informantes quienes vieron, conocen o son testigos de algo.

8. Verdadero

   Los instrumentos de recolección de datos deben ser objetivos, confiables y válidos.

9. Falso

   Una prueba piloto siempre ayudará a corregir las imprecisiones del instrumento, además sirve para mejorar el nivel de comunicación por medio de la adecuación del lenguaje al nivel del informante. La prueba piloto también evidencia las necesidades de capacitación del personal en los diferentes aspectos que enfrentará en el trabajo de campo y en su relación con los encuestados.

10. Verdadero

    La capacitación de los encuestadores es clave para reducir errores y sesgos en el trabajo de campo.

11. Falso

    Las tareas administrativas ineficientes conducirán al fracaso de la investigación, así que no pueden dejarse sin la adecuada supervisión.

12. Verdadero

    Un cuestionario autoadministrado lo llena directamente el informante, en tanto que la entrevista requiere de un intermediario que hace las preguntas y llena el formulario.

13. Verdadero

    Los costos aumentan, pero la calidad de los datos mejora.

14. Falso

    Ni los instrumentos estadísticos modernos ni los equipos informáticos sustituyen el trabajo de los investigadores y especialistas. Todos los datos tienen que ser interpretados bajo marcos teóricos y metodológicos apegados a paradigmas y buscando superarlos. Ni las máquinas ni los números interpretan nada, siempre hace falta la mente humana.

15. Verdadero

    Además de las preguntas, un buen cuestionario debe contener indicaciones breves que orienten al entrevistador y faciliten la comunicación con el informante.

16. Verdadero

    Las preguntas de un cuestionario pueden ser sobre hechos concretos, comportamientos, intenciones y opiniones de los entrevistados.

17. Verdadero

    El respeto del tiempo y la dosificación del esfuerzo mental de los informantes es parte fundamental de un buen cuestionario.

18. Verdadero

    Debe evitarse que las respuestas sean conducidas o contaminadas conforme se va desarrollando un cuestionario.

19. Falso

    Las preguntas con negación confunden al informante; además tienen mayor riesgo de que lo anotado sea contrario a la opinión que se quiso manifestar. Por ello, la recomendación general es usar frases afirmativas.

20. Falso

    La codificación es más sencilla en las preguntas cerradas o en las semiabiertas porque previamente se conocen las categorías. Los datos codificados facilitan el conteo de ocurrencias y el uso de lectores ópticos asociados a programas informáticos.

EJERCICIO 2.15

1.	Instrumento de presentación de datos, donde estos se ordenan en filas y columnas para ser comprendidos sin mucha explicación textual o gráfica.	(2)	Estado de la cuestión
2.	Es el cuerpo de publicaciones que otros investigadores han escrito sobre mi objeto de estudio o tema-problema.	(5)	Gráficos estadísticos
3.	Palabras sustitutivas de otra considerada ofensiva, de mal gusto o tabú.	(7)	Preguntas bidimensionales
4.	Numero de ocurrencias de un determinado evento o conteo de las observaciones pertenecientes a una categoría.	(1)	Cuadro estadístico
5.	Instrumentos estadísticos que comunican la información utilizando distancias, áreas y pendientes o volúmenes.	(8)	Validez
6.	Son descripciones de lo que se espera del trabajo de investigación. Se plantean en forma afirmativa y con verbos en infinitivo.	( )	Confianza
7.	Solicitan una sola respuesta en un ítem que, en realidad, contiene dos preguntas.	(9)	Pregunta de investigación
8.	Grado de confianza en que un instrumento de recolección de datos mide lo que pretende medir.	(4)	Frecuencia
9.	Cuestionamiento general y abstracto del tema por investigar.	(10)	Gráfico circular
10.	Representación estadística que informa de la participación porcentual de cada una de las categorías.	( )	Semitabular
		(3)	Eufemismos
		(6)	Objetivos

CAPÍTULO 3

EJERCICIO 3.1

a) Cálculos:

i Razón de habitantes por cada médico = $\frac{\text{Habitantes del cantón}}{\text{Total de médicos}}$

$$\frac{25\ 857}{8} = 3\ 232$$

A cada médico de este cantón fronterizo le corresponde atender potencialmente a 3232 habitantes.

ii Médicos por cada 1 000 habitantes = $\left(\frac{\text{Total de médicos}}{\text{Habitantes del cantón}}\right) \times 1000$

$$\frac{8}{25\ 857} \times 1\ 000 = 0,3$$

Este cantón fronterizo dispone 0,3 médicos por cada 1 000 habitantes. Por resultar una razón muy baja, es mejor amplificar por 10,000 porque así informaría de 3 médicos por cada 10,000 habitantes

iii Habitantes por trabajador de la salud = $\frac{\text{Habitantes del cantón}}{\text{Personal médico destacado}}$

$$\frac{25\ 857}{40} = 646$$

A cada trabajador de la salud le corresponde la atención de 646 habitantes de este cantón fronterizo.

b) Razón de enfermeros a médicos = $\frac{\text{Personal de enfermería}}{\text{Número de médicos}}$

$$\frac{124}{31} = 4$$

Se dispone de 4 enfermeros por cada médico.

c) Razón de operados por cada 10 no operados = $\frac{\text{Total de operados}}{\text{Total de no operados}} \times 10$

$$\frac{12}{20} \times 10 = 6{,}0$$

Se tienen 6 pacientes operados por cada 10 que no requieren de intervención quirúrgica.

d) Egresos anuales por cama = $\frac{\text{Total de egresos en el año}}{\text{Número de camas}}$

$\frac{864}{24} = 36$ egresos anuales por cama de psiquiatría.

e) Consultas anuales por médico = $\frac{\text{Consultas externas}}{\text{Médicos}}$

$\frac{156\ 240}{31} = 5\ 040$ Consultas anuales por médico

Luego se dividen las consultas anuales entre los 252 días laborados por cada médico.

$\frac{5\ 040}{252} = 20$ Consultas por día-médico.

EJERCICIO 3.2

a) Proporción de cirugías suspendidas $= \dfrac{\text{Cirugías suspendidas}}{\text{Cirugías programadas}} \times 100$

$$\frac{278}{3\,475} \times 100 = 8\%$$

El 8 % de cirugías suspendidas por causas no médicas activa procedimientos para identificar las barreras a los procedimientos quirúrgicos y que conducen a la suspensión de las cirugías.

b) Reintervenciones quirúrgicas $= \dfrac{\text{Número de reintervenciones quirúrgicas}}{\text{Total de cirugías}} \times 100$

$$\frac{264}{5\,280} \times 100 = 5\%$$

El 5 % de reintervenciones quirúrgicas señala deficiencias en los procesos, prescripción e implementación de tratamientos médico-quirúrgicos.

c) Construya una tabla que informe las proporciones que estos tipos de cáncer de mayor frecuencia representan del total de diagnosticados con cáncer.

Proporción de los principales tumores malignos
(Región Norte de Perú, 2011)

Tipo de Tumor	Pacientes diagnosticados	Frecuencia Relativa	Proporción
Cuello uterino	190	190/1209	15,7%
Mama	183	183/1209	15,1%
Piel	135	135/1209	11,2%
Estómago	105	105/1209	8,7%
Próstata	82	82/1209	6,8%
TOTAL	1 209		

Fuente: Elaborado con base en datos de Hernández Morón (2012)

d) Proporción de pacientes referidos a endocrinología.

$$\frac{\text{Número de referencias a endocrinología}}{\text{Total de referencias}} \times 100$$

$$\frac{463}{2\,568} \times 100 = 18\%$$

EJERCICIO 3.3

a) Calcule:

i. La sensibilidad del tacto rectal como prueba de tamizaje.

$$\frac{VP}{VP + FN} \times 100$$

$$\frac{460}{660} \times 100 = 69,7\%$$

El tacto rectal es capaz de identificar al 69,7 % de los pacientes enfermos de cáncer de próstata.

ii. La especificidad de la prueba.

$$\frac{VN}{FP + VN} \times 100$$

$$\frac{660}{740} \times 100 = 89,2\%$$

El tacto rectal es capaz de declarar sanos al 89,2 % de los pacientes que realmente no padece cáncer de próstata.

iii. El valor predictivo de una prueba positiva (que detecta agrandamiento prostático).

$$\frac{VP}{VP + FP} \times 100$$

$$\frac{460}{540} \times 100 = 85,2\%$$

Únicamente el 85,2 % de los tactos rectales positivos corresponden a pacientes con cáncer de próstata.

iv. El valor predictivo de una prueba negativa (que no revela agrandamiento de la próstata).

$$\frac{VN}{FN + VN} \times 100$$

$$\frac{660}{860} \times 100 = 76,7\%$$

El 76,7 % de los pacientes cuyas tactos rectales resultaron negativos no tiene realmente cáncer de próstata.

b) Calcule:

i. La sensibilidad de la baciloscopía como prueba de tamizaje para la detección temprana de la tuberculosis.

$$\frac{VP}{VP + FN} \times 100$$

$$\frac{19}{60} \times 100 = 31{,}67\%$$

La baciloscopía es capaz de identificar al 31,67 % de los pacientes que realmente están enfermos de tuberculosis.

ii. La especificidad de la prueba.

$$\frac{VN}{FP + VN} \times 100$$

$$\frac{639}{640} \times 100 = 99{,}8\%$$

La baciloscopía es capaz de declarar sin tuberculosis al 99,8 % de los pacientes que realmente no la tiene.

iii. El valor predictivo de una prueba positiva.

$$\frac{VP}{VP + FP} \times 100$$

$$\frac{19}{20} \times 100 = 95\%$$

Únicamente el 95 % de las baciloscopías positivas corresponden a pacientes tuberculosos.

iv. El valor predictivo de una prueba negativa.

$$\frac{VN}{FN + VN} * 100$$

$$\frac{639}{680} \times 100 = 93,97\%$$

El 93,97 % de los pacientes cuyas baciloscopías resultaron negativas no tiene tuberculosis.

c) Calcule:

i. La sensibilidad de la ultrasonografía como prueba de tamizaje.

$$\frac{VP}{VP + FN} \times 100$$

$$\frac{8}{48} \times 100 = 16,67\%$$

La ultrasonografía es capaz de identificar al 16,67 % de los pacientes que realmente padecen retraso del crecimiento fetal.

ii. La especificidad de la prueba.

$$\frac{VN}{FP + VN} \times 100$$

$$\frac{240}{242} \times 100 = 99,17\%$$

La ultrasonografía es capaz de declarar sin retaso del crecimiento fetal al 99,17 % de los bebés que realmente tiene un peso adecuado.

iii. El valor predictivo de una prueba positiva.

$$\frac{VP}{VP + FP} \times 100$$

$$\frac{8}{10} \times 100 = 80\,\%$$

Únicamente el 80 % de las ultrasonografías positivas (que detectaron retraso del crecimiento fetal) corresponden a bebés con bajo peso al nacer.

iv. El valor predictivo de una prueba negativa.

$$\frac{VN}{FN + VN} \times 100$$

$$\frac{240}{280} \times 100 = 85,7\%$$

El 85,97 % de los pacientes cuyas ultrasonografía resultaron negativas (sin retraso en el crecimiento fetal) no tendrá bajo peso al nacer.

EJERCICIO 3.4

a) Año 2000.

i. $\text{Tasa de mortalidad masculina}_{2000} = \left(\frac{8\,559}{1\,902\,614}\right) \times 100\,000 = 450$

ii. $\text{Tasa de mortalidad femenina}_{2000} = \left(\frac{6\,385}{1\,907\,565}\right) \times 100\,000 = 335$

b) Tasas brutas de natalidad, mortalidad y crecimiento natural.

Año	Población	Nacimientos	Defunciones	Crecimiento natural	Tasa Bruta de Natalidad	Tasa Bruta de Mortalidad
2005	4 263 479	71 650	16 034	55 616	16,81	3,76
2006	4 401 849	71 292	16 766	54 526	16,20	3,81
2007	4 389 139	73 634	17 067	56 567	16,78	3,89
2008	4 451 205	70 910	17 845	53 065	15,93	4,01
2009	4 509 290	68 407	18 302	50 105	15,17	4,06
2010	4 563 539	66 256	18 863	47 393	14,52	4,13
2011	4 615 518	67 034	19 469	47 565	14,52	4,22
2012	4 667 096	67 700	20 111	47 589	14,51	4,31

c) $\left(\frac{\text{Número de defunciones infantiles}}{\text{Nacidos vivos del periodo}}\right) \times 1000$

$$\text{Tasa de mortalidad infantil}_{2006} = \left(\frac{692}{71\ 292}\right) \times 1000 = 9{,}7$$

d) Año 2003, tasa de mortalidad por causa de cada 10,000 habitantes:

i. $\left(\frac{\text{Defunciones por enfermedades del Sistema Circulatorio}}{\text{Población total}}\right) \times 10\ 000$

$$\text{Tasa de mortalidad enf. Sistema Circulatorio}_{2003} = \left(\frac{4344}{4\ 136\ 250}\right) \times 10\ 000 = 10{,}5$$

10,5 habitantes por cada 10 000 fallecieron por enfermedades del sistema circulatorio en el año 2003 en Costa Rica.

ii. $\left(\frac{\text{Número de defunciones por tumores}}{\text{Población total}}\right) \times 10\ 000$

$$\text{Tasa de mortalidad por tumores}_{2003} = \left(\frac{3398}{4\ 136\ 250}\right) \times 10\ 000 = 8{,}2$$

8,2 habitantes por cada 10 000 fallecieron por tumores en el año 2003 en Costa Rica.

iii. $\left(\frac{\text{Número de defunciones por causas externas}}{\text{Población total}}\right) \times 10\ 000$

$$\text{Tasa de mortalidad por causas externas}_{2003} = \left(\frac{1909}{4\ 136\ 250}\right) \times 10\ 000 = 4{,}6$$

4,6 habitantes por cada 10 000 fallecieron por causas externas (accidentes, homicidios y suicidios) en el año 2003 en Costa Rica.

EJERCICIO 3.5

a) Tasa de incidencia de cáncer de próstata.

$$\left(\frac{\text{Casos nuevos de cáncer de próstata en el periodo}}{\text{Población en riesgo}}\right)\times 100\,000$$

$$\text{Tasa de incidencia}_{\text{cáncer de próstata}}=\left(\frac{647}{1\,434\,589}\right)\times 100\,000=45{,}1$$

La incidencia de cáncer de próstata es de 45,1 casos nuevos por cada 100 000 varones mayores de edad)

b) Tasa de prevalencia de la tuberculosis.

$$\left(\frac{\text{Personas que padecen tuberculosis}}{\text{Población}}\right)*100\,000$$

$$\text{Tasa de prevalencia}_{\text{tuberculosis}}=\left(\frac{528}{4\,664\,311}\right)*100\,000=11{,}32$$

La prevalencia de tuberculosis en esta zona geográfica es de 11,32 personas por cada 100,000 habitantes.

c) Tasa de casos de dengue.

$$\left(\frac{\text{Cantidad de casos declarados de dengue en el periodo}}{\text{Población total}}\right)\times 10\,000$$

$$\text{Tasa de casos}_{\text{Dengue}}=\left(\frac{7821}{4\,664\,311}\right)\times 10\,000=16{,}76$$

En el periodo, la tasa de casos de dengue fue de 16,76 por cada 10 000 habitantes.

d) Tasa de letalidad del VIH/sida:

$$\left(\frac{\text{Número de personas que muere de VIH/sida}}{\text{Número de personas que contrae VIH/sida en el periodo}}\right)\times 100$$

$$\text{Tasa de letalidad}_{\text{VIH/sida}}=\left(\frac{3251}{13\,974}\right)\times 100=23{,}26$$

El 23,26 % de las personas infectadas por el VIH/sida murió durante el periodo.

EJERCICIO 3.6

a) Calcule:

i $$\frac{\text{Nacidos vivos con peso al nacer menor de 2500 gramos}}{\text{Total de nacidos vivos}} \times 100$$

$$\frac{160}{2134} \times 100 = 7{,}5\,\%$$

La proporción de nacidos vivos con bajo peso al nacer es de 7,5 %

ii $$\frac{\text{Nacidos vivos con peso al nacer menor de 1500 gramos}}{\text{Total de nacidos vivos}} \times 100$$

$$\frac{47}{2134} \times 100 = 2{,}2\,\%$$

La proporción de nacidos vivos con muy bajo peso es del 2,2 %.

b) Indicadores proporcionales de mortalidad por edad:

Total de defunciones generales
Por grupo de edad y sexo
(Costa Rica, 2011)

Edad	Hombres	Mujeres	Total	Proporción
Menos de 1	374	292	666	3,54
1 a 4	54	46	100	0,53
5 a 9	32	31	63	0,34
10 a 14	65	42	107	0,57
15 a 19	164	72	236	1,26
20 a 24	225	86	311	1,65
25 a 29	293	91	384	2,04
30 a 34	273	119	392	2,08
35 a 39	266	144	410	2,18
40 a 44	366	175	541	2,88
45 a 49	457	229	686	3,65
50 a 54	630	351	981	5,22
55 a 59	669	421	1 090	5,80
60 a 64	761	512	1 273	6,77
65 a 69	859	583	1 442	7,67
70 a 74	995	710	1 705	9,07
75 a 79	1 148	860	2 008	10,68
80 a 84	1 239	1 126	2 365	12,58
85 ó más	1 790	2 244	4 034	21,46
Ignorado	6	1	7	0,04
TOTAL	**10 666**	**8 135**	**18 801**	100,00

Fuente: INEC (2012) <http://www.inec)go.cr/Web/Home/pagPrincipal.aspx>.

c) Mortalidad proporcional de 70 años o más.

$$\left(\frac{1\,705+2\,008+2\,365+4\,034}{18\,801}\right)\times 100 = 53{,}78\,\%$$

En Costa Rica, durante el año 2011, el 53,78 % de las defunciones fue de personas de 70 años o más.

Este es un buen indicador del desarrollo de la sociedad y de los resultados del sistema nacional de salud, puesto que más de la mitad de las defunciones es de personas que completaron su ciclo productivo.

d) Indicadores proporcionales de mortalidad infantil por causa de muerte.

Mortalidad infantil según causa de muerte
(Costa Rica, 2010)

Causa de muerte	Muertes	Porcentaje
Afecciones en el periodo perinatal	330	49,18
Malformaciones congénitas	234	34,87
Enfermedades del sistema respiratorio	38	5,66
Enfermedades del sistema circulatorio	14	2,09
Enfermedades infecciosas o parasitarias	12	1,79
Enfermedades del sistema nervioso	10	1,49
Otras causas	33	4,92
TOTAL	**671**	**100,00**

Fuente: INEC, Ministerio de Salud, Dirección de Vigilancia de la Salud)

e) Indicador proporcional de morbilidad$_{\text{enfermedades osteomusculares}}$

$$=\left(\frac{8\,597}{303\,789}\right)\times 100 = 2{,}83\,\%$$

La proporción de egresos hospitalarios atribuida a enfermedades osteomusculares es 2,83 %.

EJERCICIO 3.7

a) Exposición pasiva al humo del tabaco en mujeres embarazadas y riesgo de que el bebé tenga bajo peso al nacer.

i. Entre las mujeres expuestas al humo del tabaco, la oportunidad de que su bebé tenga bajo peso al nacer es:

$$\frac{\text{Bajo Peso}}{\text{Peso Adecuado}} = \frac{10}{416} = 0{,}024$$

Entre las mujeres no expuestas, la oportunidad de que su bebé tenga bajo peso al nacer es:

$$\frac{\text{Bajo Peso}}{\text{Peso Adecuado}} = \frac{7}{890} = 0{,}0078$$

$$\text{OR} = \frac{\text{Oportunidad de bajo peso estando expuestas}}{\text{Oportunidad de bajo peso sin exposición}} = \frac{0{,}0235}{0{,}0078} = 3$$

Las mujeres expuestas al humo del tabaco tienen 3 veces más probabilidades de tener bebés con bajo peso que las que no están expuestas.

ii. La tasa de incidencia de bajo peso entre las madres expuestas es:

$$\frac{\text{Bajo Peso}}{\text{Población expuesta}} \times 100 = \frac{10}{426} \times 100 = 2{,}35\,\%$$

La tasa de incidencia de bajo peso en las madres libres de la exposición al humo del tabaco es:

$$\frac{\text{Bajo Peso}}{\text{Población no expuesta}} \times 100 = \frac{7}{897} \times 100 = 0{,}78\,\%$$

Por lo que el riesgo relativo es:

$$\text{RR} = \frac{2{,}35}{0{,}78} = 3$$

El riesgo relativo de tener un bebé con bajo peso al nacer entre las mujeres expuestas al humo del tabaco es 3 veces el riesgo de mujeres no expuestas.

iii. El riesgo atribuible (RA) es la diferencia entre la tasa de incidencia en las mujeres expuestas al humo del tabaco menos la tasa de incidencia en las no expuestas.

$$\text{RA} = 2{,}35\,\% - 0{,}78\,\% = 1{,}57\,\%$$

1,57 % es la cantidad en que se reduce el riesgo de bajo peso al nacer por alejarse (las madres) del humo del tabaco.

iv. La fracción atribuible (FA) a la exposición al humo del tabaco es:

$$FA = \frac{RR-1}{RR}$$

$$FA = \frac{3-1}{3} = 0,667$$

66,7 % del bajo peso al nacer se atribuye a la exposición de la embarazada al humo del tabaco.

b) Asociaciones entre sufrir un infarto y el tratamiento con Enalapril maleato de 20 mg en pacientes hipertensos.

i. Entre los pacientes hipertensos sin tratamiento, la oportunidad de sufrir un infarto es:

$$\frac{\text{Infarto}}{\text{No infarto}} = \frac{42}{498} = 0,084$$

Entre los pacientes hipertensos con tratamiento, la oportunidad de sufrir un infarto es:

$$\frac{\text{Infarto}}{\text{No infarto}} = \frac{10}{633} = 0,0158$$

$$OR = \frac{\text{Oportunidad de infarto sin tratamiento}}{\text{Oportunidad de infarto con tratamiento}} = \frac{0,084}{0,0158} = 5,3$$

Los pacientes hipertensos sin tratamiento tienen 5,3 veces más probabilidades de sufrir un infarto que los que siguen el tratamiento.

ii. La tasa de incidencia de infarto entre los hipertensos sin tratamiento es:

$$\frac{\text{Infartados}}{\text{Total sin tratamiento}} \times 100 = \frac{42}{540} \times 100 = 7,78\%$$

La tasa de incidencia de infarto entre los pacientes con tratamiento es:

$$\frac{\text{Infartados}}{\text{Total con tratamiento}} \times 100 = \frac{10}{643} \times 100 = 1,55\%$$

Por lo que el riesgo relativo es:

$$RR = \frac{7,78}{1,55} = 5$$

El riesgo relativo de sufrir un infarto entre los pacientes sin tratamiento es 5 veces el riesgo de los pacientes medicados.

iii. El riesgo atribuible (RA) es la diferencia entre la tasa de incidencia de infarto en los pacientes sin tratamiento menos la tasa de incidencia en los que siguen el tratamiento.

$$RA = 7,78\,\% - 1,55\% = 6,23\,\%$$

6,23 % es la cantidad en que se reduce el riesgo de infarto por el hecho de seguir el tratamiento.

iv. La fracción atribuible (FA) a la falta de tratamiento es:

$$FA = \frac{RR - 1}{RR}$$

$$FA = \frac{5 - 1}{5} = 0,80$$

El 80 % de los infartos en pacientes hipertensos se atribuye a la falta de tratamiento.

c) Resultados de las pruebas de eficacia del antidepresivo.

i. Entre los pacientes tratados con el antidepresivo, la oportunidad de mejorar es:

$$\frac{\text{Mejora}}{\text{No mejora}} = \frac{66}{22} = 3$$

Entre los pacientes tratados con placebo, la oportunidad de mejorar es:

$$\frac{\text{Mejora}}{\text{No mejora}} = \frac{44}{88} = 0,5$$

$$\text{OR} = \frac{\text{Oportunidad de mejorar con antidepresivo}}{\text{Oportunidad de mejorar con placebo}} = \frac{3}{0,5} = 6$$

Los pacientes a los que se suministra el antidepresivo tienen 6 veces más probabilidades de mejorar que los que ingieren un placebo.

ii. La tasa de mejoría entre los que se tratan con el antidepresivo es:

$$\frac{\text{Mejora}}{\text{Tratados con antidepresivo}} \times 100 = \frac{66}{88} \times 100 = 75\%$$

La tasa de mejoría entre los pacientes tratados con placebo es:

$$\frac{\text{Mejora}}{\text{Tratados con placebo}} \times 100 = \frac{44}{132} \times 100 = 33,3\%$$

Por lo que el riesgo relativo es:

$$\text{RR} = \frac{75}{33,3} = 2,25$$

El riesgo relativo de mejorar entre los pacientes medicados con antidepresivo es 2,25 veces que el de los pacientes tratados con placebo.

iii. El riesgo atribuible (RA) al antidepresivo es la diferencia entre la tasa de mejoría en los pacientes tratados con el antidepresivo menos la tasa de incidencia en los tratados con placebo.

$$\text{RA} = 75\,\% - 33,3\% = 41,7\,\%$$

41,7 % es la cantidad en que los pacientes mejoran por el hecho de medicarse con el antidepresivo.

iv. La fracción atribuible (FA) de mejoría por tratamiento con el antidepresivo es:

$$\text{FA} = \frac{\text{RR} - 1}{\text{RR}}$$

$$\text{FA} = \frac{2,25 - 1}{2,25} = 0,555$$

El 55,5 % de pacientes que mejoran se atribuye al tratamiento con el antidepresivo.

EJERCICIO 3.8

a) Compare las tasas ajustadas de mortalidad tanto de A como de B, estandarizando ambas tasas con la suma de las dos poblaciones:

Edad	Tasa de A por 1000	Tasa de B por 1000	Suma de poblaciones	Muertes esperadas en A	Muertes esperadas en B
0 a 9	150,00	500,00	3 600	540	1 800
10 a 19	100,00	250,00	3 700	370	925
20 a 29	50,00	50,00	5 000	250	250
30 a 39	25,00	25,00	5 280	132	132
40 a 49	40,00	40,00	2 200	88	88
50 a 59	230,00	25,00	1 600	368	40
60 y más	250,00	138,10	620	155	86
TOTAL	100,00	83,33	22 000	1 903	3 321

$$\text{Tasa de mortalidad ajustada}_{\text{de A}} = \left(\frac{1\,903}{22\,000}\right) \times 1\,000 = 86{,}5$$

$$\text{Tasa de mortalidad ajustada}_{\text{de B}} = \left(\frac{3\,321}{22\,000}\right) \times 1\,000 = 150{,}94$$

Cuadro resumen con las tasas de mortalidad, antes y después del ajuste:

	ANTES DEL AJUSTE	DESPUÉS DEL AJUSTE
POBLACIÓN A	100,00	86,50
POBLACIÓN B	83,33	150,94
DIFERENCIA	16,67	-64,44
COCIENTE	1,2	0,57

Al estandarizar usando la suma de las dos poblaciones, la tasa ajustada de mortalidad de A es 86,5 por mil y la tasa ajustada de mortalidad de B es 150,94 por mil.

Puede decirse que la mortalidad en A es 64,44 por mil menor que la de B, o también, que la mortalidad en A es el 57 % de la mortalidad en B $\left(\frac{86{,}5}{150{,}94} \times 100 = 57\%\right)$.

b) Dos comunidades (X y Y) presentan las siguientes distribuciones de grupos laborales entre sus respectivas poblaciones:

i. Composición porcentual de cada grupo laboral.

Grupos laborales	Población X	Composición porcentual de X	Población Y	Composición porcentual de Y
Empresarios	1 000	5	3 000	10
Profesionales	1 000	5	6 000	20
Técnicos	2 000	10	6 000	20
Obreros	4 000	20	12 000	40
Campesinos	12 000	60	3 000	10
TOTAL	20 000	100	30 000	100

ii. Tasas de prevalencia de alcoholismo.

Grupos laborales	Población X	Alcohólicos	Tasa de prevalencia alcoholismo	Población Y	Alcohólicos	Tasa de prevalencia alcoholismo
Empresarios	1 000	10	10,00	3 000	60	20,00
Profesionales	1 000	40	40,00	6 000	60	10,00
Técnicos	2 000	30	15,00	6 000	120	20,00
Obreros	4 000	200	50,00	12 000	300	25,00
Campesinos	12 000	120	10,00	3 000	60	20,00
TOTAL	20 000	400	20,00	30 000	600	20,00

iii. Es necesario ajustar las tasas de prevalencia de alcoholismo porque la estructura laboral de las dos poblaciones es muy diferente. La población trabajadora de X tiene un 60 % de campesinos y solamente el 10 % es empresario o profesional; en tanto que en Y el 40 % es obrero, el 50 % es empresario, profesional o técnico y apenas un 10 % es campesino.

iv. Número esperado de alcohólicos en X y en Y.

Clase social	Tasas en X	Tasas en Y	Población TIPO	Alcohólicos esperados en X	Alcohólicos esperados en Y
Empresarios	10,00	20,00	500	5	10
Profesionales	40,00	10,00	1 000	40	10
Técnicos	15,00	20,00	1 500	22,5	30
Obreros	50,00	25,00	3 000	150	75
Campesinos	10,00	20,00	4 000	40	80
TOTAL	20,00	20,00	10 000	257,5	205

v. Calcule las tasas ajustadas de prevalencia de alcoholismo en ambas poblaciones.

$$\text{Tasa de prevalencia ajustada}_{\text{de X}} = \left(\frac{257{,}5}{10\,000}\right) \times 1000 = 2{,}575$$

$$\text{Tasa de prevalencia ajustada}_{\text{de Y}} = \left(\frac{205}{10\,000}\right) \times 1000 = 2{,}05$$

vi. Complete el cuadro siguiente y redacte las conclusiones de comparar la prevalencia de alcoholismo en estas dos poblaciones.

	ANTES DEL AJUSTE	DESPUÉS DEL AJUSTE
POBLACIÓN X	20,00	2,575
POBLACIÓN Y	20,00	2,050
DIFERENCIA	0,00	0,525
COCIENTE	1	1,256

Antes de la estandarización, ambas tasas brutas de prevalencia informaban 20 alcohólicos por cada 1000 habitantes, insinuando una igualdad de tasas brutas.

Tras la estandarización, la tasa ajustada de prevalencia de alcoholismo es 0,525 por mil mayor en la Población X. Utilizando el cociente, la tasa de prevalencia del alcoholismo en X es 25,6 % mayor que la de Y.

EJERCICIO 3.9

a) Encuentre:

i. $\text{Tasa bruta de mortalidad en X} = \frac{301}{25\,000} \times 1000 = 12{,}04 \text{ por mil}$

$\text{Tasa bruta de mortalidad en Y} = \frac{156}{20\,000} \times 1000 = 7{,}80 \text{ por mil}$

ii. Defunciones esperadas utilizando las tasas específicas de la Población C)

DEFUNCIONES ESPERADAS		
Edad	En X	En Y
0 a 14	12,00	21,00
15 a 24	7,00	10,00
25 a 34	10,00	6,00
35 a 44	25,00	10,00
45 a 54	30,00	15,00
55 a 64	42,50	17,00
65 o más	50,00	12,50
	176,50	91,50

iii. Razón de estandarización de $X = \frac{D\ O}{D\ E} = \frac{301}{176,5} = 1,705$

Razón de estandarización de $Y = \frac{D\ O}{D\ E} = \frac{156}{91,5} = 1,705$

iv. Estandarización:

$$\text{Tasa Ajustada de X} = (\text{Tasa Bruta de C}) \times (\text{Razón de Estandarización de X})$$

$$\text{Tasa Ajustada de X} = 5,7 \times 1,705 = 9,72$$

$$\text{Tasa Ajustada de Y} = (\text{Tasa Bruta de C}) \times (\text{Razón de Estandarización de Y})$$

$$\text{Tasa Ajustada de Y} = 5,7 \times 1,705 = 9,72$$

	Tasa bruta observada	Tasa bruta ajustada
Población X	12,04	9,72
Población Y	7,80	9,72

Las tasas de mortalidad ajustadas son iguales en ambas poblaciones. Esto implica que si la estructura por edades de las poblaciones X y Y fuera igual, tendrían el mismo comportamiento de mortalidad) La diferencia inicial se debe a que la Población X es más vieja que la de Y.

b) Encuentre:

i. Las defunciones esperadas de hombres y de mujeres.

Edad	Tasas de mortalidad por 100 000	Hombres	Mujeres	Población	Defunciones esperadas de Hombres	Defunciones esperadas de mujeres
0 a 9	129,1	402 730	385 058	787 788	519,9	497,1
10 a 19	43,1	418 028	403 054	821 082	180,2	173,8
20 a 29	94,0	318 086	320 394	638 480	298,9	301,1
30 a 39	121,1	287 515	298 013	585 528	348,1	360,9
40 a 49	249,9	209 176	215 715	424 891	522,8	539,2
50 a 59	534,0	123 246	127 690	250 936	658,1	681,9
60 a 69	1 357,4	77 644	82 148	159 792	1 053,9	1 115,1
70 y más	5 429,8	66 189	75 493	141 682	3 593,9	4 099,1
Total	392,2	1 902 614	1 907 565	3 810 179	7176,0	7768,0

ii. Razón de estandarización Hombres $= \frac{D\,O}{D\,E} = \frac{343}{7176} = 0,0478$

Razón de estandarización Mujeres $= \frac{D\,O}{D\,E} = \frac{367}{7\,768} = 0,0472$

	Defunciones observadas (DO)	Defunciones esperadas (DE)	$\frac{D\,O}{D\,E}$
Hombres	343	7 176	0,0478
Mujeres	367	7 768	0,0472

iii. Tasas ajustadas.

Tasa Ajustada Hombres = (Tasa Población Tipo) × (Razón de Estandarización Hombres)

Tasa Ajustada Hombres $= 392,2 \times 0,0478 = 18,75$

Tasa Ajustada Mujeres = (Tasa Población Tipo) × (Razón de Estandarización Mujeres)

Tasa Ajustada Mujeres $= 392,2 \times 0,0472 = 18,53$

	Tasa bruta observada	Tasa bruta ajustada
Hombres	14,66	18,75
Mujeres	16,13	18,53

La tasa ajustada de mortalidad por enfermedad hipertensiva sería igual, si la estructura por edades en ambos sexos fuera como la de la población tipo.

Si bien la tasa bruta observada indicaba tasas más altas para las mujeres, esto se debe a que ellas alcanzan edades mayores que los hombres y la incidencia de la enfermedad hipertensiva crece cuando se avanza en la edad)

EJERCICIO 3.10

Encierre en un círculo la opción que se ajusta al enunciado, Falsa (F) o Verdadera (V).

1. Verdadero

   Además de construirlos a partir del registro de hechos vitales, su seguimiento permite monitorear eventuales avances, estancamientos o retrocesos en la atención de salud, así como servir de guía para establecer correctivos.

2. Verdadero

   Los números relativos ayudan a resumir grandes volúmenes de datos, a poner en evidencia las relaciones entre dos o más números, permiten la comparación entre grupos de la misma población o entre poblaciones diferentes.

3. Falso

   Está invertido. Una proporción explica la relación estadística que resulta de la división de un subgrupo entre toda la población.

4. Verdadero

   Una razón informa cuantas veces es mayor una clase con respecto de otra o la fracción que representa de la población.

5. Falso

   Las tasas brutas tienen como divisor a la población total, en cambio, las tasas específicas utilizan la población expuesta a sufrir el evento.

6. Falso

   Las pruebas de tamizaje son exámenes que se hacen a cualquier grupo de una población aparentemente saludable, buscando identificar a pacientes con una patología pero que aún no presentan síntomas.

7. Falso

   La sensibilidad mide la validez de una prueba de tamizaje entre los enfermos. Se obtiene por la proporción de pacientes con la patología que se identifica por medio del primer tamizaje.

8. Falso

   La especificidad mide la validez de una prueba de tamizaje entre los sanos. Se obtiene por la proporción de pacientes sanos que resulta negativo.

9. Falso

   Dan positivo los sospechosos de estar enfermos. El valor predictivo de una prueba positiva es el número de pacientes realmente enfermo y fue diagnosticado así en la prueba, dividido entre todos que resultaron positivos en el diagnóstico inicial.

10. Falso

    Dan negativo los libres de sospecha. El valor predictivo de una prueba negativa es el número de pacientes que está sano dividido entre todos los que dieron negativo, es decir, que desde el inicio se diagnosticaron como sanos.

11. Falso

    Los informes de morbilidad de una población proporcionan indicadores negativos del estado de salud de la colectividad)

12. Falso

    Las categorías no siempre son mutuamente excluyentes porque una persona puede padecer de más de una enfermedad al mismo tiempo y difícilmente son exhaustivas porque puede haber registros incompletos o existen enfermos que no acuden al sistema de salud)

13. Verdadero

    La incidencia de una enfermedad se mide por los casos nuevos. Un caso nuevo puede comenzar y terminar en el periodo analizado o iniciarse en el periodo y seguir más allá del estudio.

14. Verdadero

    La prevalencia informa del número de pacientes que padece una enfermedad, en un periodo y zona determinada. Lo que se busca es saber el número de individuos que padece una enfermedad, sin que importe la fecha de su inicio.

15. Falso

    La metáfora de la fotografía aplica para la prevalencia.

16. Verdadero

    La tasa de casos se calcula para enfermedades que pueden atacar varias veces al mismo paciente en el periodo.

17. Falso

La oportunidad relativa (OR) es la razón del éxito terapéutico de dos tratamientos.

18. Falso

Las tasas ajustadas varían cuando se usan diferentes poblaciones tipo. Esto no es un gran problema puesto que las tasas ajustadas solamente se usan para comparar poblaciones heterogéneas.

19. Verdadero

En el método directo de ajuste de tasas se utiliza una población tipo que se multiplica por las tasas específicas para el cálculo de las muertes esperadas, mientras que en el método indirecto se desconocen las tasas específicas de la población que se va a comparar.

20. Verdadero

La razón de estandarización es la división de los casos observados (reales) entre los casos esperados (hipotéticos). El supuesto básico de los eventos hipotéticos es que sus riesgos son equivalentes a los riesgos de una población tipo, de la cual se utilizan sus tasas específicas para hacer el ajuste y poder comparar poblaciones heterogéneas. La razón de estandarización se puede usar como índice de comparación directo.

EJERCICIO 3.11

Pareo. Llene los paréntesis con el número que acompaña a cada definición de la izquierda, de forma que cada definición coincida con su concepto. Quedan dos paréntesis en blanco.

1.	Medida que indica la proporción de pacientes sanos que resulta negativo en la prueba de tamizaje.	(3)	Morbilidad
2.	Condición o rasgo detectable que se asocia con el aumento en la probabilidad de padecer una enfermedad, desarrollar un proceso mórbido, sufrir un accidente o enfrentar la muerte prematuramente.	(10)	Proporción
3.	Ocurrencia de enfermedades en una población, su registro proporciona indicadores negativos del estado de salud de la colectividad)	(7)	Tasa de letalidad
4.	Es la razón del éxito terapéutico entre los dos tratamientos.	(9)	Constante de amplificación
5.	Indicador que se calcula dividiendo las defunciones ocurridas entre las defunciones esperadas de cada población, esto es, de la división de eventos reales entre eventos hipotéticos.	(2)	Factor de riesgo
6.	Medida que indica la proporción de pacientes con un padecimiento que se identifica por medio de una prueba de tamizaje.	(8)	Fracción atribuible
7.	Número relativo que mide la probabilidad de morir por una enfermedad una vez que se ha contraído.	(4)	Oportunidad relativa
8.	Razón que estima el grado de influencia que tiene un tratamiento como causa de la curación de los pacientes y que puede ser atribuible al hecho de seguirlo.	(1)	Especificidad de la prueba
9.	Valor arbitrario por el cual se multiplican los indicadores para hacerlos más comprensibles para el gran público.	(6)	Sensibilidad de la prueba
10.	Número relativo que informa del resultado de dividir un subgrupo entre toda la población.	( )	Tasas brutas
		(5)	Razón de estandarización
		( )	Valor predictivo de la prueba

CAPÍTULO 4

EJERCICIO 4.1

a) No existe una moda definida, puesto que los analgésicos se prescriben por igual.

Ácido acetilsalicílico (Aspirina)	4
Paracetamol (Tyenol)	4
Naproxeno (Aleve)	4
Ibuprofeno (Motrin)	4

b) Es bimodal porque tanto la clorfeniramina como la loratadina se ha prescrito a igual número de pacientes = 24.

c) El promedio es:

$$\bar{x} = \frac{16+4+7+19+5+6+6}{7} = \frac{63}{7} = 9 \text{ casos de rubéola}$$

La moda es igual a 6

La posición mediana $\frac{n+1}{2} = \frac{7+1}{2} = 4$

Ordenando: 4, 5, 6, 6, 7, 16, 19

La cuarta posición indica 6 casos de rubéola

La mediana es 6.

d) Las medidas de tendencia central del contenido de glucosa en la sangre del paciente en ayunas es:

$$\bar{x} = \frac{850}{10} = 85 \; mg/dL$$

Mo = 100 mg/dL

Me = 81 mg/dL

EJERCICIO 4.2

a) Medidas de posición:

i. En esta distribución se observa que la clase modal va de 3 499,5 a 3 999,5 libras, con frecuencia de 10 observaciones, por lo que la moda es:

$$Mo = 3\,499{,}5 + \left(\frac{3}{3+4}\right) 500$$

$$Mo = 3\,499{,}5 + 214{,}3$$

Moda igual a 3 713,8 gramos.

ii. El promedio del peso de los recién nacidos es:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i f_i}{\sum_{i=1}^{n} f_i} = \frac{129\,480}{40} = 3\,237 \text{ gramos}$$

iii. El valor $\frac{n}{2} = 20$ se ubica en la clase con el límite inferior de 2 999,5, la frecuencia absoluta de la clase en donde está la mediana es 7, la frecuencia acumulada "X o menos" de la clase anterior es 14 y el ancho de clase es 500:

CLASES	FRECUENCIAS ABSOLUTAS	
(Peso en gramos)	f_i	F_a
499,5 – 999,5	1	1
999,5 – 1 499,5	2	3
1 499,5 – 1 999,5	2	5
1 999,5 – 2 499,5	4	9
2 499,5 – 2 999,5	5	14
2 999,5 – 3 499,5	7	21
3 499,5 – 3 999,5	10	31
3 999,5 – 4 499,5	6	37
4 499,5 – 4 999,5	3	40
	40	

$$Me = 2\,999{,}5 + \left(\frac{20-14}{7} \times 500\right)$$

$$Me = 2\,999{,}5 + 428{,}6 = 3\,428{,}1 \text{ gramos}$$

Este resultado informa que la mitad de los recién nacidos pesa menos de 3 428,1 gramos y la otra mitad pesa más de 3 428,1 gramos.

b) Medidas de posición:

i. En esta distribución se observa que la clase modal va de 30 a menos de 35 libras, con frecuencia de 14 observaciones, por lo que la moda es:

$$Mo = 30 + \left(\frac{2}{2+5}\right)5$$

$$Mo = 30 + 1,4$$

Moda igual a 31,4 años.

ii. El promedio de edad de las parturientas es $\bar{x} = \frac{\sum_{i=1}^{n} x_i f_i}{\sum_{i=1}^{n} f_i} = \frac{1\,505}{50} = 30,1$ años

iii. El valor $\frac{n}{2} = 25$ se ubica en la clase con el límite inferior de 30, la frecuencia absoluta de la clase en donde está la mediana es 14, la frecuencia acumulada "X o menos" de la clase anterior es 23 y el ancho de clase es 5:

CLASES	FRECUENCIA ABSOLUTA	FRECUENCIA ACUMULADA
(Edades)	f_i	F_a
10 a menos de15	2	2
15 a menos de 20	4	6
20 a menos de 25	5	11
25 a menos de 30	12	23
30 a menos de 35	14	37
35 a menos de 40	9	46
40 a menos de 45	3	49
45 a menos de 50	1	50
	50	

$$Me = 30 + \left(\frac{25-23}{14} \times 5\right)$$

$$Me = 30 + 0,714 = 30,714 \text{ años}$$

0,7144 X 12 meses = 8,57 → 8 meses

0,5714 X 30 días = 17,1 → 17días

0,714 de un año equivale a 8 meses y 17 días, por lo que la edad mediana de las parturientas es 30 años, 8 meses y 17 días.

EJERCICIO 4.3

a) La distribución de edades en esta población es asimétrica, entre otras cosas, porque no existe igualdad de las principales medidas de tendencia central: moda, mediana y promedio.

La asimetría es positiva puesto que el promedio es mayor que la moda ($\bar{x} > \text{Mo} \rightarrow 28 > 13$ años).

Es una población joven, pero su promedio se ve afectado por la presencia de valores extremos que "jalan" la cola derecha de la distribución.

b) La asimetría es negativa porque el promedio es menor que la moda ($\bar{x} < \text{Mo}$). Esto hace que la cola izquierda de la distribución se prolongue más que la cola derecha, debido a la existencia de niños con bajo peso al nacer.

$$(\bar{x} = 3\ 237) \langle\ (Me = 3\ 428{,}1) \langle\ (Mo = 3\ 713{,}8)$$

c) Estos datos se explican por qué la distribución de los salarios en la sociedad no es simétrica, antes bien, lo típico es que presente una asimetría positiva en donde la moda es menor que el promedio.

Esto ocurre por la presencia de muchos salarios bajos que producen una moda de 301 225 colones, junto con la existencia de valores extremos, es decir, pocos salarios muy altos que estiran la cola derecha de la distribución de frecuencias.

$$(Mo = 301\ 225) \langle\ (\bar{x} = 600\ 000)$$

EJERCICIO 4.4

a) Obtenga:

i. El percentil 75 $(P_{0,75})$.

$$\%n = 0,75 \times 40 = 30$$

$$L_i = 3\,499,5$$

$$f_i = 10$$

$$F_{aa} = 21$$

$$c = 500$$

$$P_{\%} = L_i + \left(\frac{(\%n) - F_{aa}}{f_i} \times c\right)$$

$$P_{0,75} = 3\,499,5 + \left(\frac{30 - 21}{10} \times 500\right)$$

$$P_{0,75} = 3\,499,5 + 450$$

$$P_{0,75} = 3\,949,5$$

ii. El primer cuartil (Q_1).

$$\%n = \frac{1}{4} \times 40 = 10$$

$$L_i = 2\,499,5$$

$$f_i = 5$$

$$F_{aa} = 9$$

$$c = 500$$

$$Q_x = L_i + \left(\frac{\left(\frac{1}{4}n\right) - F_{aa}}{f_i} \times c\right)$$

$$Q_1 = 2\,499,5 + \left(\frac{10 - 9}{5} \times 500\right)$$

$$Q_1 = 2\,499,5 + 100$$

$$Q_1 = 2\,599,5$$

b) Obtenga:

i. El percentil 20 $(P_{0,20})$.

$$\%n = 0,20 \times 50 = 10$$

$$L_i = 20$$

$$f_i = 5$$

$$F_{aa} = 6$$

$$c = 5$$

$$P_{\%} = L_i + \left(\frac{(\%n) - F_{aa}}{f_i} \times c \right)$$

$$P_{0,20} = 20 + \left(\frac{10 - 6}{5} \times 5 \right)$$

$$P_{0,20} = 20 + 4$$

$$P_{0,20} = 24$$

ii. El percentil 80 $(P_{0,80})$.

$$\%n = 0,80 \times 50 = 40$$

$$L_i = 35$$

$$f_i = 9$$

$$F_{aa} = 37$$

$$c = 5$$

$$P_{\%} = L_i + \left(\frac{(\%n) - F_{aa}}{f_i} \times c \right)$$

$$P_{0,80} = 35 + \left(\frac{40 - 37}{9} \times 5 \right)$$

$$P_{0,80} = 35 + 1,67$$

$$P_{0,80} = 36,67$$

EJERCICIO 4.5

a) Rango percentil.

i. El valor 3 949,5 libras se encuentra en la sétima clase.

$X = 3\ 949,5$ $\quad L_i = 3\ 499,5$ $\quad f_i = 10$

$F_{aa} = 21$ $\quad c = 500$ $\quad n = 40$

Sustituyendo en:

$$r = \left(\frac{(X - L_i) \times f_i}{c} + F_{aa}\right) \times \frac{100}{n}$$

$$r = \left(\frac{(3\ 949,5 - 3\ 499,5) \times 10}{500} + 21\right) \times \frac{100}{40}$$

$$r = \left(\frac{(450) \times 10}{500} + 21\right) \times \frac{100}{40}$$

$$r = (9 + 21) \times \frac{100}{40}$$

$$r = 30 \times \frac{100}{40}$$

$$r = 75$$

Un bebé que pesa 3 949,5 gramos al nacer se encuentra en el percentil 75 $(P_{0,75})$.

ii. El valor 2 599,5 gramos se encuentra en la quinta clase.

$X = 2\ 599,5$ $\quad L_i = 2\ 499,5$ $\quad f_i = 5$

$F_{aa} = 9$ $\quad c = 500$ $\quad n = 40$

Sustituyendo en:

$$r = \left(\frac{(X - L_i) \times f_i}{c} + F_{aa}\right) \times \frac{100}{n}$$

$$r = \left(\frac{(2\,599{,}5 - 2\,499{,}5) \times 5}{500} + 9\right) \times \frac{100}{40}$$

$$r = \left(\frac{100 \times 5}{500} + 9\right) \times \frac{100}{40}$$

$$r = (1 + 9) \times \frac{100}{40}$$

$$r = 10 \times \frac{100}{40}$$

$$r = 25$$

Un bebé que pesa 2 599,5 gramos al nacer se encuentra en el percentil 5 ($P_{0,25}$).

b) Rango percentil.

i. El valor 24 años se encuentra en la tercera clase.

$X = 24$ $L_i = 20$ $f_i = 5$

$F_{aa} = 6$ $c = 5$ $n = 50$

Sustituyendo en:

$$r = \left(\frac{(X - L_i) \times f_i}{c} + F_{aa}\right) \times \frac{100}{n}$$

$$r = \left(\frac{(24 - 20) \times 5}{5} + 6\right) \times \frac{100}{50}$$

$$r = \left(\frac{4 \times 5}{5} + 6\right) \times \frac{100}{50}$$

$$r = 10 \times \frac{100}{50}$$

$$r = 20$$

Una que tiene 24 años al momento de dar a luz se encuentra en el percentil 20 ($P_{0,20}$).

ii. El valor 40 años se encuentra en la sétima clase.

$X = 40$ $L_i = 40$ $f_i = 3$

$F_{aa} = 46$ $c = 5$ $n = 50$

Sustituyendo en:

$$r = \left(\frac{(X - L_i) \times f_i}{c} + F_{aa}\right) \times \frac{100}{n}$$

$$r = \left(\frac{(40 - 40) \times 3}{5} + 46\right) \times \frac{100}{50}$$

$$r = \left(\frac{0 \times 3}{5} + 46\right) \times \frac{100}{50}$$

$$r = 46 \times \frac{100}{50}$$

$$r = 92$$

Una que tiene 40 años al momento de dar a luz se encuentra en el percentil 92 ($P_{0,92}$).

EJERCICIO 4.6

Nivel de glucosa en la sangre de un paciente en ayunas:

x_i	$x_i - \bar{x}$	$\lvert x_i - \bar{x} \rvert$
75	-10	10
100	15	15
92	7	7
100	15	15
76	-9	9
80	-5	5
70	-15	15
75	-10	10
82	-3	3
100	15	15
SUMA	0	104

i. $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{850}{10} = 85\ mg/dL$

ii. $R = L_s - L_i = 100 - 70 = 30$

ii. $DM = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n} = \frac{104}{10} = 10,4$

EJERCICIO 4.7

Se tiene una muestra de observaciones acerca del nivel de glucosa en la sangre de un paciente en ayunas:

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
75	-10	100
100	15	225
92	7	49
100	15	225
76	-9	81
80	-5	25
70	-15	225
75	-10	100
82	-3	9
100	15	225
SUMA	0	1 264

Desviación estándar de esta muestra.

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{1\ 264}{10-1}}$$

$$s = \sqrt{\frac{1\ 264}{9}} = \sqrt{140,444} = 11,85$$

La desviación estándar del nivel de glucosa en la sangre de este paciente en ayunas es 11,85 mg/dL.

EJERCICIO 4.8

a) Variabilidad en la distribución de frecuencias del peso de una muestra de 40 recién nacidos en una clínica rural.

i. Variancia.

$$s^2 = \frac{\left[\sum_{i=1}^{n}(x_i^2 f_i)\right] - \left[\frac{\left(\sum_{i=1}^{n} x_i f_i\right)^2}{n}\right]}{n-1}$$

$$s^2 = \frac{458\ 370\ 510 - \left[\frac{(129\ 480_i)^2}{40}\right]}{40-1}$$

$$s^2 = \frac{458\ 370\ 510 - 419\ 126\ 760}{39}$$

$$s^2 = \frac{39\ 243\ 750}{39}$$

$$s^2 = 1\ 006\ 250$$

ii. Desviación estándar del peso de los recién nacidos.

$$s = \sqrt{1\ 006\ 250} = 1\ 003{,}12 \text{ gramos}$$

b) Edad en años cumplidos de 50 mujeres al momento de dar a luz en el Hospital de la Anexión, durante diciembre de 2014.

i. Variancia.

$$s^2 = \frac{\left[\sum_{i=1}^{n}(x_i^2 f_i)\right] - \left[\frac{\left(\sum_{i=1}^{n} x_i f_i\right)^2}{n}\right]}{n-1}$$

$$s^2 = \frac{48\ 262{,}5 - \left[\frac{(1\ 505_i)^2}{50}\right]}{50-1}$$

$$s^2 = \frac{48\ 262{,}5 - 45\ 300{,}5}{49}$$

$$s^2 = \frac{2\ 962}{49}$$

$$s^2 = 60{,}45$$

ii. La desviación estándar de la edad de la mujer al momento de dar a luz.

$$s = \sqrt{60{,}45} = 7{,}8 \text{ años}$$

EJERCICIO 4.9

i. Límite inferior = $\bar{x} - m\sigma$.

$$L_i = 3\ 237 - (1{,}23 \times 1\ 003{,}12)$$

$$L_i = 2\ 003{,}16 \text{ gramos}$$

Cuando el peso al nacer sea igual o menor a 2 003,16 gramos, deberán activarse todas las alarmas medicas para dedicarle máxima atención.

ii. Límite superior = $P_{0,75} + [m(P_{0,75} - P_{0,25})]$.

$$L_s = 200 + [1{,}8 \times (200 - 140)]$$

$$L_s = 200 + [1{,}8 \times 60]$$

$$L_s = 200 + 108 = 308 \text{ mg/dL}$$

Cuando el nivel de colesterol supere los 308 mg/dL, el paciente deberá ser intervenido clínicamente.

EJERCICIO 4.10

i. De una muestra de 300 pacientes atendidos en el servicio de emergencias en el Hospital Nacional de Niños, 96 fueron atendidos con problemas respiratorios. Calcule:

a) La proporción de pacientes atendidos por problemas respiratorios.

$$P = \frac{96}{300} = 0{,}32$$

b) La variancia de esa proporción.

$$s^2 = P\,(1 - P) = 0{,}32 \times (1 - 0{,}32) = 0{,}32 \times 0{,}68 = 0{,}2176$$

c) La desviación estándar.

$$s = \sqrt{P\,(1 - P)}$$

$$s = \sqrt{0{,}32 \times 0{,}68} = \sqrt{0{,}2176} = 0{,}466$$

ii. De una muestra de 1 000 personas, se determinó que 600 tenían sangre tipo O. Calcule:

a) La proporción de pacientes con sangre tipo O.

$$P = \frac{600}{1\,000} = 0,60$$

b) La variancia de esa proporción.

$$s^2 = P\,(1-P) = 0,6 \times (1-0,6) = 0,6 \times (0,4) = 0,24$$

c) La desviación estándar.

$$s = \sqrt{P\,(1-P)}$$

$$s = \sqrt{0,6 \times 0,4} = \sqrt{0,24} = 0,4899$$

iii. De una muestra de 10 escolares de sexto grado, 8 presentan caries dentales o secuelas de las mismas. Calcule:

a) La proporción de escolares que padece caries o secuelas de esta enfermedad)

$$P = \frac{8}{10} = 0,80$$

b) La variancia de esa proporción.

$$s^2 = \left(\frac{n}{n-1}\right)P\,(1-P)$$

$$s^2 = \left(\frac{10}{10-1}\right) \times 0,8 \times (1-0,8)$$

$$s^2 = 1,111 \times 0,8 \times 0,2 = 0,1778$$

c) La desviación estándar.

$$s = \sqrt{s^2}$$

$$s = \sqrt{0,1778} = 0,4216$$

EJERCICIO 4.11

Si al finalizar abril, se tienen reportados 16 casos o más, deberán activarse las alarmas sanitarias para enfrentar la epidemia de rubéola.

Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Ago.	Set	Oct	Nov	Dic	
0	0	0	0	4	0	0	0	0	0	0	0	
2	2	2	2	5	1	0	1	1	0	0	0	Primer Cuartil
3	4	7	7	6	4	1	2	1	1	1	0	
3	6	8	8	6	7	4	2	2	2	2	0	Mediana
3	7	10	12	7	7	4	2	2	2	3	0	
5	11	12	16	16	8	7	4	5	4	3	0	Tercer cuartil
7	12	18	22	19	10	11	4	5	4	5	0	

La epidemia se considerará controlada si para mayo solamente se tienen 5 casos de rubéola o menos.

EJERCICIO 4.12

Falso (F) o Verdadero (V).

1. Verdadero

   El análisis y la interpretación de los datos se facilita con la obtención de medidas que resumen las principales características de una distribución.

2. Falso

   La *moda* es la cantidad o cualidad que más se repite en un conjunto de datos.

   La *mediana* es el valor que divide en dos grupos iguales a una serie ordenada, la mitad de las observaciones tendrá valores menores que la mediana y la otra mitad tendrá valores mayores.

3. Falso

   Existe un único promedio para un conjunto de datos.

4. Falso

   La suma de las desviaciones respecto del promedio de todos los datos siempre da cero.

5. Verdadero

   Si bien el promedio toma en cuenta todos y cada uno de los valores del conjunto, la presencia de valores extremos "jalan" la distribución influyéndola grandemente.

6. Falso

   Los puntos medios de cada clase son solamente un valor representativo, alrededor de los puntos medios existe variabilidad y, por ello, la media aritmética para datos agrupados puede diferir ligeramente del resultado considerando todos los datos.

7. Falso

   Si la distribución es simétrica, el promedio, la moda y la mediana coincidirán en el centro de la distribución.

8. Verdadero

   Cuando el promedio es mayor que la moda, el coeficiente de asimetría planteado por Karl Pearson $\left(A_p = \frac{\bar{x} - Mo}{s}\right)$ será positivo.

9. Falso

   La asimetría "positiva" no implica un juicio de valor sobre los servicios, simplemente señala que la cola derecha de la distribución se alarga, haciendo que el promedio alcance un valor mayor que el de la moda.

10. Verdadero

    Tanto la mediana (Me) como el percentil cincuenta ($P_{0,50}$) informan del valor central de un conjunto ordenado de datos.

11. Falso

    El percentil 75 ($P_{0,75}$) es equivalente con el tercer cuartil (Q_3), también puede afirmarse que el percentil 25 ($P_{0,25}$) es equivalente al primer cuartil (Q_1).

12. Falso

    La variabilidad es un concepto fundamental del análisis estadístico también en el campo de la salud, puesto que los hechos se repiten pero lo hacen con alguna variación.

13. Falso

    El rango o recorrido solamente considera los valores extremos de un conjunto de datos y es la diferencia entre el valor mayor y el menor de la distribución.

14. Falso

    El rango o recorrido tiene como ventajas la sencillez y rapidez de cálculo, aún cuando se trate de un número grande de datos; también es de muy fácil compresión.

15. Verdadero

    La desviación media es la sumatoria de los valores absolutos de las desviaciones respecto del promedio que luego se divide entre el número de datos.

16. Falso

La desviación estándar nunca tendrá valor negativo, esto por cuanto todas las desviaciones respecto del promedio se elevan al cuadrado adoptando siempre valores positivos.

17. Verdadero

Son las medidas de posición y variabilidad más utilizadas, además, si el promedio está expresado en gramos, la desviación estándar también estará expresada en gramos; si el promedio se mide en centímetros, la desviación estándar se medirá en centímetros; si el promedio se indica en grados centígrados, lo mismo ocurrirá con la desviación estándar.

18. Falso

La variancia es la media aritmética del cuadrado de las desviaciones respecto del promedio de una distribución.

Recuerde que en el caso de poblaciones, todas las desviaciones al cuadrado se suman y su resultado se divide entre N.

19. Verdadero

Para calcular el coeficiente de variación se tiene que dividir la desviación estándar entre el promedio, con lo cual se pueden comparar dos o más conjuntos de datos en cuanto a su variabilidad)

20. Verdadero

El cálculo de límites de tolerancia permite identificar las fronteras de aceptación o de activación de alarmas.

EJERCICIO 4.13

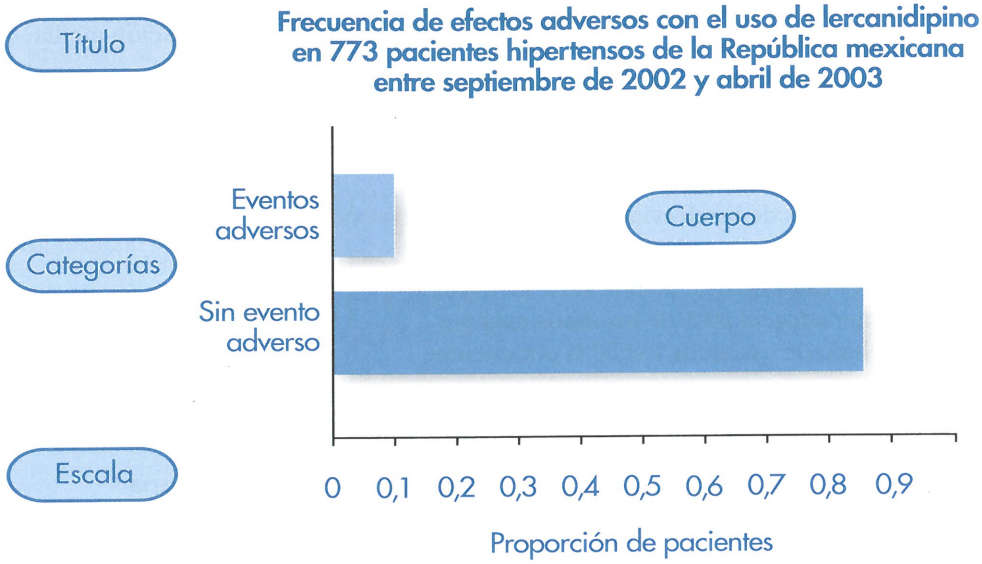
Pareo. Llene los paréntesis con el número que acompaña a cada definición de la izquierda, de forma que cada definición coincida con el concepto en la derecha. Quedan dos paréntesis en blanco.

1.	Explica la importancia de la desviación estándar de una serie de datos en relación con su promedio.	(3)	La desviación media
2.	Son los valores que dividen a un conjunto ordenado de datos en fracciones de igual magnitud)	(6)	Los percentiles
3.	Es la media aritmética de los valores absolutos de las desviaciones respecto del promedio.	(5)	La moda
4.	Es el valor central de una serie de datos ordenados, el 50 % de los datos está por debajo de ese valor y el 50 % por encima.	(8)	El rango
5.	Es la cantidad o cualidad que más se repite en un conjunto de datos; técnicamente se define como el valor más frecuente.	(7)	El promedio
6.	Son los valores que dividen un conjunto ordenado de datos en cien partes iguales.	(9)	La variancia
7.	Es el valor que se obtiene de la sumatoria de todos los datos dividida entre el número de observaciones.	(1)	El coeficiente de variación
8.	Valor que informa del intervalo contenido entre el valor mayor y el menor de los datos.	(10)	Los deciles
9.	Es la media aritmética del cuadrado de las desviaciones respecto del promedio de una distribución.	( )	La asimetría
10.	Son los valores que dividen un conjunto ordenado de datos en diez partes iguales.	( )	El apuntamiento
		(2)	Los cuantiles
		(4)	La mediana

CAPÍTULO 5

EJERCICIO 5.1

a) Elementos fundamentales que ayudan a la adecuada comprensión de un gráfico:



Fuente: Lepe Montoya (2004)

b) Elementos explicativos en la redacción del título:



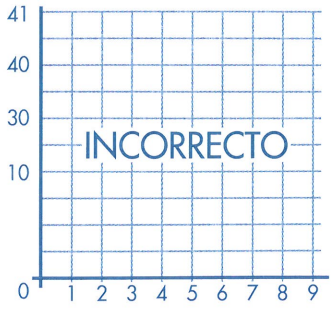
c) Elementos adicionales que puede contener un gráfico para mejorar la calidad de los informes:

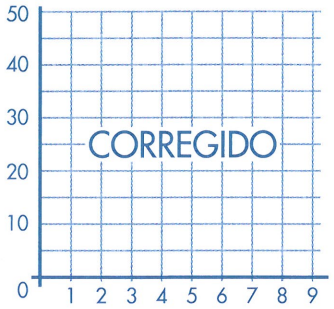
- Número de gráfico
- Nota introductoria
- Simbología
- Fuente
- Cortes en la escala de los ejes

EJERCICIO 5.2

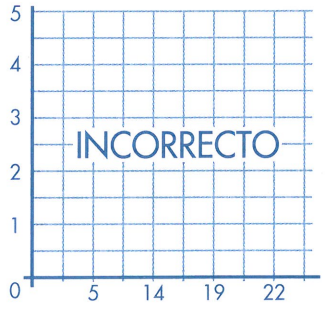
Corrija los defectos de escala que se pretende utilizar para los siguientes gráficos:

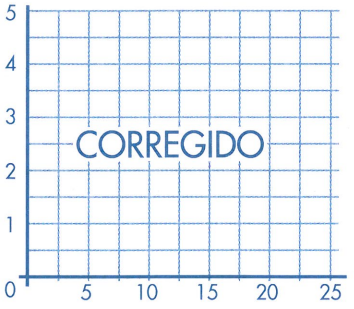
a.



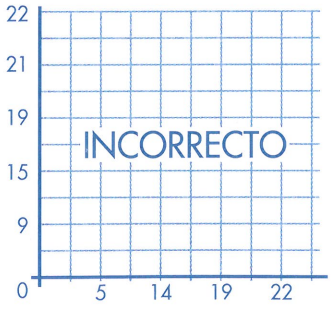


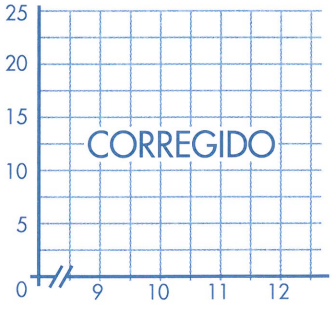
b.





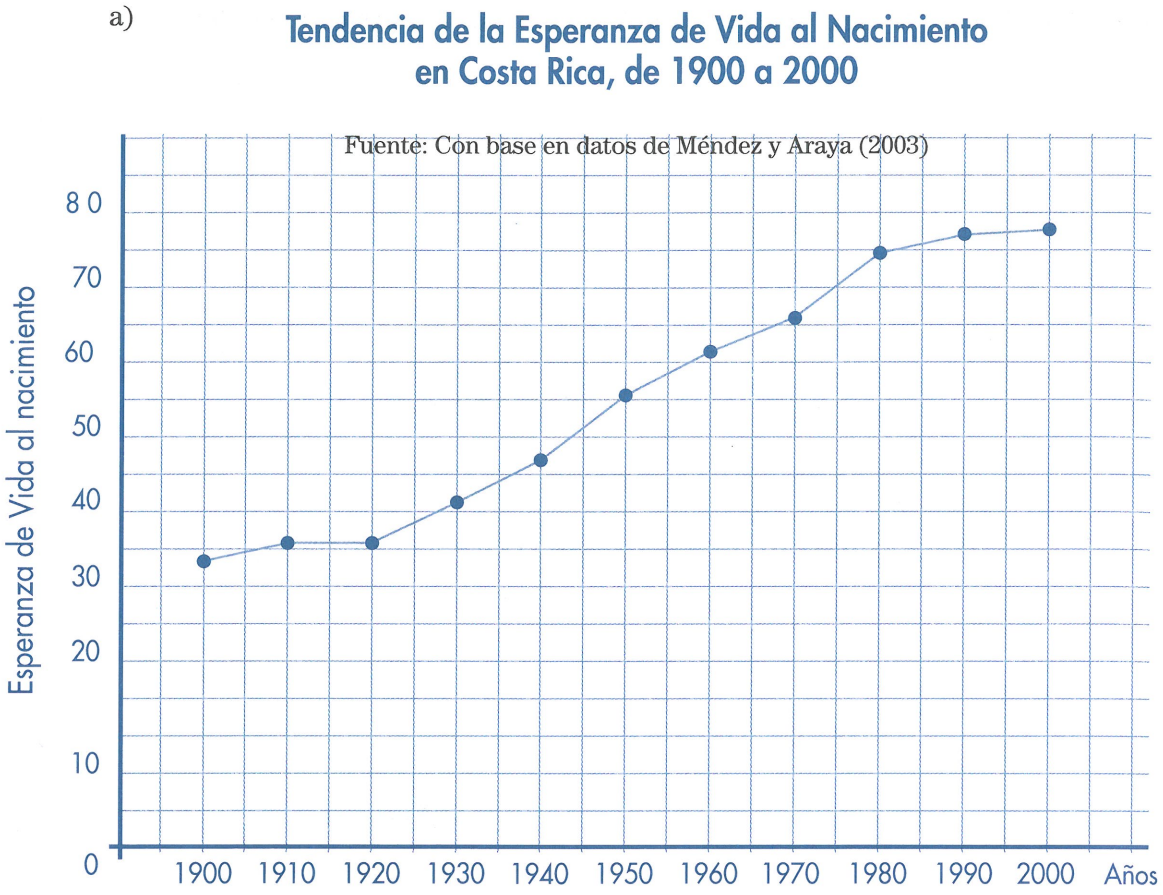
c.





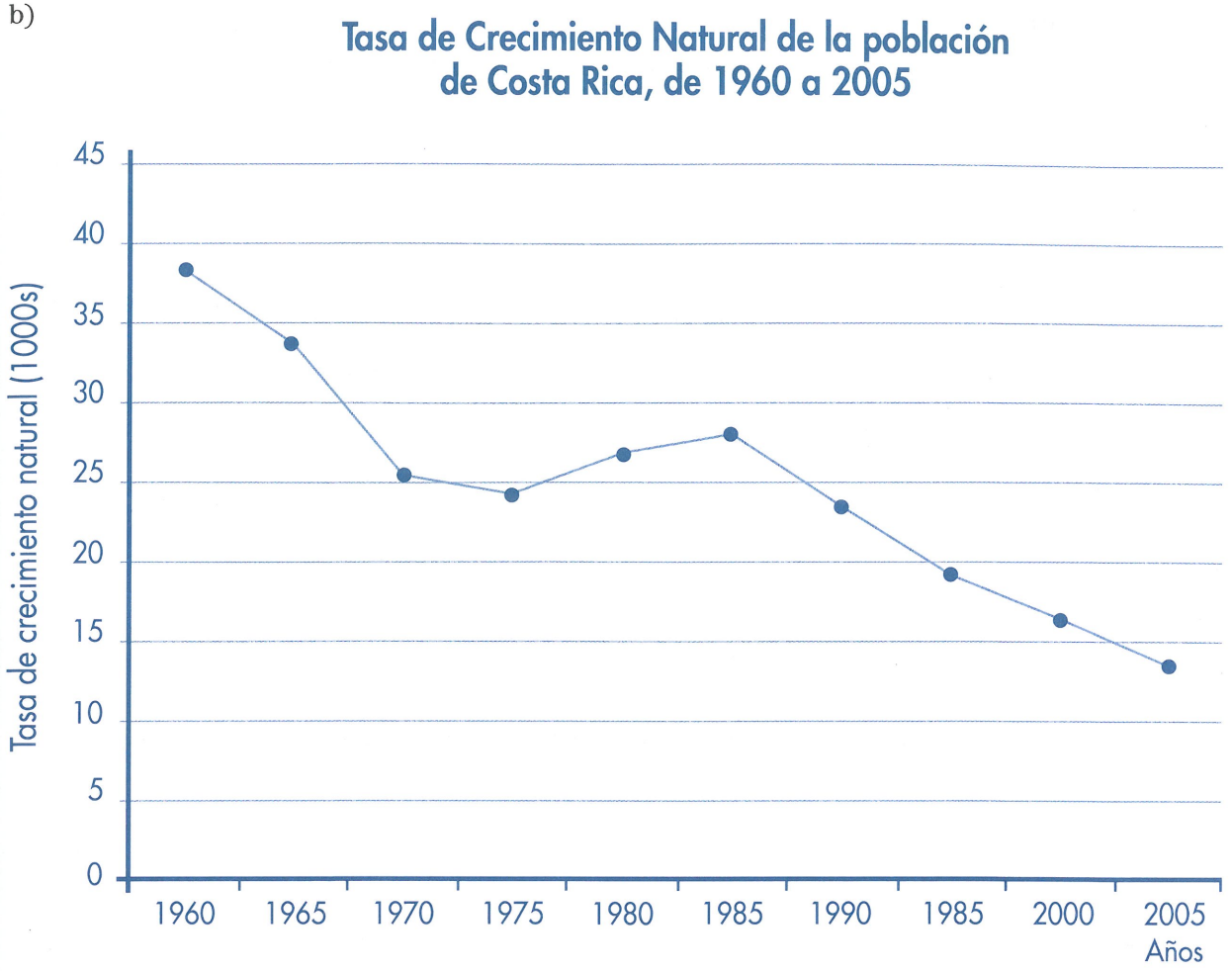
EJERCICIO 5.3

a)



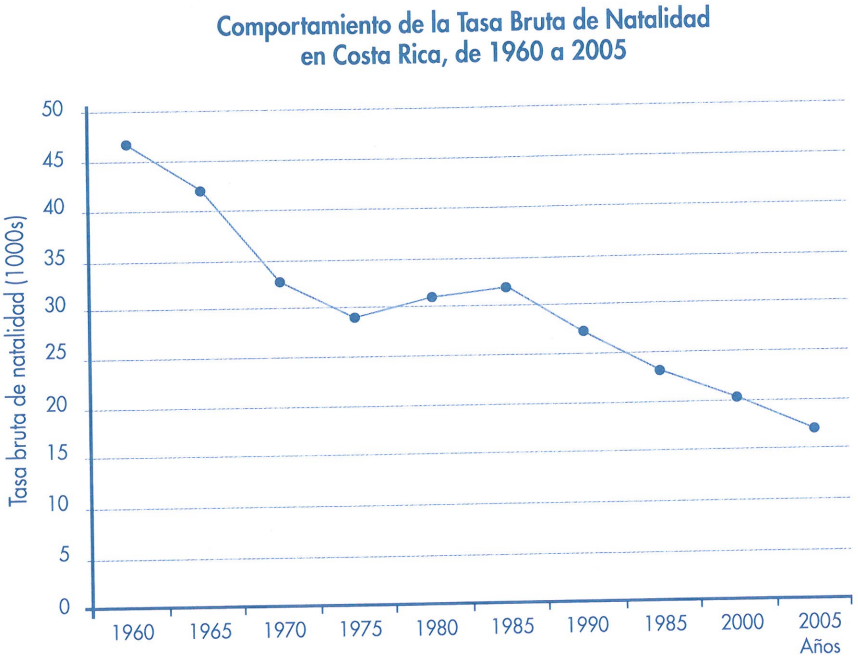
Fuente: Con base en datos de Méndez y Araya (2003)

b)



Fuente: Con base en datos del Centro Centroamericano del Población (2002)

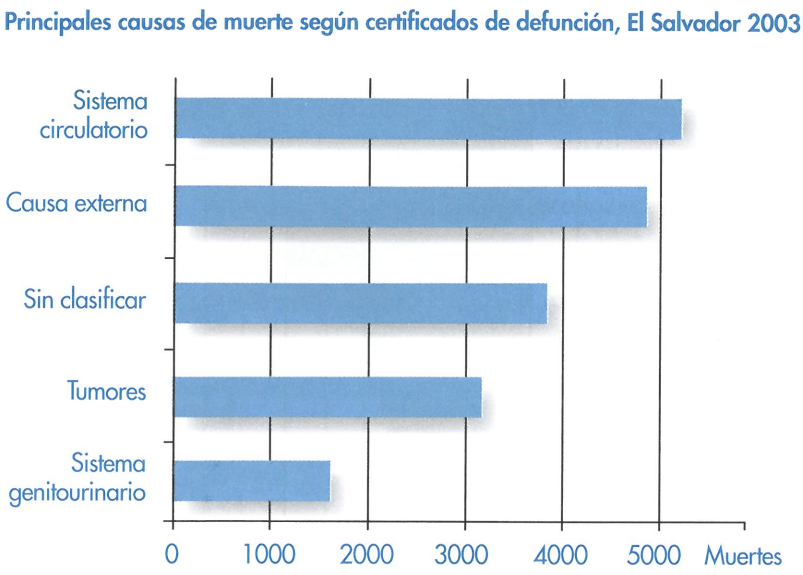
c)



Fuente: Con base en datos del Centro Centroamericano del Población (2002)

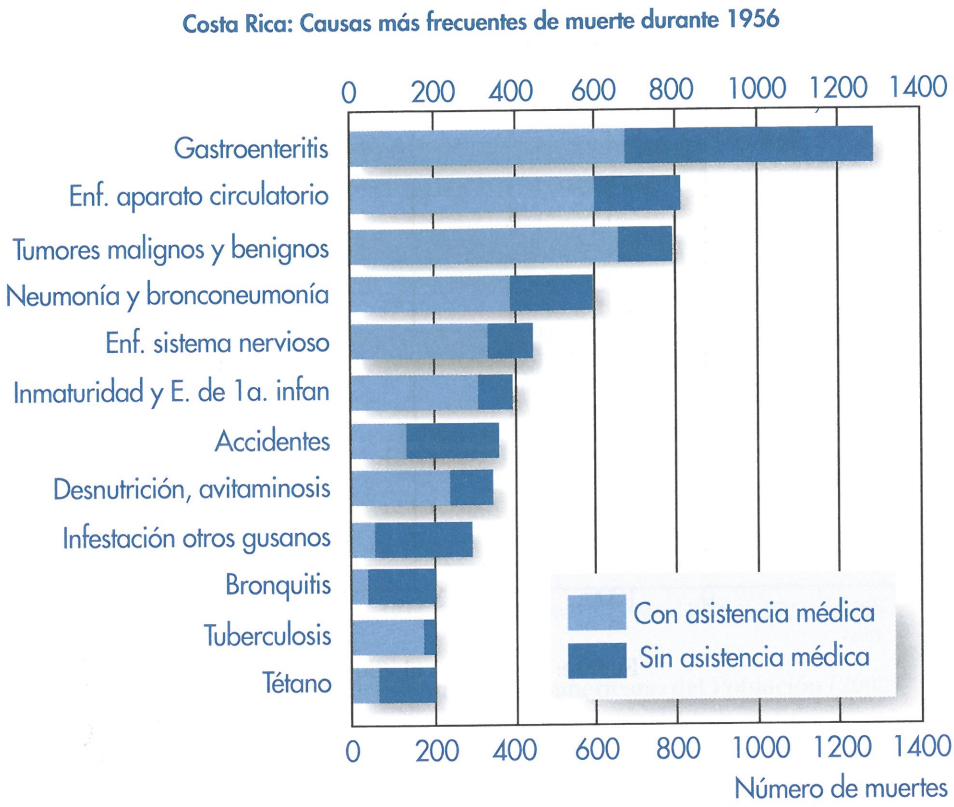
EJERCICIO 5.4

a)



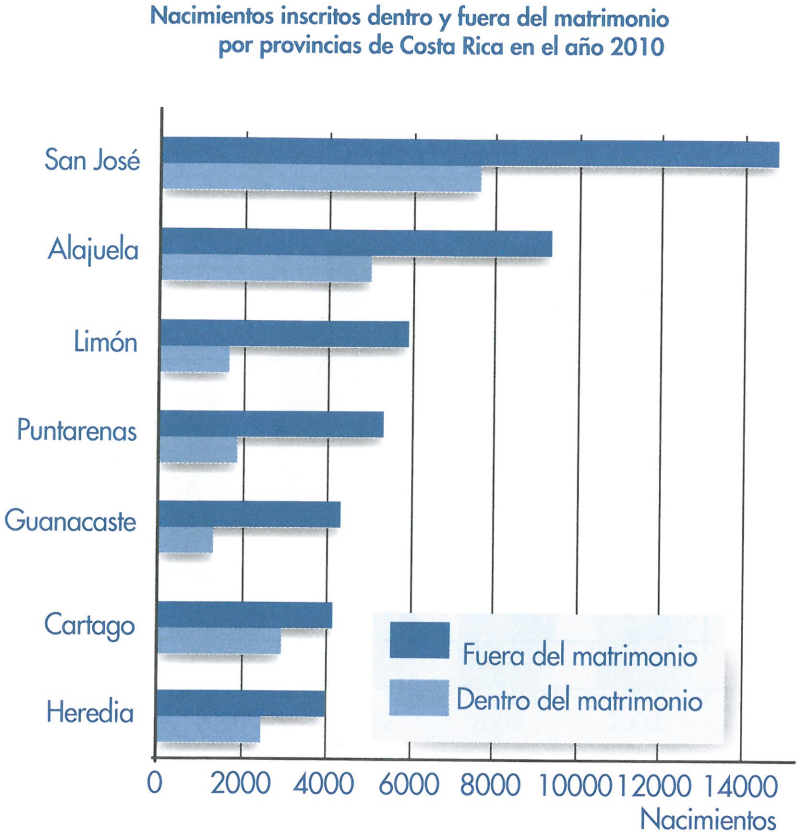
Fuente: Méndez et. al. (2006)

b)



Fuente: <http://ccp.ucr.ac)cr/bvp/pdf/anuariocr/an1956/anuario-56_03.pdf>

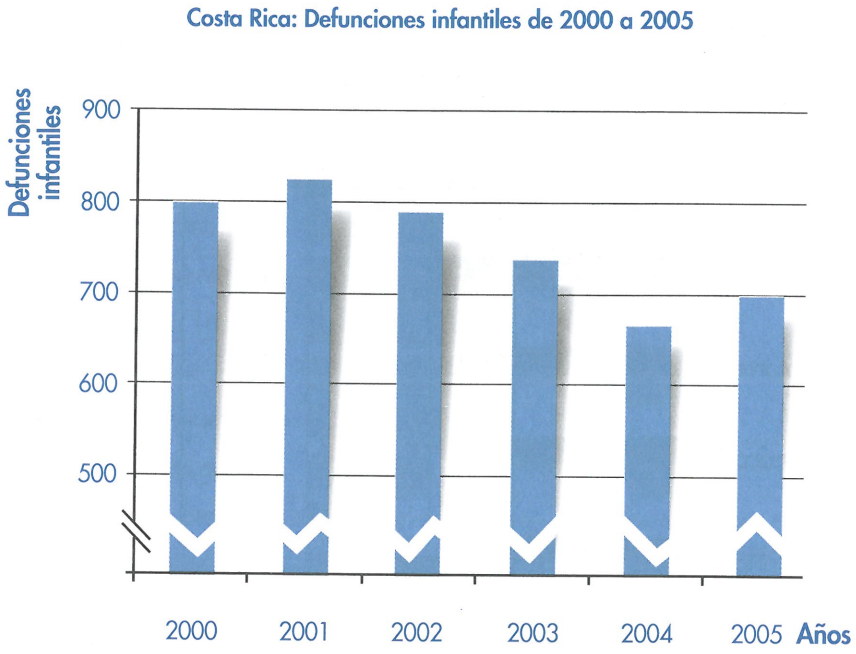
c)



Fuente: Con base en datos del INEC (2011)

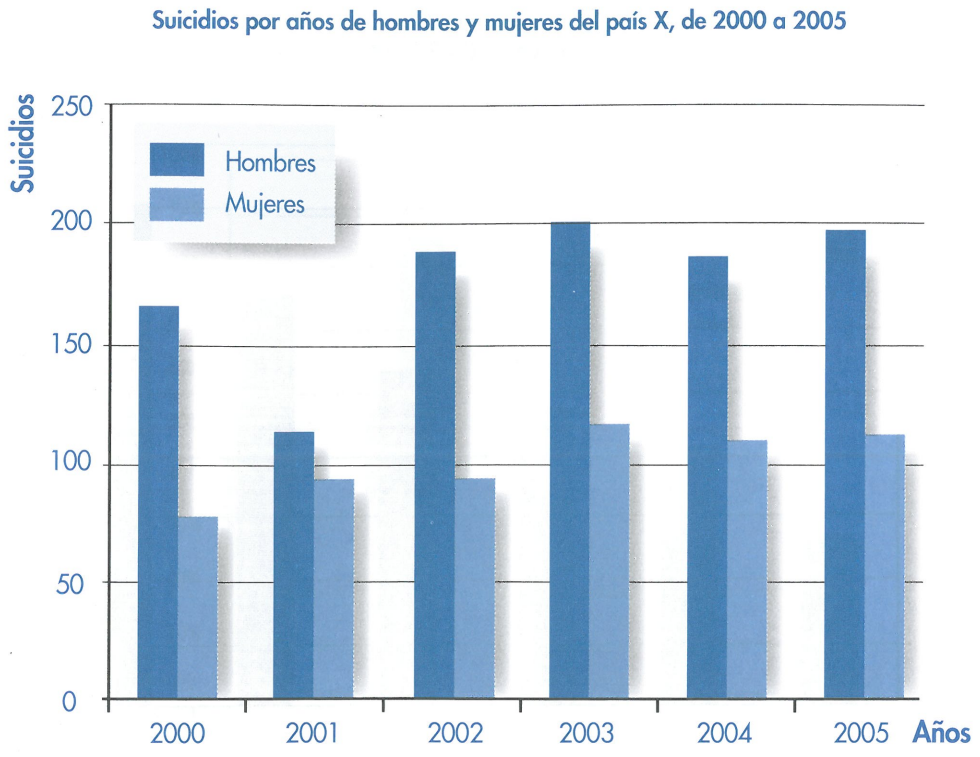
EJERCICIO 5.5

a)



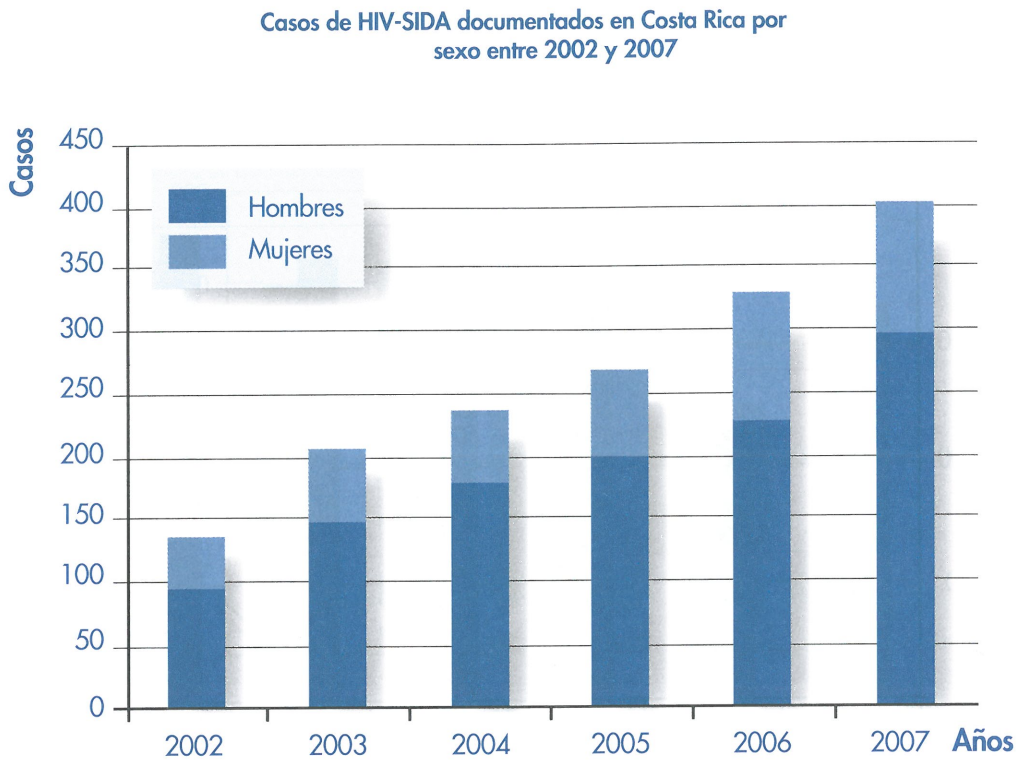
Fuente: Con base en datos del INEC (2011)

b)



Fuente: Datos hipotéticos

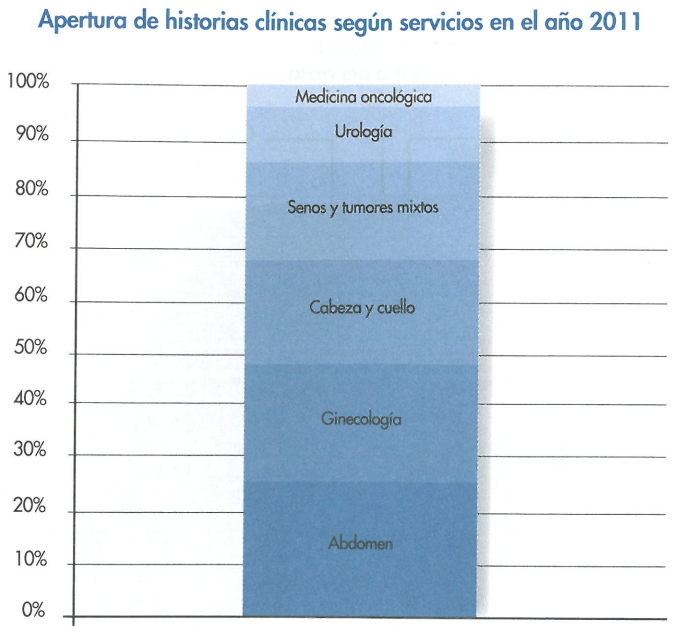
c))



Fuente: Vigilancia de la Salud, Ministerio de Salud (2010)

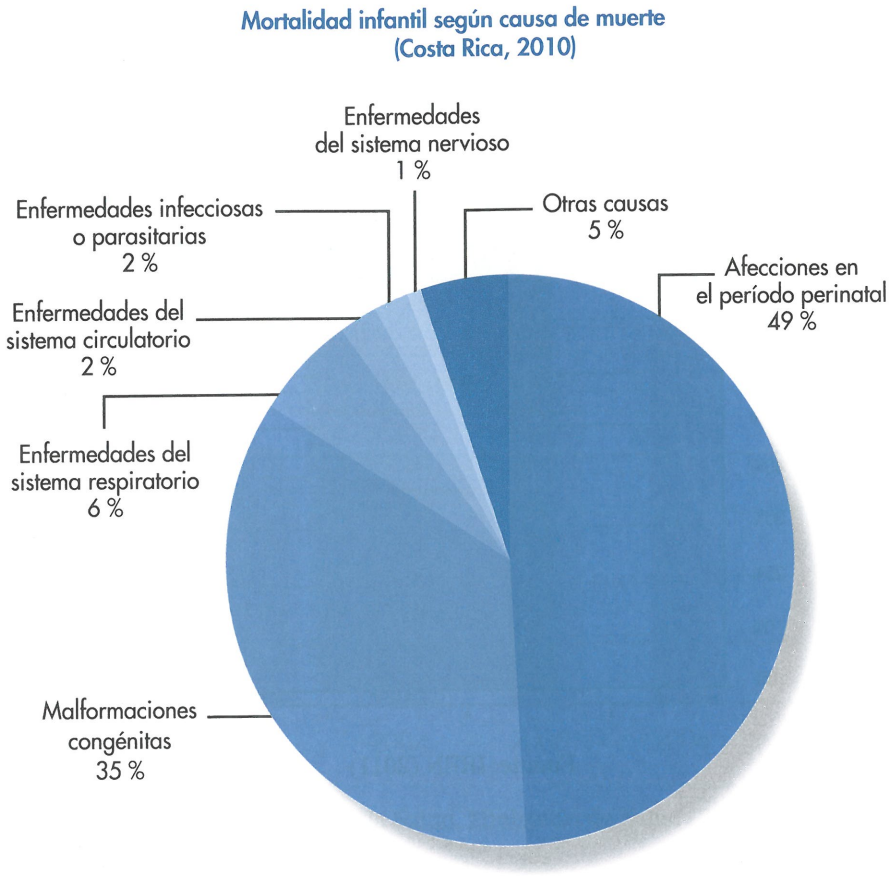
EJERCICIO 5.6

a)



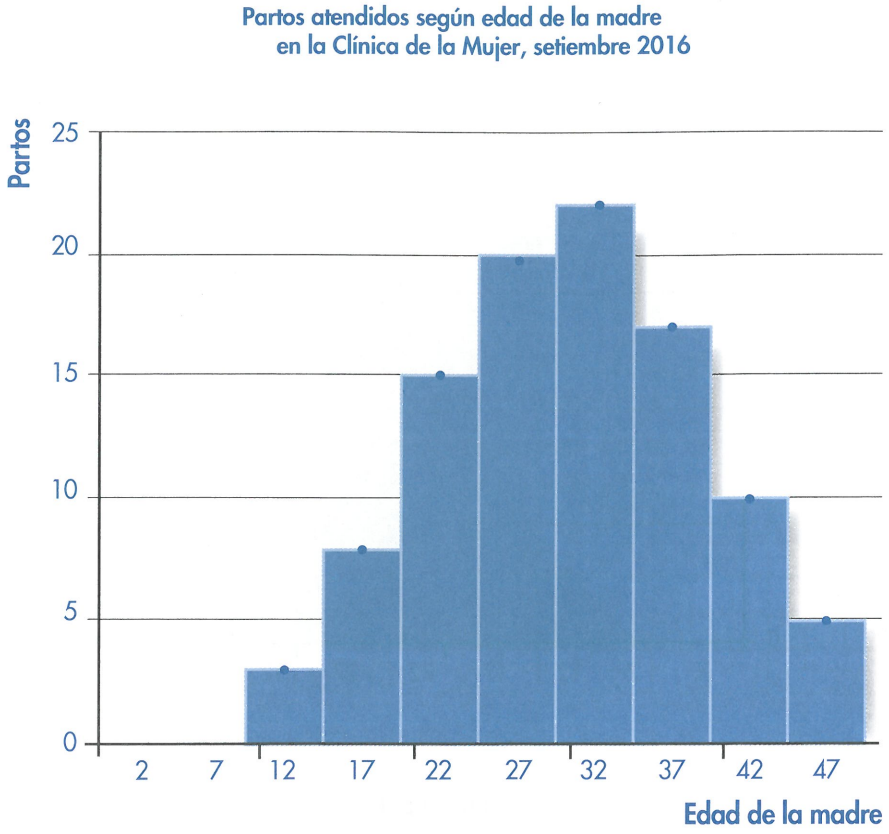
Fuente: IREN (2011)

b)

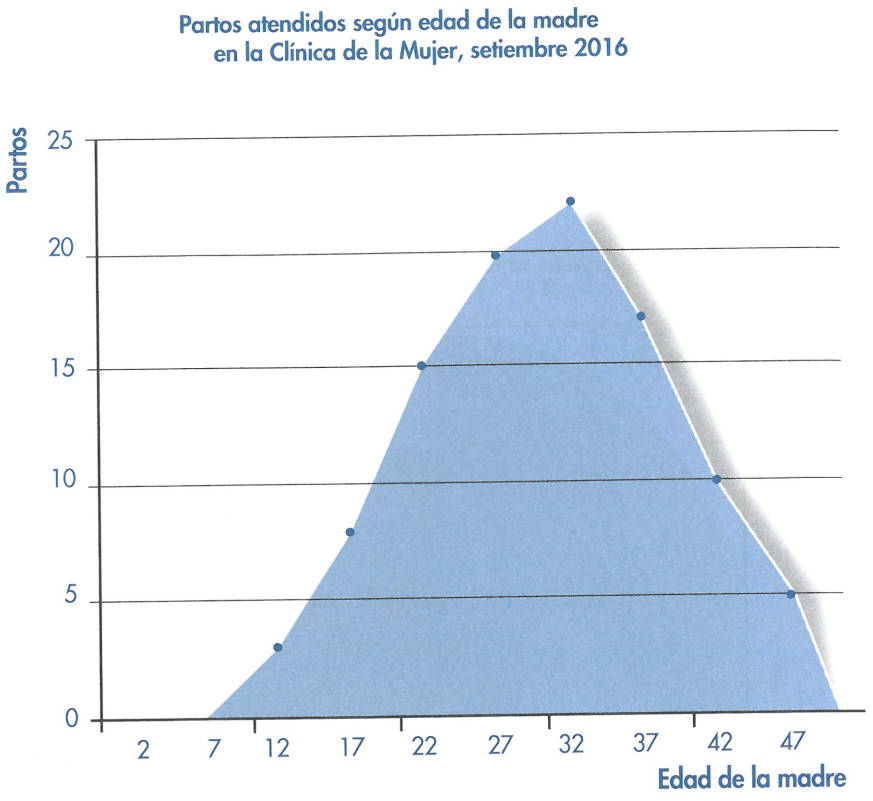


Fuente: Ministerio de Salud, Dirección de Vigilancia de la Salud (2011)

EJERCICIO 5.7



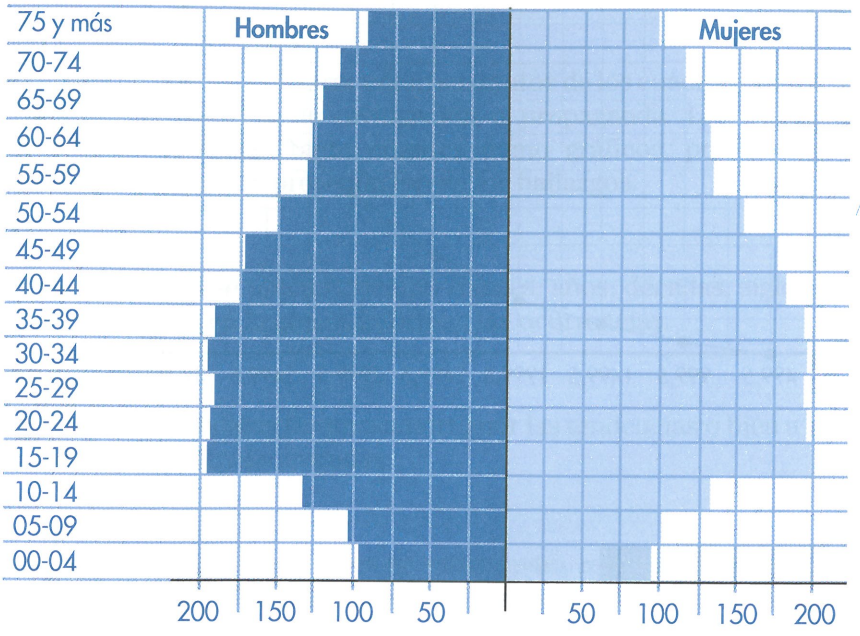
Fuente: Datos hipotéticos



Fuente: Datos hipotéticos.

EJERCICIO 5.8

Pirámide de población del país Y para el año 2010

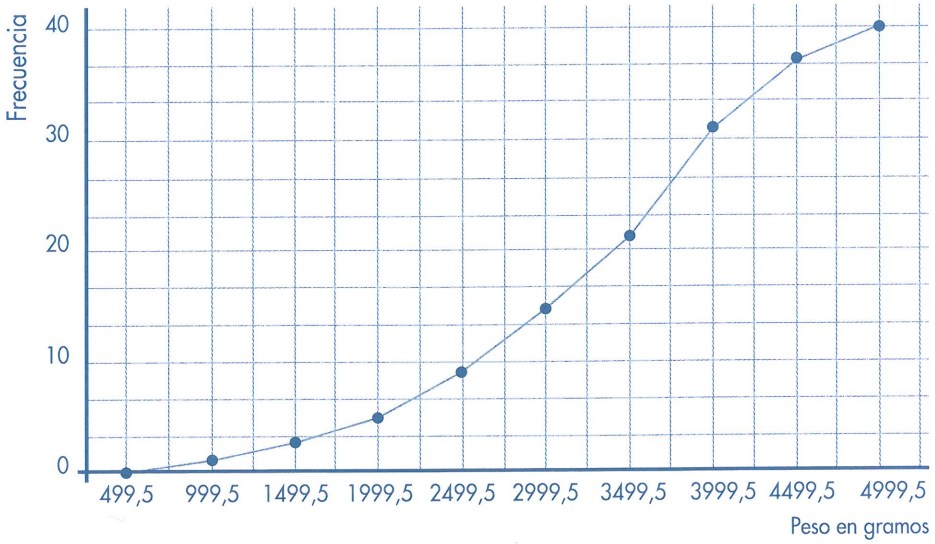


EJERCICIO 5.9

Frecuencias absolutas y acumuladas al límite superior del peso en gramos de los recién nacidos, Clínica de Turrucalito, agosto 2013

CLASES	FRECUENCIAS ABSOLUTAS	
(Peso en gramos)	Simples	Acumuladas
499,5 – 999,5	1	1
999,5 – 1 499,5	2	3
1 499,5 – 1 999,5	2	5
1 999,5 – 2 499,5	4	9
2 499,5 – 2 999,5	5	14
2 999,5 – 3 499,5	7	21
3 499,5 – 3 999,5	10	31
3 999,5 – 4 499,5	6	37
4 499,5 – 4 999,5	3	40
	40	

Fuente: Datos hipotéticos



EJERCICIO 5.10

a) Gráfico de columnas.

b) Gráfico de barras.

c) Gráfico circular, de barra 100 %.

d) Gráfico de columnas comparativas.

EJERCICIO 5.11

Encierre en un círculo la opción que se ajusta al enunciado, Falsa (F) o Verdadera (V).

1. Verdadero

   Una imagen dice más que mil palabras; sin embargo, esta imagen debe estar bien confeccionada, utilizando adecuadamente las distancias, las pendientes, las áreas y los volúmenes.

2. Falso

   El título es un elemento esencial de todo gráfico ya que debe indicar qué es lo que se grafica, cuándo de recolectaron los datos, dónde y cómo se realizó la recolección y el procesamiento de la información.

3. Falso

   La información que transmiten los gráficos solo es aproximada y se recomienda no incluir muchas variables pues lo recargarían volviéndolo confuso.

4. Falso

   Los gráficos no pueden utilizarse como sustitutos de los cuadros porque reducirían los informes a panfletos de divulgación o propaganda. Los buenos informes de investigación utilizan tanto cuadros como gráficos, por ser herramientas complementarias para la presentación de los hallazgos.

5. Verdadero

   En el eje horizontal se ubica el tiempo (meses, años, décadas, siglos) y en el eje vertical va la escala en la que se informan las ocurrencias.

6. Verdadero

   Sobrepasar estas medidas llevan a amortiguar las tendencias o bien a exagerarlas, situaciones que deben ser evitadas.

7. Falso

   El gráfico de barras horizontales se utiliza para representar series cualitativas y series geográficas, dada la posibilidad de escribir el nombre completo de las categorías a lo largo sin tener que recurrir a abreviaturas ni a códigos, que podrían hacer confusa la lectura o interpretación del diagrama.

8. Verdadero

   Esta es una licencia que tiene este tipo de gráfico, ya que interesa destacar aquellas categorías de frecuencias mayores.

9. Verdadero

   Las barras compuestas se utilizan cuando lo que interesa es mostrar los totales de cada categoría.

10. Verdadero

    Se limita la apreciación del total por categoría, pero permiten hacer comparaciones en forma rápida.

11. Falso

    Los gráficos de barras compuestas y los de barras comparativas no necesariamente compiten entres sí, antes bien, se complementan o pueden usarse alternamente aprovechando las ventajas comunicativas de unos y otros.

12. Falso

    Para ambos tipos, se recomienda no utilizar más de tres componentes puesto que cargan al gráfico y dificultan la comparación.

13. Falso

Para series cronológicas y variables cuantitativas discretas, un gráfico adecuado es el de columnas y no el de barras horizontales.

14. Falso

Lo que informa es la longitud de las barras o de las columnas, no su ancho. Es más, el ancho puede reducirse tanto hasta llegar a bastones y no se afecta la información proporcionada por el gráfico.

15. Falso

Cuando lo que interesa representar es la participación porcentual de las categorías con respecto del total, se puede utilizar la barra 100 % y el gráfico circular.

16. Verdadero

La barra 100 % busca informar la participación relativa de cada categoría con respecto del total.

17. Verdadero

El hecho de representar distribuciones de frecuencias de variable cuantitativa continua lleva a que las barras del histograma no dejen espacios entre ellas, es decir, van unidas.

18. Verdadero

La pirámide de población es un instrumento demográfico para representar la distribución por edad y sexo de una sociedad)

19. Verdadero

Las frecuencias de variable discreta pueden acumularse en los sentidos "X o menos" y "X o más". Si la variable es continua, se pueden acumular "al límite superior" o "al límite inferior".

20. Falso

Como todos los gráficos, la información solo pretende ser una buena aproximación y nunca valores exactos.

EJERCICIO 5.12

Pareo. Llene los paréntesis con el número que acompaña a cada definición de la izquierda, de forma que cada definición coincida con su concepto. Quedan dos paréntesis en blanco.

1.	Gráfico estadístico que se compone de un rectángulo suficientemente ancho cuya altura constituye el 100 %, luego se marcan las diferentes áreas para indicar los porcentajes que cada clase representa del total.	(2)	Barras comparativas
2.	Tipo especial de barras que contiene dos o más categorías que se pretende comparar.	(5)	Mapas estadísticos
3.	Representación de datos estadísticos en un plano cartesiano utilizando líneas que indican distancia, tendencia, pendiente o cambio de rumbo. Se utilizan para representar series cronológicas.	(7)	Pictograma
4.	Diagrama usado para representar distribuciones de frecuencia de variable continua. Se conforma de columnas sin separación, cuya base indica el ancho de clase y el área del rectángulo la frecuencia de clase.	(1)	Barra 100 %
5.	Forma de presentación de datos estadísticos sobre la base de un plano o un mapa, sobe el cual se sobrepone la información por medio de puntos, sombras, colores, rayas, dibujos y hasta objetos como alfileres, botones y otros dispositivos.	(9)	Gráfico de columnas
6.	Gráfico estadístico de frecuencias acumuladas que permite visualizar el número de observaciones que se encuentran por encima o por debajo de un valor determinado.	( )	Barras de dos direcciones
7.	Representación estadística que utiliza figuras alegóricas a lo que se está informando. Siguen las reglas de los gráficos de barras y se informa tanto por el número de ilustraciones como por el largo del conjunto de ilustraciones.	(10)	Diagrama de índice endémico
8.	Es una variedad de histograma que expresa gráficamente la conformación por edad y sexo de los habitantes de un país.	(4)	Histograma
9.	Gráfico adecuado para representar series cronológicas e información sobre variables cuantitativas discretas. La información se transmite por el largo de la barra y no por su ancho.	(8)	Pirámide de población
10.	Gráfico que permite dar seguimiento a la incidencia de una enfermedad para alertar la eventual llegada de brotes epidémicos.	( )	Polígono de frecuencias
		(6)	Ojivas
		(3)	Gráfico lineal

CAPÍTULO 6

EJERCICIO 6.1

Las afirmaciones siguientes están basadas en situaciones probabilísticas:

a) "La tasa de sobrevivencia de los pacientes que se someten a esta cirugía es 90%" porque ese porcentaje se construyó a partir de los registros médicos de muchos pacientes intervenidos previamente. En ellos el 10 % murió y el 90 % sobrevivió y superó el mal por el cual fueron intervenidos. Es a partir de esos datos que se establece un pronóstico para futuras operaciones, cada una con su riesgo y sus particularidades.

b) "Las primas que el instituto asegurador cobra por los seguros de vida son más altas para personas mayores de 55 años". Esto es así porque la esperanza de vida es mayor para los jóvenes. Está demostrado que a medida que se aumenta en edad, el riesgo de muerte es mayor, por lo que el instituto asegurador establece rangos de edad para establecer primas de aseguramiento.

c) "Fumar produce cáncer". Registros históricos demuestran una alta correlación entre fumar y desarrollar cáncer. El primer estudio documentado sobre la incidencia de cáncer de pulmón y su relación con el consumo del tabaco data de 1791 cuando el Dr. John Hill estudió el consumo de rapé (tabaco aspirado) y los tumores en la nariz. En 1795 Soemmerring relacionó el cáncer labial con el fumado. En 1928 Lombard y Doering asociaron el cáncer en general con el tabaco. En 1951 todos los trabajos previos fueron corroborados por el estudio retrospectivo de Doll y Hill, quienes en 22 años documentaron las causas de muerte de médicos fumadores y calcularon el crecimiento del riesgo relativo de cáncer a medida aumentaba el consumo de tabaco. Entre más recientes son los estudios, mayor uso se ha hecho de la teoría de probabilidades, por lo que esta afirmación está basada en situaciones probabilísticas.

d) "La esperanza de vida al nacer es de 77 años". La esperanza de vida al nacer se obtiene a partir de las tablas de mortalidad y sus probabilidades asociadas. La esperanza de vida es una estimación del promedio de años adicionales que una persona podría vivir, suponiendo que las tasas de mortalidad por edad específica para las otras edades no varían. Esta es una medida hipotética que se basa en las probabilidades de muerte actuales y variará cuando cambien las tasas a lo largo de la vida.

   La esperanza de vida al nacer es la más conocida y su cálculo requiere de las tasas de mortalidad de cada rango de edad y, en particular, de la tasa de mortalidad infantil, puesto que siempre será necesario conocer el número de nacidos del año.

La publicación de Floribel Méndez Fonseca y Olga Martha Araya Umaña, "Evolución de la Esperanza de Vida al Nacimiento en Costa Rica 1900 – 2000", publicado por el Instituto Nacional de Estadísticas y Censos INEC, y presentado en el año 2003, hace un análisis exhaustivo del comportamiento de la Esperanza de Vida al Nacer desde 1900 hasta el 2000. Entre sus hallazgos relevantes está:

"La esperanza de vida al nacimiento de Costa Rica en los últimos cien años ha experimentado un aumento significativo. Así, mientras una persona nacida a principios del siglo XX tenía una esperanza de vida al nacimiento de 35 años, una nacida a principios del siglo XXI tiene una de 77 años, lo que implica un aumento de 41 años" (p. 4).

EJERCICIO 6.2

a) Experimento y el espacio muestral.

i. Experimento: tomar muestras de sangre. Espacio muestral: diabéticos y no diabéticos.

ii. Experimento: pesar recién nacidos. Espacio muestral: pesos de 500 a 5000 gramos.

iii. Experimento: presentar un examen de estadística. Espacio muestral: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 y 10 (suponiendo que solamente se pueden obtener notas no fraccionales).

b) Definiciones:

i. Experimento: es un proceso controlado que permite realizar observaciones o mediciones.

ii. Espacio muestral: es el conjunto de todos los resultados posibles de un experimento.

iii. Evento: es cada resultado de un experimento.

iv. Complemento de un evento: lo conforman todos los demás resultados del Espacio Muestral.

v. Eventos mutuamente excluyentes: son los eventos que no tienen elementos en común, o que no pueden ocurrir al mismo tiempo. La existencia de uno, impide la del otro.

vi. Eventos colectivamente exhaustivos: se da cuando la suma de eventos agota todas las posibilidades del Espacio Muestral, y no es posible obtener otro resultado para un experimento dado.

c) Señale si los siguientes pares de eventos son mutuamente excluyentes.

i.	Viudo	Casado	Mutuamente excluyentes
ii.	Desempleado	Diabético	---
iii.	Esterilizado	Sucio	Mutuamente excluyentes
iv.	Ser alérgico	Tener gripe	---
v.	Menor de edad	Mayor de edad	Mutuamente excluyentes
vi.	Enfermo	Asegurado	---
vii.	Contratado	No contratado	Mutuamente excluyentes

d) Anote si los siguientes eventos son colectivamente exhaustivos.

i.	Nacional o extranjero	Colectivamente exhaustivos
ii.	Soltero o casado	---
iii.	Esterilizado o no esterilizado	Colectivamente exhaustivos
iv.	Vacunado o no vacunado	Colectivamente exhaustivos
v.	De Guanacaste o Limón	---
vi.	Ganar o perder un partido de fútbol	---
vii.	Tipos de sangre A o B	---

e) Determine el complemento de cada uno de los siguientes eventos.

i.	Empatar un partido de futbol	Ganar o perder
ii.	Tipos de sangre A y B	O y AB
iii.	No vacunados	Vacunados
iv.	Menos de tres días de incapacidad	Tres días o más de incapacidad
v.	Tres o menos días de incapacidad	Más de tres días de incapacidad
vi.	Guanacaste, Puntarenas, Cartago	San José, Heredia, Alajuela, Limón
vii.	Sacar 6 al tirar un dado	Sacar: 0, 1, 2, 3, 4, 5

EJERCICIO 6.3

a) ¿Cuál es la probabilidad de obtener el 6 en un lanzamiento de un dado? $\frac{1}{6}$ ¿Cuál es la probabilidad de su complemento? $1-\frac{1}{6}=\frac{5}{6}$

b) Una urna contiene 50 bolas plásticas: 20 amarillas, 15 rojas, 10 azules y 5 blancas. Si se saca una bola en forma aleatoria, calcule la probabilidad de que esta bola sea:

i. $P(A)=\frac{20}{50}=0,4$

ii. $P(R)=\frac{15}{50}=0,3$

iii. $P(R \text{ ó } A)=\frac{15}{50}+\frac{10}{50}=\frac{25}{50}=0,5$

iv. $P(\bar{A})=1-\frac{10}{50}=\frac{40}{50}=0,8$

v. $P(N)=\frac{0}{50}=0$

vi. $P(\bar{N})=1-\frac{0}{50}=1$

c) Registros del servicio de salas de emergencia de un hospital:

i. Supuestos:

- Que las condiciones actuales no han variado drásticamente de las existentes cuando se tomaron los datos.
- Que las probabilidades obtenidas son solamente estimaciones del valor real y nunca una garantía de tener la probabilidad exacta.
- El tamaño de la muestra utilizada para construir esa distribución de probabilidades es alta y los registros son fiables.

ii. La probabilidad de tener que atender a un paciente víctima de un accidente de tránsito es 20 %.

iii. La probabilidad de que el paciente por atender no sea víctima de un accidente de tránsito es 1 – 0,20 = 0,80

d) Exámenes para determinar el grupo sanguíneo:

i. Distribución de probabilidades de los diferentes grupos sanguíneos.

TIPO	FRECUENCIA	PROBABILIDAD
O	6 332	0,6013
A	3 027	0,2875
B	774	0,0735
AB	397	0,0377
	10 530	1,0000

ii. La probabilidad de que una persona escogida al azar sea del grupo B es 0,0735.

iii. La probabilidad de que no sea del grupo B es 1 – 0,0735 = 0,9265.

EJERCICIO 6.4

a) $P(R \text{ o } B) = P(R) + P(B) = \frac{15}{50} + \frac{5}{50} = \frac{20}{50}$

b) $P(Q \text{ o } K) = P(Q) + P(K) = \frac{4}{52} + \frac{4}{52} = \frac{8}{52}$

c) $P(S \text{ o } P) = P(S) + P(O) = \frac{3}{11} + \frac{1}{11} = \frac{4}{11}$

d) $P(H \text{ o } A) = P(H) + P(A) = \frac{3}{11} + \frac{3}{11} = \frac{6}{11}$

EJERCICIO 6.5

a) $P(M \text{ o } P) = P(M) + P(P) - P(M \text{ y } P)$

$$P(M \text{ o } P) = 0{,}25 + 0{,}30 - 0{,}10 = 0{,}45$$

b) $P(R \text{ o } J) = P(R) + P(J) - P(R \text{ y } J)$

$$P(R \text{ o } J) = 0{,}30 + 0{,}20 - 0{,}10 = 0{,}40$$

c) $P(M \text{ o } E) = P(M) + P(E) - P(M \text{ y } E)$

$$P(M \text{ o } E) = \frac{10}{10\,000} + \frac{8}{10\,000} - \frac{5}{10\,000} = \frac{13}{10\,000}$$

d) $P(S \text{ o } E) = P(S) + P(E) - P(S \text{ y } E)$

$$P(S \text{ o } E) = 0{,}70 + 0{,}40 - 0{,}30 = 0{,}80$$

EJERCICIO 6.6

a) La probabilidad de que no tenga ninguna de esas tarjetas es:

$$1 - 0{,}45 = 0{,}55$$

b) La probabilidad de que el cliente no compre ninguna de esas mercancías es:

$$1 - 0{,}40 = 0{,}6$$

c) La probabilidad que no tenga ninguno de esos 2 defectos es:

$$1 - \frac{13}{10\,000} = \frac{9987}{10\,000}$$

d) La probabilidad de que no sea llamado por ninguna de estas dos instituciones es:

$$1 - 0{,}80 = 0{,}20$$

EJERCICIO 6.7

a) Si se escoge aleatoriamente a 2 niños de esta edad, la probabilidad de que ambos sean asmáticos es:

$$0,269 \times 0,269 = 0,07$$

b) Si se escoge aleatoriamente a 3 niños de 6 a 7 años, la probabilidad de que ninguno sea asmático es:

$$0,731 \times 0,731 \times 0,731 = 0,39$$

c) La probabilidad de atender en una emergencia a dos pacientes del grupo sanguíneo B es:

$$P(B \text{ y } B) = P(B) \times P(B) = \frac{774}{10\,530} \times \frac{774}{10\,530}$$

d) La probabilidad de que al examinar 3 niños escogidos al azar, todos tengan caries o secuelas de la misma es:

$$0,844 \times 0,844 \times 0,844 = 0,60$$

EJERCICIO 6.8

a) Tabla de contingencia de probabilidades:

	Resultado de la biopsia prostática		
Resultado del tacto rectal	Cáncer	Benigno	Total
Agrandamiento	0,3286	0,0571	0,3857
Sin agrandamiento	0,1429	0,4714	0,6143
	0,4714	0,5286	1,0000

i. La probabilidad de que el Tacto Rectal detecte agrandamiento es:

$$P(A) = 0,3857$$

ii. La probabilidad de situación benigna luego de la biopsia es:

$$P(B) = 0,5286$$

iii. La probabilidad de que el Tacto Rectal resulte sin agrandamiento y la biopsia informe benigno es:

$$P(S \text{ y } B) = 0,4714$$

iv. La probabilidad de que el Tacto Rectal detecte agrandamiento o el diagnóstico final sea cáncer es:

$$P(A \text{ o } C) = P(A) + P(C) - P(A \text{ y } C)$$

$$P(A \text{ o } C) = 0,3857 + 0,4714 - 0,3286$$

$$P(A \text{ o } C) = 0,5286$$

b) Tabla de contingencia de probabilidades:

	Cultivo de muestras no pulmonares		
Baciloscopía	Enfermo	Sano	Total
Positiva	0,027	0,001	0,029
Negativa	0,059	0,913	0,971
	0,086	0,914	1,000

i. La probabilidad de que la baciloscopía resulte negativa es:

$$P(B^-) = 0,971$$

ii. La probabilidad de que el paciente esté enfermo de tuberculosis es:

$$P(E) = 0,086$$

iii. La probabilidad de que el paciente en la baciloscopía salga positivo y en los cultivos de muestras no pulmonares resulte sano es:

$$P(B^+ \text{y } S) = 0,001$$

iv. La probabilidad de que la baciloscopía salga positiva o en el diagnóstico final esté enfermo de tuberculosis es:

$$P(B^+ \text{o } E) = P(B^+) + P(E) - P(B^+ \text{y } E)$$

$$P(B^+ \text{o } E) = 0,029 + 0,086 - 0,027$$

$$P(B^+ \text{o } E) = 0,088$$

c) Tabla de contingencia de probabilidades:

	Peso del recién nacido		
Ultrasonografía	Bajo Peso	Sanos	Total
Retraso del Crecimiento Fetal	0,028	0,007	0,034
Sin Retraso en el Crecimiento Fetal	0,138	0,828	0,966
	0,166	0,834	1,000

i. La probabilidad de que la ultrasonografía indique retraso del crecimiento fetal es:

$$P(R) = 0,034$$

ii. La probabilidad de que el bebé efectivamente nazca con bajo peso es:

$$P(B) = 0,166$$

iii. La probabilidad de que la ultrasonografía indique sin retraso en el crecimiento fetal y al nacer el bebé resulte con bajo peso es:

$$P(S \text{ y } B) = 0,138$$

iv. La probabilidad de que la ultrasonografía indique sin retraso en el crecimiento fetal o al nacer el bebé resulte con bajo peso es:

$$P(S \text{ o } B) = P(S) + P(B) - P(S \text{ y } B)$$

$$P(S \text{ o } B) = 0,966 + 0,166 - 0,138$$

$$P(S \text{ o } B) = 0,994$$

EJERCICIO 6.9

a) $P(V_{2|N1}) = P(V) = \frac{1}{2}$

Son eventos independientes, esto es, lo que ocurrió en el primer parto no altera las probabilidades sobre el sexo del segundo hijo.

b) $P(C2|C1) = P(C) = 0{,}844$

Esto porque los eventos son independientes, la salud bucal del primer paciente no afecta a la probabilidad de caries del segundo.

c) $P(O_{2|AB1}) = P(O) = 0{,}6013$

Porque los eventos son independientes, el hecho que el primer paciente tenga sangre tipo AB no afecta en nada a la probabilidad sobre el tipo de sangre del segundo.

EJERCICIO 6.10

a) Tacto Rectal como prueba de tamizaje para detectar cáncer de próstata.

i. La probabilidad de que el Tacto Rectal informe agrandamiento dado que el paciente realmente tiene cáncer prostático es:

$$P(A|C) = \frac{P(A \text{ y } C)}{P(C)}$$

$$P(A|C) = \frac{0{,}3286}{0{,}4714} = 0{,}697$$

Esto indica la *sensibilidad* de la prueba de tamizaje y señala la proporción de pacientes con cáncer de próstata que se identifica por medio del tacto rectal.

ii. La probabilidad de que el Tacto Rectal indique "sin agrandamiento" dado que es benigno es:

$$P(S|B) = \frac{P(S \text{ y } B)}{P(B)}$$

$$P(S|B) = \frac{0{,}4714}{0{,}5286} = 0{,}892$$

Esto indica la *especificidad* de la prueba de tamizaje e informa la proporción de pacientes sanos que resulta sin agrandamiento en el Tacto Rectal.

iii. La probabilidad de que el paciente tenga cáncer de próstata dado que el Tacto Rectal informó agrandamiento es:

$$P(C|A) = \frac{P(C \text{ y } A)}{P(A)}$$

$$P(C|A) = \frac{0,3286}{0,3857} = 0,852$$

Esto cuantifica el *valor predictivo de una prueba positiva* (VPPP) e informa del porcentaje de pacientes que tienen cáncer de próstata y se le detectó agrandamiento en la prueba de tamizaje.

iv. La probabilidad de que el paciente esté sano (benigno) dado que el Tacto Rectal informó sin agrandamiento es:

$$P(B|S) = \frac{P(B \text{ y } S)}{P(S)}$$

$$P(B|S) = \frac{0,4714}{0,6143} = 0,767$$

Esto cuantifica el *valor predictivo de una prueba negativa* (VPPN) e informa del porcentaje de pacientes que está sano y fue diagnosticado sin agrandamiento en la prueba de tamizaje.

b) Baciloscopía como prueba de tamizaje para la detección de la tuberculosis.

i. La probabilidad de que la baciloscopía resulte positiva dado que el paciente está enfermo de tuberculosis es:

$$P(B+|E) = \frac{P(B^{+} \text{ y } E)}{P(E)}$$

$$P(B+|E) = \frac{0,0271}{0,0857} = 0,3167$$

Esto indica la *sensibilidad* de la prueba de tamizaje y señala la proporción de pacientes con tuberculosis que se identifica por medio de una baciloscopía positiva.

ii. La probabilidad de que la baciloscopía indique negativo dado que el paciente está sano es:

$$P(B^{-}|S) = \frac{P(B^{-} \text{ y } S)}{P(S)}$$

$$P(B^{-}|S) = \frac{0,9129}{0,9143} = 0,9984$$

Esto indica la *especificidad* de la baciloscopía e informa la proporción de pacientes sanos que resulta negativo en la prueba de tamizaje.

iii. La probabilidad de que el paciente esté enfermo de tuberculosis dado que la baciloscopía salió positiva es:

$$P(E|B^{+}) = \frac{P(E \text{ y } B^{+})}{P(B^{+})}$$

$$P(E|B^{+}) = \frac{0,0271}{0,0286} = 0,95$$

Esto cuantifica el *valor predictivo de una prueba positiva* (VPPP) e informa del porcentaje de pacientes que está enfermo de tuberculosis y fue diagnosticado positivo en la prueba de tamizaje.

iv. La probabilidad de que el paciente esté sano dado que la baciloscopía salió negativa es:

$$P(S|B^{-}) = \frac{P(S \text{ y } B^{-})}{P(B^{-})}$$

$$P(S|B^{-}) = \frac{0,9129}{0,9714} = 0,9397$$

Esto cuantifica el *valor predictivo de una prueba negativa* (VPPN) e informa del porcentaje de pacientes que está sano y fue diagnosticado negativo en la baciloscopía.

c) Mujeres con Síndrome Hipertensivo por Embarazo que enfrentan el riesgo de que el feto sea pequeño para la edad de gestación.

i. La probabilidad de que la ultrasonografía informe retraso en el crecimiento fetal dado que el bebé resultará con bajo peso es:

$$P(R|B) = \frac{P(R \text{ y } B)}{P(B)}$$

$$P(R|B) = \frac{0,0276}{0,1655} = 0,1667$$

Esto indica la *sensibilidad* de la ultrasonografía y señala la proporción de bebés con bajo peso al nacer que se identifica por medio de la prueba de tamizaje.

ii. La probabilidad de que la ultrasonografía indique sin retraso en el crecimiento fetal dado que el bebé está sano es:

$$\text{P (SR|S)} = \frac{P(SR\ y\ S)}{P(S)}$$

$$\text{P (SR|S)} = \frac{0,8276}{0,8345} = 0,9917$$

Esto indica la *especificidad* de la prueba que informa la proporción de bebés sanos que se predicen sin retraso en el crecimiento fetal en la prueba de tamizaje.

iii. La probabilidad de que el bebé nazca con bajo peso dado que la ultrasonografía informó retraso en el crecimiento fetal es:

$$\text{P (B|R)} = \frac{P(B\ y\ R)}{P(R)}$$

$$\text{P (B|R)} = \frac{0,0276}{0,0345} = 0,8$$

Esto cuantifica el *valor predictivo de una prueba positiva* (VPPP) e informa del porcentaje de bebés que nace con bajo peso y fue diagnosticado con retraso en el crecimiento fetal en la prueba de tamizaje.

iv. La probabilidad de que el bebé nazca sano dado que la ultrasonografía informó sin retraso en el crecimiento fetal es:

$$\text{P (S|SR)} = \frac{P(S\ y\ SR)}{P(SR)}$$

$$\text{P (S|SR)} = \frac{0,8276}{0,9655} = 0,857$$

Esto cuantifica el *valor predictivo de una prueba negativa* (VPPN) e informa del porcentaje de bebés que nace sano y fue diagnosticado sin retraso en el crecimiento fetal en la ultrasonografía.

EJERCICIO 6.11

Encierre en un círculo la opción que se ajusta al enunciado, Falsa (F) o Verdadera (V).

1. Falso

   Del rumbo de las actividades vitales no se puede tener plena certeza, aunque se pueda conocer la mayor o menor probabilidad de ocurrencia. Entender las probabilidades proporciona una valiosa herramienta para medir los riesgos y ayuda a una toma de decisiones documentada, pero no elimina completamente la incertidumbre.

2. Verdadero

   No existen probabilidades negativas ni mayores que 1. Cero informa de un evento imposible, en tanto que 1 indicaría la certeza total o del 100 % de que el evento ocurrirá.

3. Falso

   Los eventos mutuamente excluyentes no pueden tener elementos en común.

4. Verdadero

   Los eventos colectivamente exhaustivos incluyen a todos los resultados posibles, es decir, no dejan a ninguno por fuera.

5. Verdadero

   El espacio muestral lo conforman todos los eventos posibles, es decir, el 100 % de los eventos.

6. Verdadero

   Si $1-P(A)=P(\bar{A})$, entonces $P(A)+P(\bar{A})=1$. Esto es que un evento y su complemento no tienen elementos en común y juntos agotan el especio muestral.

7. Verdadero

   El enfoque clásico de probabilidades también se conoce como probabilidad a priori puesto que no es necesario hacer un experimento; basta con conocer el espacio muestral para calcular la probabilidad de ocurrencia de cualquier evento, si se sabe cuántos son favorables.

8. Falso

   Hay muchas situaciones en donde no se pueden obtener todos los eventos que conforman el espacio muestral. Vea el caso de todas las enfermedades que pueden afectar a los humanos en el mundo, su listado es grande y pueden surgir nuevas enfermedades que antes se desconocían.

9. Verdadero

   Establecer probabilidades desde el enfoque empírico, requiere de buenos registros históricos, fiables y con tamaños de muestra grande.

10. Verdadero

    Bajo este enfoque, la probabilidad es igual al número de veces en que ocurre un evento, dividido entre el número total de observaciones.

11. Falso

    El cambio drástico de condiciones afecta la asignación de probabilidades. Esta asignación se asume válida si las condiciones son básicamente las mismas o su alteración es pequeña.

12. Falso

    La determinación de probabilidades subjetivas se basa la experiencia, en estudios previos, en la evidencia del experto o en una creencia meditada.

13. Verdadero

    A nivel gerencial, estratégico y político se trabaja con situaciones únicas que requieren de una asignación de probabilidades subjetivas tomadas por expertos.

14. Verdadero

    La estimación subjetiva presenta sus desventajas propias, entre ellas está el hecho de que el porcentaje asignado es difícil de justificar si es cuestionado. Las asignaciones están afectadas por los prejuicios, por ideas preconcebidas, por temores conscientes o inconscientes y por los sentimientos acerca de lo que se quiere que suceda.

15. Verdadero

    Se da la siguiente situación: P (A o B) = P (A) + P (B)

16. Verdadero

    Para evitar la doble contabilización de eventos no excluyentes, se deberá restar la probabilidad conjunta, así: P (A o B) = P (A) + P (B) – P (A y B)

17. Verdadero

    Puesto que el complemento de un evento lo forman todos los eventos que no le son favorables, la probabilidad complementaria = $1 - P(E)$

18. Verdadero

    La probabilidad conjunta de eventos independientes es igual a la multiplicación de sus probabilidades marginales: P (A y B) = P (A) * P (B)

19. Verdadero

    Si los eventos son dependientes, la ocurrencia de uno afecta la del otro, entonces saber que un evento ha ocurrido puede ser útil para predecir la ocurrencia del otro. La probabilidad conjunta será: P (A y B) = P (A) * P (B|A)

20. Verdadero

    La probabilidad condicional es igual a la probabilidad conjunta dividida entre la probabilidad marginal de la condición, así: P (A|B) = $\dfrac{P(A\ y\ B)}{P(B)}$

EJERCICIO 6.12

Pareo. Llene los paréntesis con el número que acompaña a cada definición de la izquierda, de forma que cada definición coincida con su concepto. Quedan dos paréntesis en blanco.

1.	Son los eventos que no tienen elementos en común, o que no pueden ocurrir al mismo tiempo.	(3)	Probabilidad complementaria
2.	Informa de la probabilidad de ocurrencia de un evento, dado que otro se ha presentado previamente y lo afecta.	(5)	Evento seguro
3.	Es igual a 1 menos la probabilidad del evento de interés.	(7)	Probabilidad conjunta
4.	Es aquel evento cuya probabilidad de ocurrencia es cero.	(8)	Enfoque clásico
5.	Es aquel evento cuya probabilidad de ocurrencia es igual a uno.	(10)	Enfoque empírico
6.	Cuando la ocurrencia de un evento no afecta en nada a la ocurrencia de otro.	(1)	Mutuamente excluyentes
7.	Es la probabilidad de que dos eventos ocurran sucesivamente o al mismo tiempo.	( )	Dependencia
8.	Enfoque donde los eventos tienen una probabilidad de ocurrencia predecible teóricamente.	(9)	Enfoque subjetivo
9.	Enfoque en el cual la asignación de probabilidades está basada en la experiencia de las personas que efectúan la estimación.	( )	Experimento
10.	Enfoque que basa la asignación de probabilidades en registros históricos, es decir en las frecuencias relativas en que se ha presentado un evento.	(6)	Independencia
		(4)	Evento imposible
		(2)	Probabilidad condicional

HIPERVÍNCULOS

En estas direcciones electrónicas, usted encontrará datos oficiales, artículos científicos, estudios completos, debates y una amplia gama de material útil para respaldar sus avances de investigación y tareas, documentar estados de la cuestión, así como acceder a información actualizada.

- http://www.scielo.sa.cr/scielo.php?pid=0001-6002&script=sci_serial
- http://ajph)aphapublications.org/loi/ajph
- http://ccp.ucr.ac)cr/bvp/pdf/anuariocr
- https://apps.who.int/bulletin/digests/spanish/number3/no.3.html
- http://revistas.isciii.es/bes/index.php/bes/article/viewArticle/683/717
- http://unstats.un.org/unsd/demographic/products/dyb/dyb2.htm
- http://www.who.int/research/es/
- http://portal.ccss.sa.cr/portal/page/portal/Gerencia_Medica/Proyeccion_Servicios_Salud/Estadistica_en_Salud/Informe_Estadistico
- http://www.inec)go.cr/Web/Home/pagPrincipal.aspx
- http://www.ministeriodesalud)go.cr/
- http://revista.paho.org/
- http://www.linguee.es/ingles-espanol/traduccion/new+england+journal+of+medicine.html
- http://www.ministeriodesalud)go.cr/index.php/inicio-vigilancia-indicadores-salud-ms
- http://es.scribd)com/doc/56123631/CIE-10-COMPLETO

- http://www.coinsasalud)com.ar/~coinsasa/media/CIE_10_COMPLETO.xls
- http://www.inec)go.cr/A/MS/Encuestas/Encuesta%20Hogares%20de%20Prop%C3%B3sitos%20m%C3%BAltiples/Manuales/Manual%20de%20C%C3%B3digos%20geogr%C3%A1ficos%202007.pdf
- http://www.nortenlinea.com/inicio/wp-content/uploads/2011/05/02.-Boleta-Censo-2011-versión-para-imprimir.pdf
- http://www.ministeriodesalud)go.cr/index.php/inicio-vigilancia-indicadores-salud-ms
- http://www.inec)go.cr/Web/Home/pagPrincipal.aspx

GLOSARIO

CAPÍTULO 1

Analíticos, estudios: pretenden establecer conexiones entre variables, medir su correlación o nivel de asociación, comprobar su causalidad y plantear los escenarios futuros, las medidas preventivas y las consecuencias sobre la salud)

Atributos o variables cualitativas: características de la unidad de estudio que no pueden ser medidas pero sí pueden ser contadas u ordenadas.

Censo: una investigación sobre las características de toda la población.

Cuasi-experimentales, estudios: se dan cuando la variable en estudio no es manipulada por el investigador sino que se produce exógenamente en forma natural.

Descriptivos, estudios: buscan detallar las manifestaciones de un fenómeno a partir de su observación, es decir, lo describen.

Ecológicos, estudios: las unidades de estudio son las poblaciones, las comunidades o grupos específicos. Son útiles para establecer efectos que no se identifican en un paciente aislado y, también, son de gran ayuda para evaluar el impacto de programas, los resultados de intervenciones, la efectividad de las políticas de salud en una comunidad y permiten plantear nuevas hipótesis para futuros trabajos.

Entrevista: método de recolección de datos en el cual el informante debe contestar preguntas o llenar un cuestionario.

Error de muestreo: riesgo de que la muestra con que se trabaja no sea representativa y surge por la variabilidad del muestreo. Cuando la muestra es aleatoria, el error de muestreo se puede calcular, evaluar y hasta reducir por aumento del tamaño de la muestra o por muestras sucesivas en las cuales los errores se alternarán alrededor del parámetro.

Estadística: la ciencia que trata del desarrollo y la aplicación de las técnicas de recolección, organización, presentación, análisis e interpretación de datos numéricos obtenidos por observación o experimentación, para obtener conclusiones válidas y fundamentar la toma de decisiones.

Estadísticas: conjunto de datos numéricos que muestran las características de un fenómeno o situación.

Estadística aplicada: la estadística se hermana con otras disciplinas para generar instrumentos de trabajo, de esta forma, la demografía analiza poblaciones humanas y la bioestadística estudia a los seres vivos, además, se generan múltiples herramientas para la economía, la ingeniería, la sociología y la salud)

Estadística descriptiva: se especializa en la recolección, el resumen y la presentación de los datos, destaca sus características y relaciones para hacer la información más comprensible. No pretende hacer generalizaciones para la población.

Estadística inferencial: trabaja con muestras para hacer generalizaciones acerca de la población o, partiendo de experimentos, predice el comportamiento de las variables en situaciones similares.

Estimador: dato obtenido por medio de una muestra que sirve para inferir el valor poblacional o parámetro.

Experimentales, estudios: en ellos se da la intervención del investigador puesto que este introduce una alteración intencional para registrar los efectos y llegar hasta el control de las variables.

Fuentes primarias: datos, documentos e informes generados por la misma institución o sus dependencias.

Fuentes secundarias: información proporcionada por agencias que hacen acopio de estadísticas procesadas por otros, la hacen comparable y luego la publican. No se reducen a los anuarios y resúmenes estadísticos, puesto que también lo son los datos que se citan de revistas, tesis, informes de labores, consultorías y otras publicaciones.

Hipótesis: enunciado que propone una relación entre los datos, una explicación preliminar a un problema, una posible solución o un método de comprobación de las determinaciones entre variables.

Individuales, estudios o estudios de caso: estudian a un solo paciente. Es muy útil para estudiar enfermedades de baja incidencia, baja prevalencia o muy raras.

Longitudinales, estudios: son como una película, ya que necesitan del transcurso del tiempo para identificar posibles causas de un fenómeno, el comportamiento de una enfermedad y apreciar su tendencia.

Método deductivo: método científico que parte de leyes generales para explicar un fenómeno específico. Deducir es formular una hipótesis general para luego advertir sobre fenómenos individuales.

Método inductivo: método científico que basa sus resultados partiendo del estudio de muestras. En la inducción se obtienen conclusiones generales tras la observación y análisis de lo particular.

Muestra: subconjunto de la población que se examina para observar las características de interés.

Muestra aleatoria: subconjunto de la población cuya escogencia se basa en el azar y los elementos de la población tienen una probabilidad medible (calculable) de ser tomados en cuenta.

Muestra intencional: subconjunto de la población cuya escogencia parte del buen juicio, del conocimiento o la experiencia del investigador para identificar los elementos de la población que le pueden aportar mejores datos.

Muestra por conveniencia: subconjunto de la población cuya escogencia se basa en la comodidad para estudiar los elementos accesibles, más fáciles de encontrar, son voluntarios o se muestran más anuentes a colaborar con la investigación.

Observacionales, estudios: en ellos no se da interferencia del investigador, ya que este solamente se dedica a observar, cuantificar y medir las variables sin manipularlas ni controlarlas.

Observación: se basa en conteos o mediciones que realiza el mismo investigador en el laboratorio, en el trabajo de campo o por el examen físico directo del informante.

Población: se conforma por el total de elementos con una característica bajo estudio.

Población finita: tiene una cantidad de elementos que es posible enumerar, puede ser pequeña o muy grande pero tiene un número limitado de unidades.

Población infinita: cuenta con un número ilimitado de elementos o son parte de procesos de producción, reproducción o generación de nuevas unidades.

Prospectivos, estudios: comienzan en el presente con datos que se recolectarán en el futuro. Se diseñan hoy y parten de una causa sospechosa que estará bajo observación en el futuro, los datos se obtendrán a medida que aparezcan.

Registro: conteo de los hechos conforme se van presentando, por obligación de ley, como requisitos para tramitar indemnizaciones o por necesidades justificadas.

Retrospectivos, estudios: se realiza con datos del pasado. Generalmente, sus fuentes son secundarias y parten de una hipótesis en la cual se sospecha de una causa en el pasado que provoca una enfermedad en el presente.

Sesgos: son inclinaciones sistemáticas en un solo sentido, es decir, reflejan una tendencia por sobreestimación o subestimación insistente y constante.

Sesgo de medición: pueden ocurrir tanto en muestras como en censos, por lo que aumentar el tamaño de muestra no los corrige. En ellos hay una falla del instrumento de medición o los informantes alterna los datos consistentemente hacia el alza o hacia la baja.

Sesgo de selección: inclinación sistemática que surge en la incorrecta delimitación de la población de la cual se toma la muestra y se da en forma de: subregistro, omisiones, no respuesta o grupos no representativos de la población.

Teoría estadística: se basa en axiomas y el razonamiento lógico-matemático para dar fundamento a los métodos de análisis de la estadística.

Transversales, estudios: (*cross sectional*) son como una fotografía de una situación en un tiempo y un espacio determinado.

Variables: características de las unidades de estudio que pueden ser medidas en su longitud, peso, temperatura, precio o duración en el tiempo.

Variables continuas: pueden asumir infinito número de valores en un intervalo.

Variables dicotómicas: solamente asumen uno de dos valores: sí o no, falso o verdadero, defectuoso o aceptable, 1 o 0, contratado o no.

Variables discretas: solamente toman ciertos valores predeterminados en un intervalo, generalmente estos valores son enteros, aunque pueden usarse valores fraccionarios.

CAPÍTULO 2

Ancho de clase o amplitud de clase: Diferencia entre el límite real superior y el límite real inferior.

Clases: Agrupaciones de los diferentes valores que asume una variable.

Clases colectivamente exhaustivas: Cuando las categorías asignadas agotan el espacio muestral. Esto es que cualquier resultado posible siempre encontrará una categoría en la cual registrarse, es decir, que todas las clases juntas representan el 100 % y no hay omisiones.

Confiabilidad: Característica de un buen instrumento de recolección de datos; consiste en que si se aplica en forma repetida al mismo informante en condiciones similares, siempre producirá datos iguales o equivalentes.

Cuadro estadístico: Instrumento de presentación de datos, donde estos se ordenan en filas y columnas de manera que puedan ser comprendidos sin mucha explicación textual o gráfica.

Cuadros generales: Llamados también de referencia o primarios, presentan los datos con el mayor detalle posible, pero no pretenden destacar informaciones específicas ni demostrar aspectos concretos.

Cuadros de resumen: Son derivados de los cuadros generales y tienen como objetivo destacar aspectos específicos, facilitar el análisis y responder a preguntas de investigación concretas.

Estado de la cuestión: Es el cuerpo de publicaciones que otros investigadores han escrito sobre mi objeto de estudio o tema-problema.

Eufemismo: Palabras sustitutivas de otra considerada ofensiva, de mal gusto o tabú.

Frecuencia: Número de ocurrencias de un determinado evento. Conteo de las observaciones pertenecientes a una categoría.

Frecuencia absoluta: Contiene el número simple de elementos encontrados de cada clase.

Frecuencia acumulada: Es la suma de las frecuencias absolutas o relativas y puede hacerse en forma descendente (X o menos) o ascendente (X o más).

Frecuencia relativa: Se obtienen dividiendo cada frecuencia absoluta entre el total de las observaciones e informan el porcentaje que cada clase representa del total.

Gráficos circulares: Representación estadística circular que informa de la participación porcentual de cada una de las categorías.

Gráficos estadísticos: Instrumentos que comunican la información utilizando distancias, áreas y pendientes o volúmenes. Estas magnitudes representan a los datos numéricos con relativa facilidad, además atraen la atención de los lectores.

Gráficos de barras: La información se transmite por el largo de las barras utilizadas y se usan para representar series cualitativas y geográficas. Pueden representarse barras simples, compuestas y comparativas.

Gráficos de columnas: Al igual que los gráficos de barras, la información se transmite por el largo de las barras utilizadas, pero se usan para representar series de tiempo o variables cuantitativas discretas o categorizadas por intervalos. Igualmente se pueden confeccionar en forma compuesta o comparativa.

Gráficos lineales: Instrumentos estadísticos que transmiten la información por medio de líneas cuya distancia, tendencia, cambio de rumbo o inclinación son las claves para su interpretación. Se usan para representar series de tiempo.

Instrumentos de recolección: Son herramientas útiles para la recolección de los datos. Se dividen en instrumentos directos e indirectos.

Instrumentos de recolección directos: Obtienen el dato en la fuente, entre ellos están los experimentos, el registro y la observación, tanto cuantitativa como cualitativa.

Instrumentos de recolección indirectos: Permiten recolectar los datos utilizando intermediarios, por ejemplo, a través de entrevistas, encuestas, grupos focales, análisis de contenido o en fuentes secundarias.

Mutuamente excluyente: Situación en la cual dos datos no pueden pertenecer a la misma clase o que la ocurrencia de un evento impide que se dé el otro al mismo tiempo.

Objetivos: Descripciones de lo que se espera del trabajo de investigación. Se plantean en forma afirmativa y con verbos en infinitivo.

Preguntas abiertas: Permiten contestar libremente, se utilizan cuando la respuesta no es previsible o cuando hay muchas categorías.

Preguntas bidimensionales: Solicitan una sola respuesta en un ítem que, en realidad, contiene dos preguntas.

Preguntas de calentamiento: Son poco relevantes para el problema de investigación, pero cumplen un rol introductorio y ayudan a "romper el hielo" con el entrevistado facilitando la comunicación.

Preguntas cerradas: Las alternativas de respuesta se han limitado previamente; estas pueden ser dicotómicas: falso o verdadero, positivo o negativo, de acuerdo o en desacuerdo, sí o no (como en el test anterior), o pueden presentar varias alternativas

Preguntas de comportamiento: Buscan conocer si se ha realizado tal o cual acción, o qué se hizo frente a tal situación.

Preguntas filtro: Ayudan a clasificar a los informantes y reducen el tiempo de aplicación de un cuestionario al eliminar un grupo de ítems que no son aplicables en el caso concreto.

Preguntas de hechos: Se plantean sobre hechos concretos, de respuesta precisa.

Preguntas de intención: Indagan cómo se actuaría si se presentara tal situación, es decir, preguntan sobre algo hipotético o imaginado; el entrevistado no dice cómo se comportó sino cómo se comportaría frente a tal situación

Pregunta de investigación: Cuestionamiento general y abstracto del tema por investigar. Esta deberá traducirse en varias preguntas específicas para aplicarlas a los informantes.

Preguntas de opinión: Piden al informante que diga lo que piensa, opina o cree, no necesariamente una actuación específica, tampoco indaga qué conoce ni cómo se comporta ni qué haría en tal situación

Preguntas semi-abiertas: Establecen un listado de opciones cerradas, pero dejan una alternativa para ser llenada libremente por el entrevistado.

Preguntas tipo test: Permiten establecer un puntaje o índice que revela de lo que para el informante no es evidente o no se atreve a confesar directamente.

Punto medio: Es el valor representativo de la clase y, para ello, se calcula el valor central como un promedio de los límites reales de la clase.

Validez: Grado de confianza en que un instrumento de recolección de datos mide lo que pretende medir.

CAPÍTULO 3

Constante de amplificación: Valor arbitrario por el cual se multiplican los indicadores para hacerlos más comprensibles para el gran público. Usualmente es múltiplo de 10 (los más frecuentes son 100, 1 000 y 100,000).

Especificidad: Medida que indica la proporción de pacientes sanos que resulta negativo en la prueba de tamizaje.

Factor de riesgo: Condición o rasgo detectable (en un individuo, un grupo, una familia o una comunidad) que se asocia con el aumento en la probabilidad de padecer una enfermedad, desarrollar un proceso mórbido, sufrir un accidente o enfrentar la muerte prematuramente.

Fracción atribuible (FA): También se conoce como fracción etiológica pues estima el grado de influencia que tiene un tratamiento como causa de la curación de los pacientes y que puede ser atribuible al hecho de seguirlo.

Indicadores de salud: Cifras que se construyen a partir del registro de los hechos vitales y su posterior procesamiento para informar sobre la situación sanitaria de una sociedad y el desempeño del sistema de atención.

Morbilidad: Ocurrencia de enfermedades en una población, su registro proporciona indicadores negativos del estado de salud de la colectividad)

Oportunidad relativa (OR: Odds Ratio): Razón del éxito terapéutico entre dos tratamientos, esto es, la oportunidad de sobrevivir utilizando el tratamiento 1 entre la oportunidad de sobrevivir utilizando el tratamiento 2.

Proporción: Número relativo que informa del resultado de dividir un subgrupo entre toda la población. Las proporciones reflejan la importancia relativa de cada clase dentro de la población.

Razón: Número relativo que resulta de dividir un subgrupo de la población entre otro subgrupo o entre la población total.

Razón de estandarización: Se calcula dividiendo las defunciones ocurridas (DO) entre las defunciones esperadas (DE. de cada población, esto es, de la división de eventos reales entre eventos hipotéticos.

Riesgo atribuible: Diferencia entre las tasas de sobrevivencia de dos tratamientos.

Riesgo relativo: Razón que surge de comparar dos tasas. Se calcula dividiendo la tasa de sobrevivencia con el tratamiento 1, entre la tasa de sobrevivencia con el tratamiento 2.

Sensibilidad: Medida que indica la proporción de pacientes con un padecimiento que se identifica por medio de una prueba de tamizaje.

Tamiz: Cedazo muy tupido que sirve para colar.

Tamizaje, pruebas: Prueba no invasiva en una población saludable, para identificar a aquellos individuos que tienen alguna patología, pero que todavía no presentan síntomas.

Tasa: Número relativo que indica la frecuencia relativa con que un evento ocurre en una población y siempre se calcula con referencia a un periodo de análisis, convencionalmente de un año. Las tasas informan el riesgo (de nacer, enfermarse, morir, etc)) de que el fenómeno suceda en ese periodo y pueden ser brutas, específicas o ajustadas.

Tasas ajustadas: Indicadores estandarizados que informan "como si" las estructuras de población fueran semejantes. Se usan solamente para propósitos comparativos, pues *no* son tasas reales de la población o del periodo, sino que son tasas hipotéticas, es decir tasas de "lo que sería" si la población fuera como la población tipo.

Tasas brutas: Se obtienen de dividir el número de elementos que, en el periodo, experimentó las consecuencias del riesgo entre la población total.

Tasas específicas: Se calculan para el subgrupo que generalmente contiene la población expuesta a sufrir el evento.

Tasa de casos: Tipo particular de tasa de incidencia, se usa cuando es posible que algunas personas contraigan la misma enfermedad más de una vez en el periodo de estudio.

Tasa de incidencia: Número de *casos nuevos* de una enfermedad durante un periodo por cada 100 000 habitantes en condiciones de sufrir el riesgo.

Tasa de letalidad: Porcentaje de personas que contrae una enfermedad X y muere a causa de ella en un periodo determinado, es decir, mide la probabilidad de morir por una enfermedad una vez que se ha contraído.

Tasa de morbilidad por causa: Informa del número de casos de una enfermedad X, dividido entre la población total a mitad del periodo.

Tasa de prevalencia: Informa del número de pacientes que *padece* de una determinada enfermedad, en un periodo de tiempo y zona geográfica, por cada 100 000 habitantes.

Valor Predictivo de una Prueba Negativa (VPPN): Número de pacientes que está sano y fue diagnosticado sano, dividido entre todos los que resultaron negativos.

Valor Predictivo de una Prueba Positiva (VPPP): Número de pacientes que está realmente enfermo y que fue diagnosticado así en una prueba de tamizaje, dividido entre todos los que dieron positivo.

CAPÍTULO 4

Asimetría negativa: Existe cuando una cola de la distribución se extiende hacia la izquierda por la presencia de valores extremos. Cuando hay asimetría negativa, el promedio es menor que la moda ($\bar{x} < \text{Mo}$).

Asimetría positiva: Existe cuando una cola de la distribución se extiende hacia la derecha por la presencia de valores extremos. Cuando hay asimetría positiva, el promedio es mayor que la moda ($\bar{x} > \text{Mo}$).

Coeficiente de variación: Explica la importancia de la desviación estándar de una serie de datos en relación con su promedio.

Cuantiles: Los valores de una variable que dividen a un conjunto ordenado de datos en fracciones de igual magnitud)

Cuartiles: Los valores de una variable que dividen a un conjunto ordenado de datos en cuatro partes iguales, donde cada una contiene el 25 % del conjunto de datos.

Deciles: Los valores que dividen un conjunto ordenado de datos en diez partes iguales.

Desviación estándar: La raíz cuadrada de la media aritmética de las desviaciones respecto del promedio elevadas al cuadrado.

Desviación media: La media aritmética de los valores absolutos de las desviaciones respecto del promedio.

Mediana: El valor central de una serie de datos ordenados, el 50 % de los datos está por debajo de ese valor y el 50 % por encima.

Medidas de dispersión: Informan la variabilidad de un conjunto de datos alrededor de un valor determinado. Entre las medidas de dispersión más utilizadas están: el rango, la desviación media, la desviación estándar, la variancia y el coeficiente de variabilidad)

Medidas de forma: Indicadores de la simetría o asimetría de una distribución de frecuencias, así como del grado de apuntamiento o concentración de datos.

Medidas de posición o de tendencia central: Medidas descriptivas de la tendencia o concentración de los datos de una distribución. Entre ellas están: la moda, el promedio, la mediana y los percentiles.

Moda: La cantidad o cualidad que más se repite en un conjunto de datos, técnicamente se define como el valor más frecuente.

Percentiles: Los valores que dividen un conjunto ordenado de datos en cien partes iguales.

Promedio o media aritmética: El valor que se obtiene de la sumatoria de todos los datos dividida entre el número de observaciones.

Punto medio de una clase: es el valor que representa a una clase, generalmente se obtiene de promediar la suma de los límites superior e inferior.

Quintiles: Los valores de una variable que dividen a un conjunto ordenado de datos en cinco partes iguales. Corresponden a dos deciles o a veinte percentiles.

Rango: Recorrido o amplitud, informa del intervalo contenido entre el valor mayor y el menor de los datos. Posibilita una idea general de la dispersión de los datos de una distribución.

Simetría: Armonía en la posición de los datos con relación a un centro o eje. Proporción adecuada de las partes de un todo. Cuando una distribución es simétrica, presenta el mismo número de valores a la derecha y a la izquierda del promedio.

Variancia: La media aritmética del cuadrado de las desviaciones respecto del promedio de una distribución.

CAPÍTULO 5

Barra 100 %: Gráfico estadístico que se compone de un rectángulo suficientemente ancho cuya altura constituye el 100%, luego se marcan las diferentes áreas para indicar los porcentajes que cada clase representa del total.

Diagrama de dispersión: Es un gráfico que permite visualizar las posibles relaciones entre dos o más variables. A cada eje del plano cartesiano se le asigna una variable y las observaciones se representan como puntos del plano conteniendo su información particular asociada a las variables.

Diagrama de Pareto: Es una variación del histograma y se utiliza para clasificar los eventos (enfermedades, problemas, defectos) comenzando por la de mayor frecuencia y, en forma descendente, hasta los eventos menos significativos.

Gráfico de barras: Representación estadística que informa los datos por medio de la longitud de las barras. Convencionalmente, las barras horizontales se usan para representar series cualitativas y series geográficas que deben ordenarse por magnitud).

Gráfico de barras comparativas: Tipo especial de barras que contiene dos o más categorías que se pretenden comparar.

Gráfico de barras compuestas: Tipo particular de gráfico de barras que se origina en tablas de doble entrada, lo cual genera dos o más componentes en una misma barra y que juntos informan los datos por el largo total.

Gráfico circular: Se usa para representar la participación porcentual de las categorías con respecto del total. Es el área de cada una de las secciones la que indica la participación relativa de la categoría.

Gráfico de columnas: Adecuado para representar series cronológicas e información sobre variables cuantitativas discretas. La información se transmite por el largo de la barra y no por su ancho.

Gráfico estadístico: Representación visual de la información estadística por medio de líneas, tendencias, pendientes, áreas y volúmenes.

Gráfico lineal: Representación de datos estadísticos en un plano cartesiano utilizando líneas que indican distancia, tendencia, pendiente o cambio de rumbo. Se utilizan para representar series cronológicas.

Histograma: Diagrama usado para representar distribuciones de frecuencia de variable continua. Se conforma de columnas sin separación, cuya base indica el ancho de clase y el área del rectángulo la frecuencia de clase.

Mapas estadísticos: Forma de presentación de datos estadísticos sobre la base de un plano o un mapa, sobe el cual se sobrepone la información por medio de puntos, sombras, colores, rayas, dibujos y hasta objetos: alfileres, botones y otros dispositivos.

Ojivas: Gráfico estadístico de frecuencias acumuladas que permite visualizar el número de observaciones que se encuentran por encima o por debajo de un valor determinado.

Pictograma: Representación estadística que utiliza figuras alegóricas a lo que se está informando. Siguen las reglas de los gráficos de barras y se informa tanto por el número de ilustraciones como por el largo del conjunto de ilustraciones.

Pirámide de población: Variedad de histograma que expresa gráficamente la conformación por edad y sexo de los habitantes de un país.

Polígono de frecuencias: Gráfico que sirve para representar variables continuas uniendo los puntos medios de cada clase con líneas rectas y cerrando en los extremos con frecuencia cero.

CAPÍTULO 6

Colectivamente exhaustivos: Se da cuando la suma de eventos agota todas las posibilidades del espacio muestral y no es posible obtener otro resultado para un experimento dado.

Complemento: Lo que falta para completar el 100 %. El complemento lo conforman todos los demás resultados del espacio muestral que no son favorables al evento de interés.

Dependencia: Cuando la ocurrencia de un evento afecta a favor o en contra de la ocurrencia de otro. Así, conocer la ocurrencia de uno es muy útil para mejor predecir la del otro.

Enfoque clásico: Enfoque de probabilidades que considera que los eventos tienen una probabilidad de ocurrencia medible que se obtiene al dividir un evento favorable entre todos los eventos del espacio muestral:

$$P(A) = \frac{Evento\ favorable}{Número\ total\ de\ resultados\ posibles}$$

Enfoque empírico: Enfoque de probabilidades que basa la asignación de probabilidades en registros históricos, es decir, en las frecuencias relativas en que se ha presentado un evento. La probabilidad es igual al número de veces que ocurre el evento A, dividido entre el número total de observaciones.

$$P(A) = \frac{Número\ de\ veces\ que\ ocurre\ A}{Número\ total\ de\ observaciones}$$

Enfoque subjetivo: Enfoque en el cual la asignación de probabilidades está basada en la experiencia de las personas que efectúan la estimación. Así, un individuo puede basarse en la evidencia que tenga disponible, en experiencias previas o en una creencia meditada.

Espacio muestral: Conjunto de todos los eventos posibles en un experimento.

Evento: Los posibles resultados de un experimento.

Evento imposible: Aquel cuya probabilidad de ocurrencia es cero. Se está completamente seguro de que nunca ocurrirá.

Evento seguro: Aquel cuya probabilidad de ocurrencia es igual a uno. Existe plena certeza de su ocurrencia.

Experimento: Proceso controlado que permite realizar observaciones o mediciones.

Independencia: Cuando la ocurrencia de un evento no afecta en nada a la ocurrencia de otro. Cuando dos eventos son independientes, entonces la probabilidad de que ambos ocurran es igual a la multiplicación de sus probabilidades individuales o marginales.

Mutuamente excluyentes: Eventos que no tienen elementos en común o que no pueden ocurrir al mismo tiempo. Uno y solo uno de ellos puede tener lugar a la vez, es decir, la ocurrencia de uno evita la del otro.

Probabilidad condicional: Informa de la probabilidad de ocurrencia de un evento, dado que otro se ha presentado previamente y lo afecta o condiciona.

$$P\,(A|B) = \frac{P(A\ y\ B)}{P(B)}$$

Probabilidad complementaria: Se calcula con todos los eventos que no le son favorables, de allí que la probabilidad complementaria sea igual a 1 menos la probabilidad del evento de interés.

$$\text{Probabilidad complementaria} = 1 - P(E)$$

Probabilidad conjunta: Probabilidad de que dos eventos ocurran sucesivamente o al mismo tiempo.

$$P\,(A \text{ y } B) = P\,(A) * P\,(B)$$

Si los eventos son independientes

$$P\,(A \text{ y } B) = P\,(A) * P\,(B|A)$$

Si hay dependencia

Probabilidad marginal: Igual a la probabilidad de ocurrencia de eventos individuales.

$$P(A)$$

REFERENCIAS BIBLIOGRÁFICAS

Aedo, S., Pavlov, S. y Clavero F) (2010). "Riesgo relativo y Odds ratio ¿Qué son y cómo se interpretan?" Santiago de Chile: *Revista de Obstetricia y Ginecología* (5), pp. 51 – 54.

Agra, Y., De la Cueva, P. Guasch, E. y Santamaría, R. (2009). *Maternidad hospitalaria.* Madrid: Ministerio de Sanidad y Política Social.

Anaya, L. (2011). *Indicadores de salud*) México: INDESOL.

Alba, E., Fernández, A., Manchado, C) y Tenorio, S. (2010). *Estudio encuesta: Métodos de investigación.* San José: UAM.

Baltodano, VJ., Burgos, E. y Rodríguez, R. (2012). *Ejercicios y prácticas de probabilidad y estadística.* Heredia: Editorial EUNA.

Burgos, E. (2006). *Introducción a la estadística inferencial.* Heredia: EIESTRA.

Burgos, E. (2013). *Demografía.* San José: Editorial UNED)

Burgos, E.; Dopslaf, M. y Rojas, R. (2007). *Aplicaciones informáticas de la estadística inferencial.* Heredia: EIESTRA.

Burgos, E.; Dopslaf, M. y Rojas, R. (2008). *Aplicaciones en Excel para la estadística descriptiva.* Heredia: EIESTRA.

Caja Costarricense del Seguro Social. (2003). *Cambios en la mortalidad y morbilidad por edad y sexo, Costa Rica 1987, 1992, 1997 y 2002.* Recuperado de http://www.ccss.sa.cr/html/organizacion/gestion/gerencias/medica/germed/dis/diess/intcam.htm

Caja Costarricense del Seguro Social. (2004). *Egresos hospitalarios por parto y aborto según estado conyugal y edad para el año 2003.* San José: Departamento de Estadísticas de Servicios de Salud, CCSS.

Centro Centroamericano de Población, Instituto Nacional de Estadística y Censos. (2002). *Costa Rica: estimaciones y proyecciones de población 1970-2100 actualizadas al año 2000 y evaluación del Censo 2000 y otras fuentes de información: informe metodológico.* San José: CCP.

Cuero, O. (2012). *Prueba de tamizaje.* Diaporama recuperado el 18 de Julio de 2012 de < http://www.slideshare.net/malebranche18/prueba-de-tamizaje>.

Chaves, S., Moya, R., Peralta, S., Quirós, I., Sancho, X. y Umaña, F) (2008). *Guía de atención integral a las mujeres, niños y niñas en el período prenatal, parto y posparto.* San José: CCSS.

Daniel, W. (1988). *Bioestadística.* México: Editorial Limusa.

Díaz, C) (2010). *Pruebas diagnósticas. Medidas de asociación.* Diaporama recuperado de <http://www.slideshare.net/cristiandiazv/pruebas-diagnsticas-11196171>.

Díaz G), Gran, M., Labrada, L. Torres, R. (2010). *Indicadores básicos para el análisis del estado de salud de la población.* La Habana: Ministerio de Salud Pública.

Durán, F) (2001). *Equidad y acceso a los servicios de salud en Costa Rica.* San José: CCSS.

Fernández, P., Vila, A. y Montero, C) (2002). *Determinación de factores de riesgo.* La Coruña: Complejo Hospitalario Juan Canalejo.

Galván, J. (2009). *Pruebas de tamizaje.* México: UNAM, Red de comunicación e Integración Biomédica.

Gallardo, H) (2005). *Elementos de investigación académica.* San José: Editorial UNED)

García, JJ. (1998). "Medición del riesgo en epidemiología". México: *Revista Mexicana de Pediatría.* (65) 2, pp. 76-83.

Gómez, M. (1998). *Elementos de estadística descriptiva.* San José: Editorial UNED)

Hernández, P. (2012). *Indicadores hospitalarios.* Lima: IREN.

Instituto Nacional de Estadísticas y Censos (2009). *Estadísticas vitales 2008.* San José: INEC)

IREN (2012). *Indicadores hospitalarios 2011.* Lima: Instituto Regional de Enfermedades Neoplásicas.

Jiménez, JG) (2004). *Obesidad: problema de salud pública.* San José: Acta Médica Costarricense, v.46 Suplemento 1.

Lepe, L. (2004). Utilidad de lercanidipino en pacientes con hipertensión arterial esencial. *Revista Mexicana de Cardiología,* 15, 75-82.

Lind, D), Marchal, W. y Wathen. S. (2008). *Estadística aplicada a los negocios y la economía.* México: Editorial McGraw-Hill.

Mann, J., Tarantola, D) y Netter, T. (1992). *Aids in the world)* Massachusetts: Harvard University Press.

Meléndez, M. (2009). *La salud corre por sus venas*. San José: Perfil 596.

Méndez, C), Espinoza, E., Lara, W. y Elas, M. (2006). *Reforma al sistema de salud de El Salvador*. San Salvador: Rectoría Universidad de El Salvador.

Méndez, F) y Araya, O. (2003). *Evolución de la Esperanza de Vida al Nacimiento en Costa Rica 1900 – 2000*. San José: Instituto Nacional de Estadísticas y Censos INEC)

Ministerio de Salud, República de Costa Rica. (2011). *Indicadores Básicos de Salud 2010*. San José: Dirección de Vigilancia de la Salud)

Ministerio de Salud, República de Costa Rica. (2012). *Informe de Dengue y Malaria a la semana 20 terminada el 19 de mayo 2012*. San José: Dirección de Vigilancia de la Salud)

Molinero, L. (2001). *Odds ratio, riesgo relativo y número necesario a tratar*. Madrid: ASEH)

Montero, O., Ulate, J., Rodríguez, A., Méndez, C), Monge, L. y Elías, A. (2007). *Estudio de prevalencia de caries dental en niños y niñas escolares*. San José: Colegio de Dentistas. Recuperado de <www.colegiodentistas.org>

Moral, M. S.; Ortega, J.; López, M. J. y Pellicer, P. (2003). *Perfil de riesgo de morbilidad psíquica en cuidadores de pacientes interesados en su domicilio*. Servicio Andaluz de Salud)

Moya, L. (2009). *Introducción a la estadística de la salud*) San José: Editorial UCR.

Organización Panamericana de la Salud) (2006). *Situación de la salud en Costa Rica: Indicadores básicos*. San José: OPS.

Organización Panamericana de la Salud) (2002). "La estandarización: Un método epidemiológico clásico para la comparación de tasas". *Boletín Epidemiológico*, (23) 3. pp. 9–12.

Pérez, C) (2002). *Estadística aplicada a través de Excel*. Madrid: Editorial Prentice Hall.

Rada, G) (2007). *Ajuste de tasas*. Santiago: Universidad Católica de Chile.

RITEVE. (2012). *El impacto de una década con RITEVE*. San José: Boletín Informativo.

Rojo, C) (2004). *Indicadores de salud de una comunidad*) Sevilla: Departamento de Medicina Preventiva y Salud Pública.

Roselló, M., Arauz, A., Padilla, G) y Morice, A. (2004). *Prevalencia de diabetes mellitus auto-reportada en Costa Rica, 1998*. San José: Acta Médica Costarricense v.46 n.4.

Ross, S. (1996). *Introductory Statistics*. Los Ángeles: Editorial McGraw Hill.

Sánchez, M. (2005). "Indicadores de gestión hospitalaria". México: *Revista del Instituto Nacional de Enfermedades Respiratorias*, (18) 2, pp. 132-141.

Sawyer, RG), Raymond, DP., Pelletier, SJ., Crabtree, TG), Gleason, TG) y Pruett TL. (2001). *Implications o 2457 consecutive infections entering year 2000.* University of Virginia: Annual Surgery.

Schapiro, M. (1972). Tipo de sangre, Factor Rh y determinaciones de genotipos en un grupo heterogéneo de la población nativa de Honduras. *Revista Médica de Costa Rica*, 438, 375-394.

Soto-Martínez, M. y Soto-Quirós, M. (2004). Epidemiología del asma en Costa Rica. *Revista Médica del Hospital Nacional de Niños*, 39 (1).

Spiegel, M. Stephens, L. (2009). *Estadística.* México: Editorial McGraw-Hill.

Stevenson, W. (1996). *Estadística para administración y economía.* Nueva York: Editorial Harla.

UNICEF) (2001). *Bajo peso al nacer.* Boletín informativo.

Varela, A. y Brenes, W. (1998). *Efectividad de algunas medidas preventivas y educativas en la reducción de la caries dental.* San José: INISA, UCR.

Wilson, J. y Jurger, G) (1968). *The principles and practice of screening for disease.* Washington: Public Health Papers, WHO, N. 34.

ELIO BURGOS GÓMEZ

Economista, administrador y lingüista. Catedrático de la UNA y de la UNED. También ha sido profesor visitante en la Universidad de Ciencias Aplicadas de Münster, Alemania, la UNAN-León y la Universidad para la Paz de la ONU. Sus publicaciones con EUNA y EUNED abordan temas de microeconomía, demografía, muestreo y estadística tanto descriptiva como inferencial.

Comprado el 28/09/2018

Estadística descriptiva. Aplicada a las ciencias de la salud.
se terminó de imprimir en el mes de abril del 2016
en los Talleres Gráficos de la Editorial EUNED.
Su edición consta de 1000 ejemplares
impresos en papel bond 75 gramos
con forro de cartulina barnizable.

Estuvo al cuidado
de la Dirección de Producción de Materiales Didácticos.

EUNED

La estadística es una poderosa herramienta que ha coadyuvado a importantes avances en epidemiología, demografía, bioestadística, muestreo, diseño de experimentos, pruebas de hipótesis y registros médicos.

La estadística permite potenciar los avances en las ciencias de la salud para prevenir la enfermedad, enfrentar sus causas, combatir formas de contagio, generar pronósticos y patrones de ocurrencia, además de brindar datos útiles a fin de establecer políticas generales de salud, sistemas de vigilancia y campañas de divulgación que lleven a mejorar la salud de la comunidad.

El libro que usted tiene en sus manos aborda la parte descriptiva de la estadística: la captura, la presentación y el análisis de datos numéricos; en él encontrará procedimientos que le permitirán elaborar informes textuales, tabulares y gráficos, también le mostrará técnicas para establecer las medidas de tendencia central, variabilidad y forma, métodos para calcular números relativos como las razones, proporciones, tasas e índices y maneras de estimar probabilidades, riesgos y oportunidades relativas.

ISBN 978-9968-48-195-0