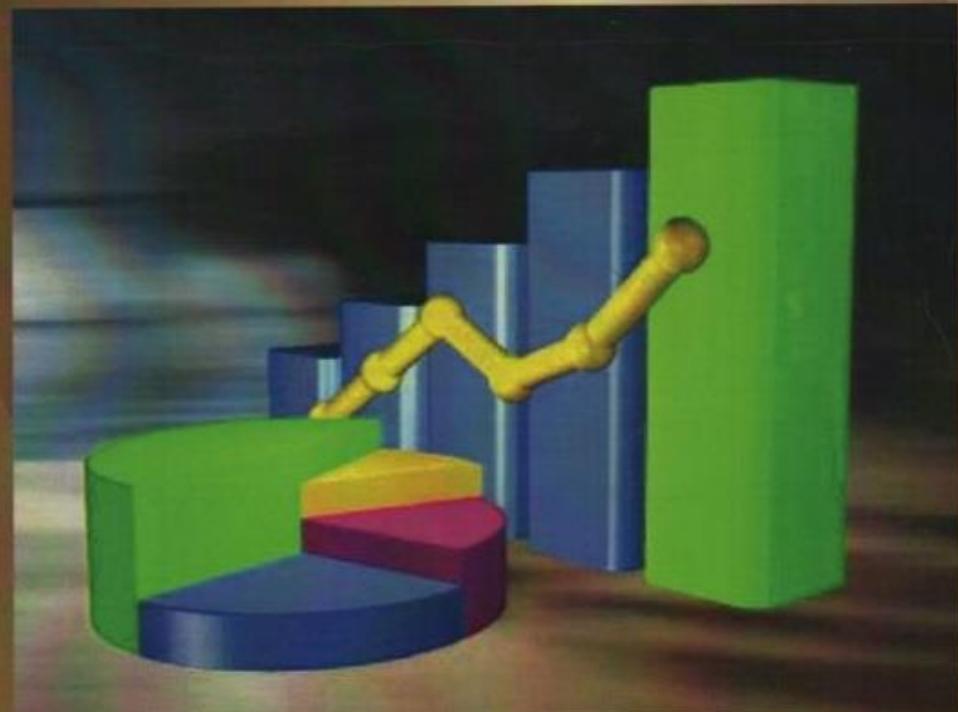


ESTADÍSTICA Y PROBABILIDAD

con aplicaciones

TOMO 1



Lic. Gabriel Leandro Oviedo, MBA

PUBLITEX

Grupo Editorial S.A.

Prefacio

Hay muchos factores que influyen en el desempeño de un estudiante en un curso de estadística. Entre ellos están la clara exposición de los conceptos por parte del profesor, la motivación y empeño por parte del estudiante, los conocimientos previos del alumno, materiales apropiados para el curso, etc. Este texto y todos los materiales digitales que lo acompañan han sido diseñados para ayudar al estudiante y al profesor en todos estos aspectos, pues provee al profesor de materiales que puede emplear en su clase para exponer los conceptos con claridad, presenta al estudiante materiales con ejercicios paso a paso, aplicaciones y amplio uso de la tecnología, de modo que pueda sentirse más motivado al disponer de recursos para adquirir los distintos conceptos y procedimientos, a la vez que se le ofrece gran cantidad de ejercicios resueltos, presentaciones interactivas, videos, entre otras ventajas.

Características didácticas

A lo largo del texto, las definiciones se presentan en recuadros de modo que el estudiante pueda efectuar un repaso posterior de estos conceptos, los cuales son de gran importancia para comprender los ejercicios e interpretar los resultados obtenidos.

UNIDAD ESTADÍSTICA: Es aquella unidad u objeto de interés en la investigación y de ella se desprenden las observaciones, o sea, de ella se derivan los datos para el análisis.

Adicionalmente cada tema incluye gran cantidad de ejemplos. Los ejemplos en el texto impreso se presentan en cuadros que incluyen primero el planteamiento del ejercicio y su solución explicada paso a paso.

Ejemplo:

Suponga que una empresa posee quince vendedores de un determinado producto. Cuatro de los vendedores lograron vender 50 unidades, 6 vendieron 40 unidades, tres vendieron 35 unidades y 2 vendieron 20 unidades. ¿Cuál es el número de unidades promedio de cada vendedor?

Solución:

Dado que existen valores repetidos, entonces se aplica la fórmula:

$$\bar{x} = \frac{\sum_{i=1}^k x_i f_i}{n} = \frac{4 \cdot 50 + 6 \cdot 40 + 3 \cdot 35 + 2 \cdot 20}{15} = 39$$

Es decir, el número de unidades vendidas por cada vendedor es 39 unidades.

Luego de los ejemplos se incluye un ejercicio de revisión, que es similar al ejemplo. La finalidad es que el estudiante pueda efectuar ese ejercicio luego de leer y realizar por cuenta propia el ejemplo. Las soluciones de los ejercicios de revisión se proveen en la página de internet del texto www.auladeeconomia.com/raeep.html.

Ejercicio de revisión:

En un muelle hay 20 contenedores que pesan 15 toneladas cada uno, 25 que pesan 20 toneladas cada uno. ¿Cuál es el peso promedio de los contenedores?

Ver solución de este ejercicio en la página de internet de este texto
www.auladeeconomia.com/raeep.html

Al final de cada capítulo se proveen muchos ejercicios. Estos ejercicios se dividen en dos bloques. Los primeros son ejercicios de desarrollo. Los segundos son ejercicios de selección única y se les llama examen del capítulo. Las soluciones de estos ejercicios se proveen a través de la página de internet del texto.

Ejercicios:**Ejercicios de desarrollo:**

1. A continuación se presenta una lista de variables obtenidas en distintas investigaciones, en cada caso indique cuál tipo de gráfico debería efectuarse para representar en forma adecuada los datos.
 - a. Tiempo medio de espera de los pacientes de un hospital en el servicio de consulta externa obtenido mes a mes durante un año.

Examen del capítulo

1. Se tiene un grupo de n libros. El número de diferentes órdenes posibles de los n libros en una mesa no equivale a:

(a) $P(n,n)$ (b) $n!$

(c) $C(n,n)$ (d) Ninguna de las anteriores

2. Suponga que los n libros se van a conformar en grupos de 3 libros ($n > 3$). El número de diferentes grupos con distinto orden, equivale a:

(a) $P(n,3)$ (b) $n! / 3!$

(c) $C(n,3)$ (d) $P(n,n - 3)$

Por otro lado, cabe destacar que Recursos para el Aprendizaje Efectivo de la Estadística y la Probabilidad es en realidad más que solo un libro, pues se compone de una gran cantidad de recursos en línea de gran utilidad tanto para el estudiante como para el profesor.

The cover features a black and white photograph of a person's hand holding a pen, pointing towards a line graph on a computer screen. The graph shows a bell-shaped curve. To the right of the image, there is a vertical column of text in Spanish. At the bottom left, it says '¿Qué necesita para aprobar un curso de Estadística con excelente nota y comprender claramente los contenidos de esta materia?'. In the center, it says 'Tenemos la respuesta: Recursos para el Aprendizaje Efectivo de la Estadística y la Probabilidad (RAEEP)'. Below that, it says 'RAEEP es una colección completa de una gran cantidad de recursos para el aprendizaje de la estadística descriptiva, la probabilidad y la inferencia estadística.' On the right side, there is a small illustration of a person sitting at a desk with a laptop, and the text 'Cada pasamos más conectados a la real. Los haremos para todo, incluso estudiar. RAEEP te lo facilita'.

Con la factura de compra del texto cada lector podrá recibir una contraseña de ingreso a la página del texto www.auladeeconomia.com/raeep.html y así acceder a una gran cantidad de

- Más de 2000 **diapositivas** que puede ser presentadas por el profesor para impartir su clase, o bien, empleados por el estudiante para repasar posteriormente.
- **Videos** que explican los conceptos y que exponen el uso del software.
- **Ejercicios interactivos.** El texto y las diapositivas contienen más de mil ejercicios con sus soluciones
- **Uso de software Excel y Minitab y uso de la calculadora.** El uso de la tecnología es fundamental en el procesamiento, presentación y análisis de datos y por eso se le concede gran importancia en el texto.

- Además se incluye una gran colección de **otros recursos** disponibles en la web, tales como links a páginas con bases de datos, simulaciones (applets), calculadoras de probabilidades y medidas estadísticas, entre otros recursos disponibles en internet.

El lector debe escribir a info@auladeeconomia.com y así recibirá una contraseña para acceder a los recursos completos.

A lo largo del texto se incluyen cuadros que indican algunos de los principales recursos audiovisuales que pueden ser empleados por estudiantes y profesores.

Material audiovisual:

En la página de internet de este texto podrá encontrar una presentación de diapositivas que expone este tema y es un complemento a este texto.

Solicite su usuario y contraseña escribiendo al correo electrónico info@auladeeconomia.com. Para ello deberá indicar el número de su factura de compra del texto.

En la página los temas están agrupados de modo similar al que se emplea en el texto, tal como se muestra en la imagen, donde se pueden elegir los diferentes recursos.



**Introducción a la
Estadística descriptiva**
Por Lic. Gabriel Leandro, MBA
www.auladeeconomia.com



Medidas de posición en datos no agrupados

En esta segunda semana vamos a pasar a aspectos prácticos, pues empezaremos a calcular las medidas de posición en datos no agrupados.

Para esto emplearemos los siguientes materiales:



Lecciones:

- [Medidas de posición central](#)
- [Medidas de posición: cuantílos](#)
- [Medidas de variabilidad](#)
- [Ejercicios sobre medidas de posición y variabilidad](#)



Videos:

- [Video: Conceptos de medidas de posición](#)
- [Video sobre medidas de posición central](#)
- [Video: Notación de sumatoria o notación sigma](#)
- [Video: Cálculo de medidas de posición y variabilidad en Excel](#)
- [Video: Medidas de posición y variabilidad con Excel](#)

Adicionalmente, en cada capítulo se presenta uno o varios cuadros de Aplicación, los cuales consisten en algunos ejemplos de aplicaciones de la estadística en distintos ámbitos, como las ciencias económicas, la ingeniería, las ciencias sociales y las ciencias de la salud, entre otras áreas del conocimiento humano.

Aplicación:

Índice Dow Jones

Possiblemente en alguna oportunidad usted habrá escuchado noticias sobre el mundo de los negocios que se relacionen con la Bolsa de Valores de Nueva York y entonces habrá escuchado hablar sobre el Dow Jones, el cual es el indicador bursátil más conocido en el planeta.

En términos generales, este material pretende ser una valiosa ayuda para estudiantes y profesores de modo que puedan seguir un mejor proceso de enseñanza – aprendizaje, ya que al disponer de gran cantidad de recursos a través de internet ofrece una serie de ventajas que los textos tradicionales no proveen:

- **Está siempre disponible.** Es como tener al profesor disponible las 24 horas del día y en cualquier lugar donde haya conexión a internet.
- **Es interactivo.** El texto tradicional no es interactivo, pero Recursos para el Aprendizaje Efectivo de la Estadística y la Probabilidad es totalmente interactivo. Muestra una explicación paso a paso y ejercicios solucionados paso a paso. Usted da clic y el material lo lleva a su ritmo.
- **Es motivador e innovador.** Actualmente los jóvenes emplean intensivamente las tecnologías de información y la comunicación. De hecho, la mayoría de la población joven emplea internet como su principal herramienta para investigar y estudiar.
- **Se ajusta a distintos estilos de aprendizaje.** Bien es sabido que cada persona posee diferente modo de aprender. Algunos aprenden más por lo que ven, otros aprenden más por lo que oyen, y aprendemos mucho por lo que hacemos. Recursos para el Aprendizaje Efectivo de la Estadística y la Probabilidad ofrece presentaciones con múltiples imágenes, videos, herramientas tecnológicas, entre otros recursos que ayudan al estudiante a aprender según su estilo de aprendizaje.
- **Fomenta el uso de la tecnología.** Tanto a través del uso de software como Excel y Minitab como por los recursos disponibles en la web se realiza un uso intensivo de las tecnologías de la información y la comunicación.

- **Facilita aprender haciendo.** Recursos para el Aprendizaje Efectivo de la Estadística y la Probabilidad promueve una modalidad de estudio en la que se aprende haciendo, pues usted puede ir desarrollando los ejercicios en su cuaderno o computadora conforme se presentan en el material.
- **Expone aplicaciones diversas:** Recursos para el Aprendizaje Efectivo de la Estadística y la Probabilidad posee ejemplos y ejercicios que se relacionan con la ingeniería, las ciencias económicas, las ciencias de la salud, las ciencias sociales, etc.
- **Siempre está actualizado:** Gracias a internet los lectores podrán tener acceso a nuevas actualizaciones del material disponible en línea.
- **Apoyo a los docentes:** A través de la página del texto se proveen recursos para que los docentes puedan desarrollar mejor sus cursos, a la vez que se propicia el intercambio de experiencias entre profesores.

CONTENIDOS

Agradecimiento	5
Prefacio	7

Capítulo 1

Introducción a la estadística

1.1 Estadística	16
1.2 Estadística descriptiva e inferencia estadística	16
1.3 Importancia del análisis de la información estadística	17
1.4 Fases de una investigación estadística	28
1.5 Fuentes de información o conocimiento	30
1.6 Métodos e instrumentos de recolección de información	31
1.7 Escalas de medida	35
1.8 Ejercicios	39

Capítulo 2

Presentación de la información estadística

2.1 Presentación de los datos	54
2.2 Presentación textual	54
2.3 Cuadros estadísticos	55
2.4 Construcción de cuadros	55
2.5 Representación gráfica	58
2.6 Construcción de gráficos	58
2.7 Clases de gráficos	60
2.8 Ejercicios	75

Capítulo 3

Análisis descriptivo de la información estadística

3.1 Medidas de posición	88
3.2 Medidas de posición central	88
3.3 Los cuantílicos	104
3.4 Medidas de variabilidad	121
3.5 Rango o amplitud total	122
3.6 La varianza y la desviación estándar	123
3.7 Medidas de variabilidad relativa: El coeficiente de variación	132
3.8 Ejercicios	137

Capítulo 4**Distribuciones de frecuencias**

4.1 Necesidad de resumir la información	154
4.2 Gráficos de las distribuciones de frecuencias	161
4.3 Cálculo de medidas de posición y variabilidad en distribuciones de frecuencias	164
4.4 Ejercicios	174

Capítulo 5**Introducción a las probabilidad**

5.1 Probabilidad	184
5.2 Conteo	185
5.3 Definiciones de probabilidad	191
5.4 La frecuencia relativa como probabilidad	192
5.5 Concepto intuitivo o subjetivo de probabilidad	193
5.6 Axiomas y teoremas de probabilidad	193
5.7 Probabilidad condicional	198
5.8 Teorema de Bayes	202
5.9 Ejercicios	205

Apéndice**Tablas y fórmulas****Apéndice 1**

Fórmulas de estadística descriptiva	219
Medidas de variabilidad	220

Apéndice 2

Fórmulas e probabilidad	221
-------------------------------	-----

en el que se presentan los datos en el orden que surgen no tienen alcance si ésta es la única forma de organizar estos resultados. Pero si se lleva a cabo una investigación, ésta debe ser sistemática y lógica para que los resultados sean más fáciles de interpretar y más útiles para las conclusiones que se sacan de ellos. Si se sigue una metodología sistemática, los resultados se obtendrán de una manera más precisa y más útil.

CAPÍTULO 1

INTRODUCCIÓN A LA ESTADÍSTICA

OBJETIVOS

Al concluir el capítulo, será capaz de:

- Explicar el concepto de estadística y sus funciones principales
- Reconocer la diferencia entre estadística descriptiva e inferencial
- Identificar las fases básicas de una investigación estadística
- Conocer las escalas de medición de las variables estadísticas

Al finalizar este capítulo, el lector deberá saber:

- Diferenciar entre estadística descriptiva e inferencial.
- Identificar las fases básicas de una investigación estadística.
- Conocer las escalas de medición de las variables estadísticas.

1.1 Estadística

Con mucha frecuencia la gente habla de estadísticas. Por ejemplo, las estadísticas sobre accidentes en las carreteras, sobre la economía, sobre enfermedades, entre muchos temas. Además, es posible observar en la televisión, los periódicos, internet, libros, revistas, boletines y otros medios distintos gráficos, cuadros con datos, proyecciones, a la vez que se habla de encuestas, realización de experimentos, entre otros temas relacionados con la estadística.

ESTADÍSTICA: La estadística es un conjunto de métodos y teorías aplicadas a la recolección, descripción y análisis de datos, los cuales constituyen evidencia numérica para la toma de decisiones en condiciones de incertidumbre.

Esta definición es particularmente útil en los distintos campos de acción, en donde los profesionales en las distintas áreas diariamente tienen que tomar decisiones sobre situaciones que comprenden tanto datos numéricos como incertidumbre, en situaciones que van desde la contratación de personal cuyo desempeño no puede ser anticipado, el requerimiento de materias primas para la producción de artículos cuya demanda es incierta, las decisiones en materia de política pública cuando no se conoce la evolución futura de las variables económicas y sociales, las prescripciones que un profesional de la salud puede realizar aun cuando no tiene certeza de la reacción del paciente ante el tratamiento, entre muchas otras posibles situaciones.

Material audiovisual:

En la página de internet de este texto podrá encontrar una presentación de diapositivas que expone este tema y es un complemento a este texto.

Solicite su usuario y contraseña escribiendo al correo electrónico info@auladeeconomia.com. Para ello deberá indicar el número de su factura de compra del texto.



Introducción a la
Estadística descriptiva
Por Lic. Gabriel Leandro, MBA
www.auladeeconomia.com

1.2 Estadística descriptiva e inferencia estadística

El énfasis sobre la importancia de la estadística en la toma de decisiones es bastante reciente. En sus primeros tiempos los estudios estadísticos consistían primordialmente de metodologías para recopilar y describir datos numéricos, casi siempre en forma de tablas y diagramas. A esta área de la disciplina se le conoce ahora como *Estadística descriptiva*. Los métodos mencionados contrastan con el enfoque moderno en el cual se realizan generalizaciones acerca de la totalidad de los datos, los cuales constituyen la *población*, a partir

de las investigaciones realizadas en una parte de la población, a la que se llama *muestra*. El estudio de cómo realizar estas generalizaciones constituye la *estadística inferencial* o *inferencia estadística*.

Estos métodos inferenciales se requieren para pronosticar las tendencias de la demanda de los artículos (con base en períodos anteriores a las ventas) o para estimar las necesidades de capital de trabajo (basándose en crecimientos o decrecimientos de los costos) o la necesidad de ajuste de una máquina (analizando las tendencias de ciertas mediciones importantes) o los porcentajes defectuosos, etc. En todos estos ejemplos existe incertidumbre puesto que no se cuenta con la información completa y es mediante la *inferencia estadística* que se puede llegar a encontrar el curso de acción que facilita la toma de decisiones.

1.3 Importancia del análisis de la información estadística

Tanto individuos como organizaciones necesitan información para realizar sus distintas actividades, sin embargo, para que ésta información sea válida y confiable, es necesario que haya sido obtenida y procesada apropiadamente. Ahí precisamente es donde entra en juego la estadística, y por eso es de gran relevancia en todo proceso de investigación científica.

Aplicación

Utilidad de la estadística

En prácticamente todos los campos del conocimiento humano se emplea la estadística. Sin tratar de ser exhaustivos, se mencionan algunas aplicaciones importantes de la estadística en algunos áreas distintas:

En el ámbito de las empresas se emplea la estadística en:

- El control de la rentabilidad de la empresa, de sus ventas y de sus costos, entre otras partidas.
- La investigación de mercados, la medición de la satisfacción de los clientes y las proyecciones de la demanda.
- La evaluación de oportunidades de inversión y la gestión de los riesgos presentes en toda inversión.
- La evaluación de desempeño, la evaluación del clima organizacional, el control de la seguridad ocupacional, etc.
- La verificación del alcance de las metas, la toma de decisiones bajo condiciones de riesgo e incertidumbre, etc.

En el ámbito de la ingeniería:

- El control de la calidad, de la productividad, de los inventarios, etc.

- Al desarrollar un proyecto urbanístico se requieren registros demográficos, datos sobre el abastecimiento de agua potable, etc.
- Son necesarios registros del tránsito para diseñar el tiempo que duran los semáforos.
- En la descripción de modelos termodinámicos complejos (termodinámica estadística).
- Al tratar de comprender la variabilidad de sistemas de medición y el control de procesos.

En el ámbito de la psicología, la educación y otras ciencias del comportamiento, se emplea el método científico tratando de cuantificar distintos fenómenos y aplicando técnicas estadísticas como la prueba de hipótesis y la realización de experimentos.

- En el contexto de la psicología laboral son muy importantes los aspectos psicológicos presentes en el comportamiento individual y colectivo en el lugar de trabajo de las personas.
- El estudio de los fenómenos del aprendizaje y de ese modo cómo mejorar los procesos de enseñanza y aprendizaje.
- Las investigaciones sobre los distintos trastornos mentales.
- La utilización de diversos tipos de pruebas para obtener información tanto a nivel individual como colectivo.

En el ámbito de las ciencias de la salud:

- Tanto el diagnóstico, e pronóstico como la terapia que un profesional en salud pueda realizar se basan en datos estadísticos que han determinado la eficacia y riesgos de las distintas acciones posibles.
- En la administración sanitaria son fundamentales los datos, a fin de realizar la mejor utilización de los recursos disponibles.
- En la realización de ensayos clínicos, fundamentales para que luego se puedan emplear en diversos tratamientos médicos.
- En decisiones de políticas públicas son fundamentales los indicadores del estado de la salud para los países y sus distintas regiones, sean estados, distritos, etc.

En el ámbito de las ciencias de la comunicación y otras ciencias sociales:

- Se pueden medir las opiniones de distintos grupos, o medir variables importantes como la intención de voto.
- Cuantificar la audiencia de un programa de televisión o de un evento específico.
- Monitorear distintos medios de comunicación y monitorear la publicidad.
- Presentar información cuantitativa de forma accesible y adecuada para los consumidores de noticias.
- Determinar la efectividad de las comunicaciones realizadas por una empresa, un partido político, un gobierno, etc.

En general, resulta evidente que la Estadística provee la información necesaria y herramientas para analizar datos y así facilita la toma de decisiones y la solución de problemas.

Funciones de la estadística

Se pueden establecer como las funciones básicas de la estadística las siguientes:

1	2	3	4	5
Recopilar los datos	Procesar los datos	Presentar los datos	Analizar los datos	interpretar los datos

1. **Recolección:** La estadística busca recolectar en forma adecuada la información. Con este fin muchas veces se emplean cuestionarios, la observación, experimentos y otros tipos de instrumentos, a fin de recopilar los datos requeridos.
2. **Procesamiento:** Una vez que se han recopilado los datos, se desea organizarlos, lo cual significa que hay que corregir, clasificar y tabular los datos que se obtuvieron de la etapa anterior, para lo cual también es necesario hacer uso de las técnicas estadísticas apropiadas.
3. **Presentación:** Luego es necesario presentar adecuadamente los datos. Para esto generalmente se hace uso de cuadros estadísticos y gráficos, y se calculan diferentes tipos de medidas, como los promedios, las proporciones, varianzas, se elaboran distribuciones de frecuencias, etc.
4. **Análisis:** Esto es la aplicación y desarrollo de modelos, establecimiento de correlaciones, estimaciones, proyecciones, etc. El análisis puede ser de una sola variable en un momento dado o podría ser de series de tiempo, o sea, analizar cómo ha evolucionado la variable a lo largo del tiempo. El análisis podría ser de la relación que una variable tiene con otras variables, por ejemplo, determinar si una variable se correlaciona con otra.
5. **Interpretación:** Una vez obtenidas las distintas medidas, estos nuevos datos son empleados para interpretar el comportamiento de las variables en estudio, para realizar proyecciones, tomar decisiones, etc.

Conceptos básicos de estadística descriptiva

A continuación se describen algunos de los más importantes conceptos de la estadística descriptiva.

UNIDAD ESTADÍSTICA: Es aquella unidad u objeto de interés en la investigación y de ella se desprenden las observaciones, o sea, de ella se derivan los datos para el análisis.

La clara definición de la unidad estadística es el aspecto más importante del proceso de obtención de información, pues la unidad estadística es la generadora de dichos datos, no es la información misma (observaciones), sino que es la fuente u origen de donde se obtienen los datos.

Por ejemplo, si un investigador desea estudiar la proporción de hogares pobres en un país, entonces las unidades estadísticas serían las familias, ya que el objetivo de su estudio es determinar cuáles familias son pobres y cuáles no.

Dependiendo del estudio, la unidad estadística puede ser una persona, una familia, una empresa, un árbol, un país, un ecosistema, etc.

Cabe recalcar la importancia de su delimitación en al menos 4 dimensiones básicas:

1. ¿Qué/Quién?
2. ¿Dónde?
3. ¿Cuándo?
4. ¿Cómo?

Realmente decir que la unidad estadística es una persona es insuficiente, pues hay que agregar cuáles son las cualidades que definen a esa unidad estadística, por ejemplo, según su edad, su lugar de residencia, el periodo y cualquier otro aspecto relevante. Entonces sería mejor decir que la unidad estadística es un “profesional en periodismo que haya residido en la ciudad capital del país durante los últimos 12 meses”.

CARACTERÍSTICAS O VARIABLES: Cada unidad estadística posee una o varias características, también llamadas variables, que son distintas de una unidad estadística a otra y que el estudio tiene interés en conocer.

De una unidad estadística a otra hay una serie de diferencias de interés para el investigador, por eso la característica o variable es cambiante. Por tanto, las variables son las características de los sujetos, que debido a su variabilidad, se requiere conocer.

Continuando con el ejemplo del estudio de la pobreza, se espera considerar ciertas “características” o “variables” de las familias del país, como lo pueden ser el ingreso del hogar, el número de personas que laboran en la familia, las condiciones de la vivienda (acceso a agua potable, disponibilidad de baño, etc.), el nivel educativo de los miembros de la familia, etc.

Tal como se ha podido observar las características o variables pueden ser variables **cuantitativas**, cuando producen directamente una magnitud numérica, o bien, pueden ser variables **cualitativas**, cuando no pueden cuantificarse. En este último caso se recurre a una clave o un código para su proceso. Más adelante en este capítulo se ampliará con respecto a este último punto.

Cuando una variable cuantitativa se puede determinar por conteo, se dice que es una variable **discreta**, como es el caso del número de hijos que una persona tiene, el número de cursos que un estudiante universitario está realizando, el número de empleados de una empresa o el número de artículos defectuosos obtenidos por día en un proceso de producción, entre muchos otros posibles ejemplos. Por otro lado, cuando la variable se obtiene por medición, se dice que es una variable **continua**. En estos casos se emplea un instrumento de medición, tal como una balanza, un termómetro o un cronómetro, y la unidad se expresa con unidades de medida, tales como los gramos, grados centígrados, segundos, etc. Algunos ejemplos son el peso de paquetes de azúcar, la temperatura máxima diaria en una ciudad, tiempo de producción de un artículo, etc.

Ejercicio de revisión

Clasifique las siguientes variables como cualitativas o cuantitativas, y en caso de ser cuantitativas señale si son discretas o continuas:

1. Marca de un refresco producido en el país.
2. Grado académico de un profesional.
3. Ingreso mensual familiar.
4. Número de hijos.
5. Talla de una camiseta (pequeña, mediana, grande).
6. Número de la talla de un pantalón (10, 12, etc.).
7. Tiempo de espera en una fila para recibir un servicio.
8. Ciudad de residencia.
9. Calidad de un producto (sin defectos, con defectos menores o con defectos mayores).
10. Peso de un paquete de harina.
11. Nombre del principio activo de un medicamento.
12. Número de personas en una fila.
13. Cantidad de energía eléctrica consumida por mes en una empresa.
14. Número de artículos defectuosos por línea de ensamble.
15. Consumo de calorías por día.

Ver solución de este ejercicio en la página de internet de este texto:
www.auladeeconomia.com/raeep.html

OBSERVACIÓN: Cada característica bajo estudio se expresa numéricamente, ya sea porque se mide, se cuenta o se le asigna un código, por lo que cuando se recopilan los datos se tiene una serie de valores numéricos, y a cada de ellos se le llama observación.

O sea, una observación es un dato. No es el resultado de la investigación como un todo. Tomando el mismo ejemplo de la pobreza, entonces una observación para la variable *Ingreso mensual del Hogar* puede ser \$500. En este caso por ser una característica cuantitativa su medición es simple. En otros casos lo que hace es un recuento, como lo sería para la variable *Número de miembros del hogar que laboran*, donde una observación podría ser 2. En otros casos se asigna un código, por ejemplo si se desea considerar la variable *Grado académico del jefe del hogar*, donde se podría establecer una codificación como: 1. Ninguno. 2. Primaria. 3. Secundaria. 4. Técnica o para universitaria. 5. Universitaria. En este caso una observación podría ser 3 para un jefe de hogar que realizó sus estudios de secundaria.

POBLACIÓN: Es el conjunto total de elementos de referencia o unidades estadísticas sobre el que se interesaría realizar la observación en una investigación estadística.

El investigador desea entonces conocer cuáles serían esas observaciones para todas las familias del país, y poder analizar esos datos. Ese conjunto total de observaciones constituye su población.

La población bajo estudio puede ser **infinita** o **finita**. Es infinita cuando tiene un número infinito de elementos, es decir, no termina. Por ejemplo, el agua que corre por un río. Al contrario, una población finita tiene un número limitado de elementos. Una población infinita podría convertirse en una finita si se le establece un intervalo de tiempo. Por ejemplo, el número de habitantes de una ciudad podría ser una población infinita si se considera a lo largo de todo el tiempo, pero si se considera el número de habitantes en un momento dado, entonces será una población finita.

Las poblaciones pueden ser **homogéneas** o **heterogéneas**, según si los valores de las características en estudio son iguales o diferentes, respectivamente. Así, la sangre que una persona posee será una población homogénea, pero los empleados de una empresa serán una población heterogénea en cuanto a sus salarios, por ejemplo.

Ejercicio de revisión

El encargado de recursos humanos de una empresa va a realizar un estudio de clima organizacional. Para ello va a aplicar un cuestionario a todos los funcionarios que actualmente laboran en la empresa y les va a pedir que den su opinión con respecto a la comunicación dentro de la empresa, el liderazgo de los gerentes, las relaciones interpersonales, entre otros aspectos. Con respecto a esta situación indique:

1. ¿Cuál es la unidad estadística?
2. ¿Cuál puede ser un ejemplo de una característica o variable cuantitativa que pueda interesar en este estudio?

3. ¿Cuál puede ser un ejemplo de una característica o variable cualitativa que pueda interesar en este estudio?
4. ¿Cuál es la población?

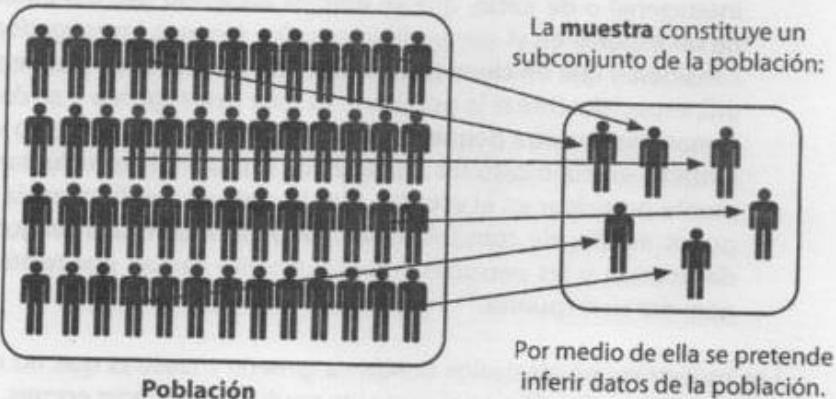
Ver solución de este ejercicio en la página de internet de este texto:
www.auladeeconomia.com/raeep.html

PARÁMETRO: Es una medida descriptiva de la población total de todas las observaciones de interés para el investigador.

Si el investigador tuviera toda esa población de datos desearía calcular algunas medidas que le ayuden a describir el total de datos. Por ejemplo, desearía conocer parámetros como el ingreso promedio de esa población, la proporción de familias que satisfacen todas sus necesidades básicas, etc.

Dado que muchas veces es demasiado costoso o requiere demasiado tiempo obtener las observaciones para toda la población, e incluso a veces es imposible, entonces se emplea una muestra, la que si es bien seleccionada, va a representar a la población, y entonces las conclusiones que en ella se obtengan luego podrán ser generalizadas al resto de la población.

MUESTRA: Es una parte de la población que se selecciona para ser estudiada, ya que estudiar toda la población es demasiado costoso, toma demasiado tiempo o es imposible.



ESTADÍSTICO: Elemento que se obtiene como una función de las observaciones obtenidas en la muestra, que describe dicha muestra y sirve como una estimación del parámetro de la población correspondiente.

Tal como se mencionó anteriormente, el investigador desea conocer el ingreso de todas y cada una de las familias de la población, pero como obtener ese dato sería demasiado costoso y toma demasiado tiempo, entonces toma una muestra representativa y obtiene el ingreso promedio de las familias de la muestra. Este último dato es un estadístico que servirá como estimador del parámetro poblacional.

En la práctica pueden usarse dos tipos de muestra, según el método que se utilice para hacer la selección de los elementos:

- **Muestreo aleatorio o muestreo estadístico:** en este tipo de muestreo se emplea un procedimiento que conceda a cada unidad de la población una probabilidad conocida de formar parte de la muestra. Por ejemplo, en el muestreo simple al azar, todas las unidades de la población tienen la misma probabilidad de ser incluidos en la muestra. Además del muestreo simple al azar, existen otras técnicas como el muestreo estratificado, en etapas, sistemático, de conglomerados, etc. En el capítulo 9 se amplían estos conceptos sobre muestreo.
- **Muestreo no aleatorio:** En este tipo de muestreo se emplea un procedimiento mediante el cual no es posible calcular la probabilidad de que cada unidad de la población sea seleccionada en la muestra. Existen varias modalidades, por ejemplo, la selección por conveniencia, en la cual el investigador incluye a aquellas unidades que le facilitan más su trabajo, lo cual puede introducir grandes sesgos en la investigación. También existe el muestreo intencional o de juicio, que se emplea haciendo uso del criterio de un experto en el campo bajo estudio. Algunos investigadores consideran que en ciertos casos este tipo de muestreo puede ser útil, especialmente si la población es muy heterogénea y se debe tomar una muestra pequeña. Otra posibilidad es el muestreo voluntario, en cuyo caso las unidades de estudio eligen voluntariamente participar en el estudio. Esto lo realizan con frecuencia algunos medios de comunicación cuando formulan una pregunta de opinión y las personas voluntariamente llaman por teléfono para dar su respuesta.

Los muestreos no aleatorios tienden a generar muestras que no son representativas de la población y por tanto los resultados incluyen errores. Estos errores en la elección de los individuos o grupos que se dan en los muestreos no aleatorios se conocen como sesgos de selección. La principal ventaja del muestreo aleatorio es que limita ese sesgo de selección, e incluso es posible cuantificarlo y controlarlo.

Muestra	No aleatoria	Por conveniencia De juicio Voluntario Otros	Implican sesgos de selección, o sea, muestras no representativas y la imposibilidad de hacer inferencias.
	Aleatoria	Simple al azar Sistématico Estratégico De conglomerados	

Cuando en una investigación la información se obtiene mediante el uso de una muestra, es posible que aparezcan dos tipos de errores, que son los errores de muestreo y los sesgos. Los errores de muestreo solo pueden aparecer en las investigaciones en las que se emplea una muestra. Este error surge porque la muestra se obtiene al azar, entonces no va a ser "perfectamente" representativa de la población. El único modo de eliminar estos errores es utilizar la población completa. La mayoría de los investigadores están dispuestos a aceptar un error de muestreo relativamente pequeño, pues esto permite reducir los costos del estudio y controlar mejor el trabajo de la recopilación de datos.

ERROR DE MUESTREO: Es la diferencia, causada por el azar, entre el valor verdadero de una característica y el valor estimado mediando el uso de una muestra.

Si un investigador desea saber qué tan grande puede ser el error de muestreo, puede utilizar los modelos de probabilidad apropiados para determinar su magnitud. Después, si se quisiera reducir ese error de muestreo, entonces puede emplear una muestra de mayor tamaño.

Ahora bien, los sesgos pueden aparecer tanto en estudios por muestreo como en aquellos que realizan un censo, o sea, que abarcan la población completa. Por ejemplo, si se ha diseñado un cuestionario inadecuado, entonces habrá ciertos errores que no se corregirán ni siquiera empleando la población completa. Los errores de muestreo surgen por la naturaleza aleatoria de la muestra, mientras que los sesgos son errores sistemáticos que se dan en el proceso de investigación.

SESGO: Error sistemático, no debido al azar, entre el valor verdadero de una característica y el valor estimado en una investigación.

Los sesgos son errores de naturaleza sistemática, o sea que producen errores que se dan en el mismo sentido. Por ejemplo, si en un estudio de satisfacción de los usuarios de un servicio, las encuestas se aplican en los

momentos en que hay menor cantidad de usuarios y el servicio es más rápido, entonces los resultados tenderán a dar una evaluación del servicio mejor que la real. Los sesgos generalmente no son cuantificables, pero deben evitarse mediante las medidas administrativas adecuadas, que procuren que todo el proceso de recolección, procesamiento y análisis de los datos se realice del mejor modo posible.

Son muchas las posibles causas de los sesgos, pero algunos sesgos se producen por las siguientes razones:

- Se define la unidad estadística en forma inadecuada y, por tanto, no se identifica claramente cuál es la población.
- Se selecciona inadecuadamente la muestra.
- Los instrumentos de recolección de datos no se diseñaban correctamente o no se aplican del modo adecuado.
- La recolección de los datos no se realiza bien, por ejemplo, cuando los entrevistados no contestan todas las preguntas que se les aplican, o cuando el entrevistador influye en las respuestas.
- No se emplea el método de estimación correcto, o bien, se cometen errores en los cálculos, etc.



Ejemplo

Un funcionario de un banco desea hacer una evaluación de cliente interno, es decir, una evaluación de ciertos servicios que los distintos departamentos del banco se prestan entre sí. Con ese fin ha elaborado un cuestionario, el cual, por el nivel de sus contenidos deberá ser aplicado al personal que ocupa puestos de jefatura. El banco posee 5.000 empleados, pero solo 350 ocupan puestos de jefatura. Con base en la información anterior, determine:

1. ¿Cuál es la unidad estadística?
2. ¿Cuál puede ser un ejemplo de una característica o variable cuantitativa que pueda interesar en este estudio?

3. ¿Cuál puede ser un ejemplo de una característica o variable cualitativa que pueda interesar en este estudio?
4. ¿Cuál es la población?
5. ¿Vale la pena emplear una muestra o es mejor aplicar el cuestionario a toda la población?
6. Suponga que se aplica el cuestionario a la población, ¿habría error a la hora de hacer la estimación de los parámetros investigados?
7. ¿Cuál podría ser un posible sesgo?

Solución

1. Dado que el cuestionario solo debe ser aplicado a los puestos de jefatura, entonces la unidad estadística no corresponde a un empleado del banco, sino a un empleado que ocupe un puesto de jefatura en el periodo en el cual se va a realizar el estudio.
2. Una característica o variable cuantitativa que pueda interesar en este estudio puede ser el tiempo que tiene el funcionario de laborar para el banco, o el número de empleados que tiene como subordinados, entre muchas otras posibles respuestas.
3. Una característica o variable cualitativa que pueda interesar en este estudio puede ser la valoración que hace del servicio que presta otro departamento (calificándolo como muy bueno, bueno, regular, malo o muy malo), o el departamento para el cual labora el funcionario que contesta el cuestionario, entre muchas otras respuestas posibles.
4. Dada la definición que se hizo en la pregunta 1 de la unidad de estudio, la población correspondería al conjunto de empleados que ocupen puestos de jefatura en el periodo en el cual se va a realizar el estudio.
5. Una población está compuesta por 350 personas no es demasiado grande, por lo que podría emplearse la población. Sin embargo puede ser que las oficinas se encuentren distribuidas a lo largo de todo el país, y que por aspectos de costo y tiempo sea mejor emplear una muestra.
6. Si se aplica el cuestionario a la población, entonces no habrá error de muestreo a la hora de hacer la estimación de los parámetros investigados. Este error aparece solo cuando se utiliza una muestra en el estudio.
7. Existen muchos posibles sesgos, pero uno muy frecuente es el diseño inadecuado del cuestionario. Por ejemplo, que contenga preguntas mal redactadas, que sugieran la respuesta, etc.

Ejercicio de revisión

Un investigador está interesado en conocer el impacto de las relaciones entre padres e hijos sobre el desempeño académico de los niños en edad escolar. Para realizar su estudio ha diseñado un cuestionario que desea aplicar a una muestra de niños en varias escuelas de la ciudad capital. Con respecto a esta situación indique:

1. ¿Cuál es la unidad estadística?
2. ¿Cuál puede ser un ejemplo de una característica o variable cuantitativa que pueda interesar en este estudio?
3. ¿Cuál puede ser un ejemplo de una característica o variable cualitativa que pueda interesar en este estudio?
4. ¿Cuál es la población?
5. ¿Vale la pena emplear una muestra o es mejor aplicar el cuestionario a toda la población?
6. ¿Cuáles ventajas y desventajas tendría realizar el estudio empleando una muestra no aleatoria?
7. ¿Cuáles ventajas y desventajas tendría realizar el estudio empleando una muestra aleatoria?
8. Dé un ejemplo de un posible sesgo que podría presentarse en un estudio de este tipo.

Ver solución de este ejercicio en la página de internet de este texto:
www.auladeeconomia.com/raeep.html

Apoyo audiovisual

En la página de internet www.auladeeconomia.com podrá encontrar una presentación de diapositivas que expone este tema y es un complemento a este texto. Puede hallar este material dentro del tema 1: Introducción a la estadística. Adicionalmente puede encontrar algunos videos explicativos.

Solicite su usuario y contraseña escribiendo al correo electrónico info@auladeeconomia.com. Para ello deberá indicar el número de su factura de compra de texto.

1.4 Fases de una investigación estadística

La realización de una investigación estadística es un proceso planeado que requiere efectuar una serie de etapas o fases claramente establecidas. A continuación se exponen las etapas básicas para el desarrollo de una investigación estadística.

I. Planteamiento del problema de investigación

La primera fase comienza con el descubrimiento o planteamiento deliberado de un problema relacionado con el campo de actividad del investigador, y cuya solución es de interés para él individualmente, para el equipo con el que trabaja o para una institución en particular. Para poder efectuar el planteamiento del problema se debe aclarar previamente lo que se entiende por objeto de estudio en una investigación, o sea, la unidad estadística, y lo que se entiende por problema.

II. Diseño y selección de la muestra

En la mayoría de los estudios se necesita emplear una muestra, pues trabajar con toda la población generalmente resulta más costoso, puede demandar mucho tiempo, resultar imposible e inclusive innecesario (pues con una muestra se pueden obtener resultados suficientemente buenos y es posible supervisar mejor el trabajo). Diseñar y seleccionar la muestra es una fase clave, pues es necesario determinar el tamaño apropiado de la muestra y la técnica de muestreo adecuada, para que evitar sesgos en el estudio y que se pueda llegar a conclusiones útiles.

III. Diseño de instrumentos de recolección

Con frecuencia los datos no existen, es necesario emplear un instrumento para recolectarlos. Esta fase es importantísima para una investigación, pues es fundamental el diseño adecuado de los instrumentos de recolección de los datos. Se requiere que un instrumento de recolección de datos sea válido y confiable. Cuando el instrumento no está bien diseñado, los datos obtenidos estarán sesgados, afectando las conclusiones que se obtengan a partir de ellos. En la sección 1.6 se amplía este tema.

IV. Presentación de los datos finales

Los investigadores deben presentar en sus informes los datos obtenidos de manera sintética, clara y ordenada. Las formas más comunes son los cuadros o tablas y las gráficas.

- **Cuadros:** La presentación tabular o en forma de cuadro o tabla se usa cuando el número de datos que se quiere presentar es relativamente grande. Un cuadro sirve para dar énfasis a ciertas cifras, hacer comparaciones, facilitar el análisis de los datos y como fuente de información.
- **Gráficos:** Los gráficos son figuras que sirven para representar mediante elementos geométricos a un conjunto de datos estadísticos. Tienen como ventaja el ser más eficaces para llamar la atención que las otras formas de presentación de datos. Si están bien construidos son más fáciles de comprender que un cuadro y facilitan las comparaciones. En general permiten una comprensión más clara y rápida de lo que presentan los datos, lo cual favorece el análisis.

V. Análisis cuantitativo e interpretación de los datos

El análisis cuantitativo de los datos finales consiste en el tratamiento matemático de la información que permite realizar comparaciones y establecer relaciones. Para ello se pueden emplear simples técnicas descriptivas, o bien análisis propios de la estadística inferencial.

- El **análisis descriptivo** consiste en el cálculo de medias, medianas, varianzas, entre otras medidas que ayudan a describir el comportamiento de las variables estudiadas.
- El **análisis estadístico inferencial** se emplea cuando los datos provienen de una muestra, y se requiere efectuar una inferencia con respecto a la población, obteniendo a su vez una medida de la probabilidad de error implícito en dicha inferencia. En este sentido hay dos resultados principales que se espera llevar a cabo, que son la diferencia entre grupos y la correlación entre variables.

VI. Presentación del informe final de una investigación

En el informe final de la investigación se busca comunicar principalmente los siguientes aspectos:

- Problema que dio origen a la investigación.
- Las hipótesis planteadas y cómo se pusieron a prueba.
- Metodología empleada en la recolección y análisis de los datos.
- Resultados obtenidos.
- Conclusiones finales.

1.5 Fuentes de información o conocimiento

Cuando se realiza una investigación es posible que la información ya exista, como es el caso en el cual se ha efectuado alguna investigación previa sobre el tema, o bien, como cuando alguna institución del gobierno recopila periódicamente los datos de interés. En estos casos se dice que se emplea una fuente de información. Las fuentes consisten en documentos, obras o elementos que proveen información o datos que ya han sido recopilados.

Según la relación con la investigación las fuentes de información pueden ser:

- **Fuentes primarias:** Es aquel documento, obra o elemento publicado o creado por una persona o una institución mediante el cual proporciona datos que han sido recopilados por ella misma. Por ejemplo, en la mayoría de los países existe un instituto de estadística que determina datos como los niveles de desempleo, de pobreza y otras variables. Cuando se consulta un informe de uno de estos institutos de estadística, entonces se está consultando una fuente primaria.
- **Fuentes secundarias:** Es aquel documento, obra o elemento publicado o creado por una persona o una institución mediante el cual proporciona datos que no han sido recopilados por ella misma. Con respecto al mismo ejemplo dado en el caso de las fuentes primarias, si los datos del desempleo se consultan en un medio de comunicación, como un periódico, entonces este medio es una fuente secundaria, pues el periódico no determina el nivel de

desempleo en el país, sino que obtiene sus datos de otra fuente (un informe del instituto de estadística, por ejemplo).

Ejercicio de revisión

Clasifique las siguientes fuentes de información como primarias o secundarias:

- Artículo de un periódico sobre el crecimiento de las exportaciones del país.
- Reporte del instituto de estadística del país sobre la evolución del desempleo a nivel nacional.
- Informe del Fondo Monetario Internacional sobre las tasas de inflación de los países de América Latina.
- Estado de pérdidas y ganancias de una compañía entregado a sus accionistas.
- Folleto de la Organización Panamericana de la Salud sobre la prevalencia de las enfermedades cardiovasculares en los países de América Latina.
- Artículo de una revista científica en que un investigador presenta los hallazgos que obtuvo sobre la salud bucodental de una comunidad rural del país.
- Anuario estadístico del Banco Interamericano de Desarrollo sobre la infraestructura vial en los países de América Latina.
- Anuario estadístico del Ministerio de Hacienda sobre la recaudación fiscal.
- Reporte sobre el control de la calidad de una línea de producción.
- Informe sobre las mercaderías en existencia de una tienda.

Ver solución de este ejercicio en la página de internet de este texto:
www.auladeeconomia.com/raeep.html

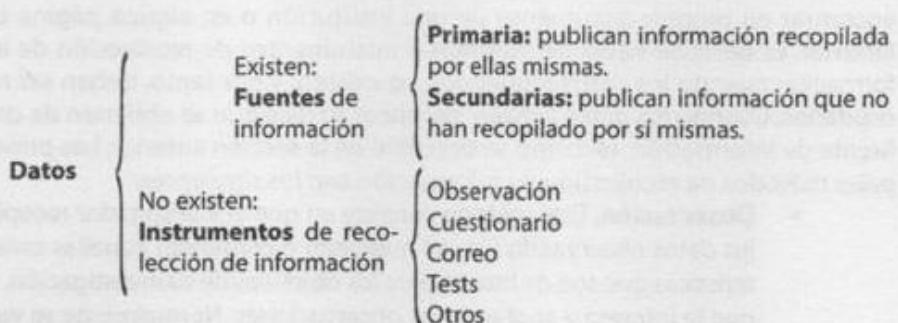
1.6 Métodos e instrumentos de recolección de información

Los métodos e instrumentos de recolección de información se emplean cuando la información no ha sido recopilada, por ejemplo, si se desea determinar la demanda de un nuevo producto, esa información no se va a encontrar en ningún documento de una institución o en alguna página de internet. Es decir, se habla de métodos e instrumentos de recolección de información cuando los datos requeridos no existen, y por tanto, deben ser recopilados. Cuando los datos existen, entonces se dice que se obtienen de una fuente de información, tal como se describió en la sección anterior. Los principales métodos de recolección de información son los siguientes:

- Observación.** Este método consiste en que el investigador recopila los datos observando (ya sea midiendo o contando) aquellas características que son de interés para los objetivos de su investigación, lo que le interesa y anotando las observaciones. Normalmente se va a

apoyar en algún instrumento de medición y va a registrar las observaciones en un registro apropiado.

- **Entrevista personal.** Este método se emplea principalmente en las ciencias sociales y las ciencias económicas. El investigador visita a las unidades de estudio, tales como empleados de una empresa, familias de una zona, pacientes de un hospital, etc. (o a las unidades de información, pues en algunos casos la unidad estadística no puede suministrar la información, tal como en el caso de un niño, por lo que la unidad de información podría ser alguno de sus padres) y les plantea una serie de preguntas, las cuales se registran en un cuestionario previamente diseñado para este fin.
- **Entrevista telefónica.** Es similar al método anterior, pero la diferencia es que los informantes son contactados por medio de una llamada telefónica. Tiende a ser más económico que la entrevista personal.
- **Correo.** En este caso también se emplea un cuestionario, pero se envía por correo, ya sea correo postal o correo electrónico. El cuestionario luego se devuelve por la misma vía. Actualmente existe software que facilita la aplicación de encuestas a través de internet. Tiende a ser un método bastante económico.
- **Registro.** Si usted desea saber cuántas personas han contraído matrimonio durante el último año o cuánto ha exportado el país durante este semestre, no va a necesitar ir a aplicar alguno de los métodos anteriores, ya que en la mayoría de los países existe alguna institución que registra dichos eventos.
- **Test.** Es una técnica derivada de la entrevista y la encuesta. Tiene como objeto lograr información sobre rasgos definidos de la personalidad, la conducta o determinados comportamientos y características individuales o colectivas (inteligencia, interés, actitudes, aptitudes, rendimiento, memoria, manipulación, etc.).
- **Otros métodos** pueden ser: panel, estudios de caso, estudios históricos, análisis de contenido, investigación acción, investigación participativa, entrevistas en profundidad, grupos de discusión y los diseños no-experimentales.



De alguna manera el investigador debe garantizarse que el instrumento de medición que está empleando muestra lo que realmente se desea saber. En este sentido se evalúa la calidad de los instrumentos de medición considerando dos aspectos:

- **Confiabilidad:** Se relaciona con la capacidad del instrumento de dar medidas "certeras", es decir, que se acerquen a la verdadera medida de la variable considerada.
- **Validez:** Se refiere a que el instrumento mida realmente lo que se quiere medir. Para que un instrumento sea válido debe ser confiable, sin embargo podría ser confiable sin ser válido.

Aplicación

Psicometría

Possiblemente en algún momento de nuestra vida hemos realizado algún test de personalidad o algún otro tipo de prueba psicométrica, por ejemplo, un test sobre nuestro estilo de aprendizaje.

Estas pruebas tienen por objetivo la medición de alguna variable o conducta psicológica. Claro está que estas variables son no fácilmente observables, por lo que la elaboración y aplicación de estos test no es tarea sencilla. Precisamente de eso se encarga la psicometría, la cual busca cómo construir mecanismos válidos y confiables de medición de estas variables psicológicas. Así, se han desarrollado pruebas relacionadas con la medición de la inteligencia, rasgos de personalidad, habilidades, actitudes, creencias, problemas de aprendizaje, entre muchos otros aspectos.

La psicometría se apoya en una serie de teorías sobre el proceso de medición, el diseño de escalas y la elaboración de test. La finalidad es poder crear herramientas de medición válidas y confiables, o sea, que midan los que se supone que deben medir (validez) y que lo midan bien (confiabilidad). Por ejemplo, cuando se desarrolla un instrumento sobre estilos de aprendizaje, la validez y la confiabilidad consistirían en que dicho instrumento realmente sea capaz de determinar cuál o cuáles son los estilos de aprendizaje que una persona posee.

Elaboración del cuestionario

El cuestionario es uno de los métodos que más se emplea en las ciencias sociales y las ciencias económicas, e incluso también en las ciencias de la salud y otros campos. Tiene la ventaja de que puede ser más económico y que permite obtener información que sería muy difícil o imposible obtener por otros medios. Ahora bien, el cuestionario debe estar muy bien confeccionado para que sea un instrumento válido y confiable. La validez consiste en que el instrumento mida lo que se supone que debe medir. Por ejemplo, si una empresa desea saber si sus empleados han llegado a convertir los valores organizacionales en parte integral

de su vida (al menos en la empresa), entonces podría aplicar un cuestionario para preguntarles al respecto. Pero qué pasaría si el cuestionario se limita a preguntar: "¿Conoce usted los valores organizacionales?". En este caso la pregunta se refiere a si conoce los valores organizacionales, no si los "vive" o los aplica en su trabajo. Este no sería un cuestionario válido, pues no está midiendo lo que se espera que mida. Por otro lado, además de válido, el instrumento debe ser confiable. La confiabilidad consiste en que el instrumento mida bien. Suponiendo que la empresa anterior solo deseé saber si sus empleados conocen los valores organizacionales, entonces la pregunta anterior tampoco sería confiable, pues muy probablemente los empleados respondan que sí los conocen, aunque no sea cierto.

Para lograr que un cuestionario sea válido y confiable debe confeccionarse muy bien su estructura y su contenido, es decir, el orden de las preguntas (generalmente las más sencillas primero), los temas o contenidos a que hacen referencia, el vocabulario que emplean (preferiblemente palabras sencillas y claras) de modo que se eviten ambigüedades, que no sugieran las respuestas y que no influyan en los entrevistados para dar ciertas respuestas.

Otro detalle es la extensión del cuestionario, pues si es demasiado extenso las personas no van a querer contestarlo, por lo que el investigador debe elegir preguntar aquello que es lo más importante y necesario para su investigación. Finalmente, también es importante el formato de las preguntas. En este sentido existen preguntas abiertas y preguntas cerradas:

- **Pregunta cerrada:** Este tipo de pregunta limita la respuesta del entrevistado a un conjunto de opciones previamente establecidas por el investigador. Las opciones deben ser diseñadas de manera que no sugieran la respuesta. Estas preguntas pueden tener solo dos opciones, como las preguntas de Sí o No, o bien, varias opciones, e incluso ser de múltiple escogencia. Una ventaja de este tipo de preguntas es que es más fácil tabular los resultados.
- **Pregunta abierta:** Este tipo de pregunta no limita la respuesta a un conjunto de opciones definidas por el investigador, por lo cual dan la posibilidad de obtener datos que no se hayan previsto. Pero su desventaja es que es más difícil tabular las respuestas y que a la hora de interpretar las respuestas podrían aparecer algunos sesgos.

Aplicación

Estudios de intención de voto

Con mucha frecuencia los medios de comunicación publican los resultados de las distintas encuestas de intención de voto que se realizan durante los períodos electorales. Generalmente estos estudios aplican cuestionarios a una muestra de electores. La metodología es en realidad la misma que se utilizaría en un estudio de mercado o cualquier otra investigación social, pues al igual que en cualquier otro estudio, debe definirse la población y la unidad estadística para el

análisis, determinar las variables que se requiere conocer, la técnica de encuesta (personal, por teléfono, etc.), el tipo de muestreo y el tamaño de la muestra, y el tipo análisis estadístico a realizar.

Ahora bien, las encuestas sobre temas políticos muchas veces generan importantes polémicas, pues los partidos políticos o los candidatos no reconocen los resultados, especialmente cuando no les favorecen, e incluso algunos aseguran que se emplean para manipular la opinión pública e influir sobre los resultados de las elecciones.

En otras ocasiones los pronósticos efectuados por las encuestas se alejan de los verdaderos resultados de las elecciones, e incluso encuestas realizadas por distintas firmas de investigación se contradicen. Sin embargo, hay que ser cauteloso a la hora de juzgar los resultados de estos estudios, pues hay muchos factores que pueden influir en ellos:

- **La metodología del estudio:** Es importante conocer cómo se realizó el estudio, especialmente en los factores que pueden elevar el margen de error, como el tamaño de la muestra, o introducir posibles sesgos, como el tipo de cuestionario empleado, la aplicación del cuestionario, el momento en que se realizaron las encuestas, etc.
- **La naturaleza del estudio:** En las ciencias sociales los fenómenos que se estudian son de carácter inmaterial, o sea, no son observables, cambian rápidamente y están muy relacionados con muchos otros elementos sociales. Si bien, las encuestas son el principal instrumento empleado, hay muchos aspectos de estas variables que son difíciles de captar por medio de cuestionarios.

1.7 Escalas de medida

Se ha mencionado que las variables pueden ser cuantitativas o cualitativas, lo cual implica la necesidad de asignarles valores numéricos, y para tal fin se emplean distintos tipos de escalas, por ejemplo, la escala nominal.

ESCALA NOMINAL: Nombres o clasificaciones que se utilizan para datos cualitativos en categorías no ordenadas distintas y separadas.

Estas se emplean, por ejemplo, al hablar del sexo de las personas, donde se podría codificar como "1" si es masculino y "2" si es femenino. Pero puede observarse en este caso que ese 2 no es el doble del 1, ni tampoco tendría sentido realizar operaciones matemáticas con esos valores.

Esta escala nominal también puede usarse para procesar datos como: marcas, productos, zonas geográficas, etc.

Propiedades de la escala Nominal

- Solo se realiza conteo, no se da ningún tipo de medición. Por ejemplo, se cuenta cuántos hombres o cuántas mujeres fueron entrevistadas.
- Las categorías no poseen un orden natural, ya sea de mejor o peor, ni de mayor a menor, etc. Así, si "1" si es masculino y "2" si es femenino, no puede decirse que ese 2 sea mayor o mejor que el 1.
- Al emplearse en variables cualitativas, entonces no existe diferencia cuantitativa entre una categoría y otra.
- No se puede realizar operaciones matemáticas.

ESCALA ORDINAL: Clasifican las observaciones en categorías con un orden natural o un orden significativo.

Por otro lado, las mediciones de escala ordinal indican un orden de preferencia o una secuencia de datos. Por ejemplo, si se le preguntara a alguien evaluar un servicio como 1. Excelente, 2. Bueno, 3. Regular y 4. Malo. Puede observarse que 2 no es la mitad de 4 y que no tiene sentido realizar operaciones con estos números.

Propiedades de la escala Ordinal

- Solo se realiza conteo, no se da ningún tipo de medición. Por ejemplo, se cuenta cuántas personas responden "excelente", "bueno", etc.
- Las categorías poseen un orden natural, ya sea de mejor o peor, ni de mayor a menor, etc. En otras palabras, existe una relación de orden entre las categorías, es decir, relaciones se pueden expresar en términos algebraicos de desigualdades (mayor que o menor que). Así, si "1" si es excelente y "2" si es bueno, puede decirse que ese 1 es mayor o mejor que el 2.
- Al emplearse en variables cualitativas, entonces no existe diferencia cuantitativa entre una categoría y otra.
- No se puede realizar operaciones matemáticas.

Aplicación

Medición de la satisfacción del cliente

Para las empresas que desean mejorar cada día es fundamental poder medir qué tan satisfechos están sus clientes con sus productos o servicios. Incluso, la medición de la satisfacción del cliente se incluye en normas de gestión de la calidad, como ISO 9000.

Lo más común es que, con cierta periodicidad, se aplique un cuestionario en el cual el cliente expresa su satisfacción con respecto a diferentes aspectos del producto o del servicio. Por ejemplo, en el caso de un servicio se puede preguntar si en trato fue amable, si la persona que atendió estaba preparada, si las instalaciones de la empresa le parecen adecuadas, entre otros.

Existen diversas formas de medir la satisfacción del cliente, pero en el caso de los cuestionarios, se plantean una serie de afirmaciones (reactivos) con respecto a los cuales el cliente debe indicar su opinión (o reacción). Por ejemplo, se plantea la afirmación: "Al llegar a la caja fui atendido con amabilidad". Y entonces se presentan opciones en una escala ordinal, tales como: "Muy de acuerdo", "De acuerdo", "Ni de acuerdo ni en desacuerdo", "En desacuerdo" y "Muy en desacuerdo".

Cuando se procesa esta información es importante tener presente que esta es una escala ordinal, o sea, que es cualitativa y tiene un orden natural, y que por tanto tiene la limitación de que no se puede efectuar operaciones matemáticas con los datos recopilados, es decir, no tendría sentido calcular un promedio u otros resultados. A pesar de estas limitaciones, la escala ordinal es la más adecuada para este tipo de variable.

ESCALA DE INTERVALO: Medidas en una escala numérica en la cual el valor de cero es arbitrario pero la diferencia entre valores es importante.

Tal como la definición lo indica, en la escala de intervalo las diferencias de valores son importantes, pero el cero es un valor arbitrario. Un buen ejemplo es el caso de las temperaturas, las cuales son medidas en grados centígrados. Así una temperatura de 40° es mayor que una de 20° , pero eso no indica que 40° sea el doble de 20° , ya que la temperatura 0° es arbitraria y no indica cero calor o ausencia de temperatura.

Propiedades de la escala de Intervalo

- Al aplicarse en variables cuantitativas, entonces implica la cuantificación de los datos.
- Se emplean unidades de medida (como los grados Fahrenheit o centígrados) los cuales producen intervalos iguales entre puntos de la escala.
- Las mediciones en esta escala generan valores que indican diferencias entre las unidades (individuos u objetos) medidos.
- El punto cero y la unidad de medida son arbitrarios.
- Al ser una variable cuantitativa se pueden aplicar diferentes operaciones matemáticas, de modo que se pueden calcular las principales medidas estadísticas.

ESCALA DE RAZÓN Medidas numéricas en las cuales cero es un valor fijo en cualquier escala y la diferencia entre valores es importante.

Esta es la única medida en la que el cero es significativo, lo cual implica que los datos pueden ser manipulados empleando cualquier operación matemática. Tal es el caso de variables como tiempo, ingreso, participación de mercado, distancia, etc.

Propiedades de la escala de razón

- La distancia o diferencia entre los números es conocida y constante.
- Tiene un punto cero absoluto, no arbitrario.
- Dado su carácter cuantitativo y que posee un cero absoluto, entonces puede efectuarse cualquier tipo de cálculo y prueba estadística.
- Se pueden hacer comparaciones entre los números tanto por su diferencia como por su razón, o sea, si es válido decir que si una cuerda mide de longitud 40 cm, entonces mide el doble de otra cuya longitud es de 20 cm.

Resumen de las escalas de medición de las variables

Cualitativas		Cuantitativas	
Escala nominal	Escala ordinal	Escala de intervalo	Escala de razón
No tiene orden natural	Tiene orden natural	Tiene un cero arbitrario	Tiene un cero absoluto
Ejemplo: Género: 1. Masculino 2. Femenino	Ejemplo: Opinión sobre gobierno: 1. Bueno 2. Regular 3. Malo	Ejemplo: Temperatura: La temperatura cero grados no indica ausencia de temperatura	Ejemplo: Ingreso: Un ingreso de \$0 indica ausencia de ingreso

Ejercicio de revisión

Clasifique las siguientes variables según su nivel de medición (nominal, ordinal, de intervalo o de razón):

1. Marca de un refresco producido en el país.
2. Grado académico de un profesional.
3. Ingreso mensual familiar.
4. Número de hijos.
5. Talla de una camiseta (pequeña, mediana, grande).
6. Número de la talla de un pantalón (10, 12, etc.).
7. Tiempo de espera en una fila para recibir un servicio.
8. Ciudad de residencia.
9. Calidad de un producto (sin defectos, con defectos menores o con defectos mayores).
10. Peso de un paquete de harina.

Ver solución de este ejercicio en la página de internet de este texto:

www.auladeeconomia.com/raeep.html

Aplicación

Software estadístico

Existen en la actualidad una serie de distintas aplicaciones creadas especialmente para resolver problemas estadísticos. Sin duda estos programas son de gran utilidad, pues permiten el procesamiento y almacenamiento de grandes cantidades de información, facilitan la realización de procedimientos complejos, efectuar gran cantidad de operaciones en poco tiempo y ayudan a mejorar el aprendizaje y la comprensión de los conceptos estadísticos.

Algunos de los paquetes estadísticos más importantes en la actualidad son Minitab, SPSS, Statistica, Stata, Systat, SAS, R, e incluso Excel, que aunque no es una aplicación exclusivamente estadística, sí es sumamente útil en el manejo de datos y en la realización de muchos cálculos estadísticos. Incluso, existen algunos programas estadísticos que corren sobre Excel, como XLStat.

Existe software libre de calidad que también se puede emplear, como es el caso de R, Calc (OpenOffice), entre otros.

En este texto se van a describir procedimientos para Excel y Minitab, que son dos herramientas especialmente útiles para el aprendizaje de la estadística.

Apoyo audiovisual y uso de la tecnología

En la página de internet www.auladeeconomia.com podrá encontrar una presentación de diapositivas que expone este tema y es una parte importante de este texto. Esta presentación presenta el tema en forma visual, pues emplea fotografías, esquemas u otros recursos visuales.

Adicionalmente puede encontrar algunos videos explicativos.

Solicite su usuario y contraseña escribiendo al correo electrónico info@auladeeconomia.com. Para ello deberá indicar el número de su factura de compra de texto.

1.8 Ejercicios

Ejercicios de desarrollo

Resuelva los ejercicios que a continuación se formulan (las respuestas a los ejercicios se encuentran en la página de internet del texto: www.auladeeconomia.com/raeep.html).

1. A continuación se presenta una lista de investigaciones, en cada caso indique cuál considera usted que es la unidad de estudio, y también identifique dos posibles variables en cada caso:
- En un hospital se desea determinar el tiempo medio de espera de los pacientes en el servicio de consulta externa.
 - Se desea conocer la efectividad de un programa de prevención de enfermedades cardiovasculares aplicado en la población de la ciudad capital de julio a noviembre del año 2012.
 - Un laboratorio farmacéutico ha desarrollado un nuevo medicamento y desea evaluar la posible interacción con otras sustancias, para lo cual posee una muestra de ratas en las cuales va a efectuarse el experimento.
 - Un investigador está preocupado por el continuo uso de computadoras, televisores y otros aparatos y su efecto en la agudeza visual de los niños de 6 a 12 años. El estudio se realizará a nivel nacional.
 - Como consecuencia del calentamiento global se considera que cierto tipo de ranas han ido desapareciendo, por lo que se requiere conocer cuántas ranas habitan aun en las zonas montañosas del país.
 - Se requiere conocer qué porcentaje de la población del país alcanza ciertos niveles de peso, para saber cuánta es la prevalencia de la obesidad en el país durante el último año.
 - Se va a determinar el estado de las viviendas del país. El estudio se efectuará de enero a marzo del próximo año.
 - En un proceso de producción se requiere saber la proporción de artículos defectuosos, además de conocer las posibles causas de los defectos.
 - Se desea saber la resistencia de determinados componentes electrónicos ante ciertas circunstancias extremas (sobrecargas, calentamiento, etc.).
 - Se va a realizar un estudio sobre las medidas de seguridad informática que las empresas financieras implementan. Para esto se va a efectuar una encuesta a los gerentes de tecnologías de información de bancos, cooperativas, mutuales, etc.
 - Ante el lanzamiento al mercado de un nuevo sistema operativo se va a estimar la proporción de empresas que están dispuestas a implementar este nuevo sistema en el transcurso de los próximos doce meses.
 - Un inversionista está buscando nuevas opciones para colocar su capital, por lo que decide invertir en la Bolsa de Valores de Nueva York. Ha reunido información de distintas empresas para conocer cómo han variado los precios de sus acciones y los pagos de dividendos.
 - Una nueva empresa de telecomunicaciones está efectuando un estudio de mercado para conocer las preferencias de los consumidores nacionales en lo que respecta al uso del teléfono celular, uso de internet, etc.
 - El encargado de recursos humanos de una empresa va a estimar el nivel de satisfacción de los empleados de la empresa para la cual labora.

- o. En una auditoría se van a examinar las cuentas por cobrar de una empresa.
- p. Se requiere conocer cómo afectaría un nuevo paquete de impuestos a los diferentes sectores productivos del país. Para tal fin se efectuará una encuesta entre empresarios, líderes de organizaciones empresariales y expertos en el tema.
- q. El gobierno desea saber qué tan efectiva ha sido una campaña para promocionar al país como destino turístico. Para ello se aplicará un cuestionario a turistas que ingresan al país por los distintos aeropuertos y fronteras terrestres.
- r. Una educadora está investigando el impacto de las tecnologías de la información y la comunicación en los hábitos de estudio de los estudiantes de secundaria del país.
- s. Un psicólogo en una empresa aplica un test para identificar distintos tipos de personalidad de los candidatos a una serie de puestos vacantes en la empresa.
- t. Se está investigando el efecto que tienen las redes sociales en internet en las relaciones familiares, para esto se tomará una muestra de familias del área metropolitana durante los próximos 6 meses.
- u. Un investigador de una institución desea conocer el estado general (salud, emocional, etc.) de los niños dados en adopción durante el último año en el país.
- v. El gobierno está analizando el efecto que han tenido algunas manifestaciones estudiantiles en su imagen. Para esto va a realizar una encuesta a nivel de las zonas urbanas del país entre personas mayores de 18 años.
- w. Un periodista está preocupado porque considera que algunas acciones del gobierno afectan la libertad de prensa. Va a realizar una encuesta aplicada a una muestra representativa de sus colegas para valorar la percepción que tiene de esos mismos sucesos.
- x. Para determinar la efectividad de una campaña publicitaria, se va a evaluar la exposición mensual de los consumidores a los comerciales pautados y el volumen mensual de ventas de la compañía en el mismo período. Se tomará una muestra de consumidores del área metropolitana.

2. A continuación se da una lista de variables:

- a. Clasifique las siguientes variables como cualitativas o cuantitativas:
 - Género (sexo)
 - Estatura de una persona
 - Nota en el examen de estadística
 - Ciudad de residencia
 - Número de habitaciones de la casa
 - Número de la camiseta de un jugador

- b. Clasifique las siguientes variables cuantitativas como discretas o continuas:
- Temperatura del horno
 - Número de hijos
 - Número de materias matriculadas por un estudiante
 - Distancia de la casa al trabajo
- c. Con respecto a las siguientes variables, indique su nivel de medición (nominal, ordinal, de intervalo, de razón):
- Género (sexo)
 - Estatura de una persona
 - Nota en el examen de estadística
 - Ciudad de residencia
 - Número de habitaciones de la casa
 - Número de la camiseta de un jugador
 - Temperatura del horno
 - Distancia de la casa al trabajo
3. Un médico está interesado en la relación que tiene la tasa de mortalidad infantil con distintos aspectos de la atención prenatal.
- a. Indique cuáles podrían ser 5 variables de este estudio (clasifíquelas por su naturaleza y según su nivel de medición).
 - b. Señale dos posibles fuentes de información que podría utilizar.
4. El siguiente es un extracto de un estudio estadístico:
El fenómeno de la migración de trabajadores/as de la salud en Costa Rica debe entenderse en el contexto más amplio de los cambios que ha experimentado la economía mundial y no como una simple relación de oferta y demanda, concluye el estudio "Trabajadores de la salud y migraciones en Costa Rica". Los responsables de la investigación señalaron que, además de las razones individuales de las personas migrantes, "entre los factores que la explican la migración de los profesionales de la salud, se cuenta la nueva dinámica económica internacional y sus implicaciones en el campo de la salud, que han llevado a la adopción de las tendencias industriales en el servicio de salud, como son la incorporación de nueva tecnología, el ahorro de personal y el aumento de las cargas de trabajo dentro de una lógica de costo-beneficio, llevando al desplazamiento de personal, sobre todo de enfermería".

El estudio fue efectuado por la Organización Internacional del Trabajo (OIT) y publicado en http://dwt.oit.or.cr/index.php?option=com_content&view=article&id=470:estudio-de-oit-sobre-trabajadores-de-la-medicina-y-la-enfermeria&catid=198:noticias&Itemid=347

- Con relación a la información anterior, conteste:
- ¿Cuál es la unidad estadística del estudio?
 - En el estudio se aplicó un cuestionario a 1.200 profesionales. ¿Cuál pudo haber sido la principal razón para emplear una muestra en vez de la población? Explique.
 - Mencione una posible variable cuantitativa que se haya medido en el estudio.
 - Mencione una posible variable medida en una escala ordinal en el estudio.
5. El siguiente es un extracto del resumen del proyecto de investigación titulado "Estrés laboral en profesionales de Enfermería: Estudio sobre evaluación cognoscitiva y afrontamiento" elaborado por María de los Ángeles Meza-Benavides (Escuela de Enfermería de la Universidad de Costa Rica):

Se presentan los resultados de la investigación sobre estrés laboral en enfermeras, centrada en el establecimiento de la relación entre evaluación cognoscitiva y afrontamiento del estrés, de acuerdo con la teoría de Lazarus y Folkman. Además, se examina esta relación con las manifestaciones psicológicas y psicofisiológicas del estrés. Para dicho estudio, se elaboraron dos instrumentos de medición: uno explora la evaluación cognoscitiva de las situaciones de estrés en el trabajo de los y las enfermeras (os) y el otro obtiene información sobre varias estrategias de afrontamiento.

Con relación a la información anterior, conteste:

- ¿Cuál es la unidad estadística?
- ¿Cuál pudo haber sido la principal razón para emplear una muestra en vez de la población en este estudio? Explique con detalle
- Mencione una posible variable cuantitativa que se haya medido en el estudio.
- Mencione una posible variable medida en una escala nominal en el estudio.
- Una posible variable de interés que se mida en una escala de razón. Explique.
- Un posible sesgo que podría presentarse en este estudio. Explique.
- El método de recolección de datos a emplear. Justifique su elección.

6. La siguiente información fue publicada en el periódico La Nación el 18 de setiembre de 2012:

CCSS mostró proceso de control de calidad de medicamentos

Antes de que los medicamentos que le receten a usted en un Ebais u hospital público lleguer a sus manos, éstos pasan entre dos y 20 días de análisis de control de calidad para cerciorarse de que el medicamento no producirá efectos adversos mayores y además tendrá la eficacia esperada para tratar determinado dolor o enfermedad.

Funcionarios de la Caja Costarricense de Seguro Social (CCSS) mostraron esta mañana cuáles son los procedimientos con los que se evalúa los fármacos. De cada lote se toma una muestra representativa y se le pasa por varios tipos de análisis. Primero se les hace un estudio fisico en donde se ven diferencias en forma, color, composición y si tienen algún tipo de mancha, o están más duras o suaves de lo esperado.

También se les pone en lupas y máquinas especiales para determinar la cantidad de partículas y el tamaño de éstas. Si no se pasa este análisis el medicamento es rechazado y el lote es devuelto al proveedor. Si el medicamento aprueba esta fase, pasa a una fase de pruebas químicas.

Aquí, en varias máquinas se les hacen pruebas de disolución de químicos, de impurezas, un análisis de a nivel de átomos, otro a nivel de luz, y finalmente se coloca en un simulador del comportamiento del fármaco en el organismo humano y suponer cómo se diluye y absorbe en el cuerpo humano.

Fuente: <http://www.nacion.com/2012-09-18/AldeaGlobal/ccss-mostro-proceso-de-control-de-calidad-de-medicamentos.aspx>

Con base en esta lectura determine, con respecto al control de calidad de los medicamentos descrito, indique:

- a. ¿Cuál sería la unidad estadística?
 - b. Indique cuatro variables que satisfagan las siguientes condiciones:
 - Una variable cualitativa que se mida en escala ordinal.
 - Una variable cuantitativa discreta.
 - Una variable cualitativa que se mida en escala nominal.
 - Una variable continua que se mida en escala de razón.
 - c. ¿Cuál es el método de recolección de la información y cuál sería un instrumento de recolección de la información? Explique.
7. Una institución está interesada en conocer el efecto de distintos tipos de fármacos en pacientes geriátricos, dado que esta población ha experimentado un mayor crecimiento en años recientes. En la investigación se tomó una muestra de adultos mayores y se estudiaron variables como: edad (en años), peso (en kilogramos), estado nutricional (categorizado como muy adecuado, adecuado, inadecuado, o muy inadecuado), tipo de fármacos que emplea con regularidad (analgésicos, anticoagulantes, psicofármacos, etc.), masa corporal magra (en porcentaje), y si tiene algún padecimiento hepático (si o no). Clasifique todas las variables anteriores

como cualitativas o cuantitativas. Las variables cuantitativas clasifíquelas como discretas o continuas. Además, indique el nivel de medición de cada una de las variables.

8. Cada vez más personas recurren al quirófano para librarse de las gafas. Se desea realizar un estudio para determinar cuáles instituciones están ofreciendo ese servicio, qué técnicas emplean, cuáles son las calificaciones del personal que realiza estos servicios y cuáles son los costos que tienen para los pacientes. Identifique cuál es la unidad de estudio en esta situación y señale cuatro variables de interés.
9. Una empresa de desarrollos habitacionales está estudiando la posibilidad de construir un condominio en una zona urbana. Se sabe que es necesario conocer la disponibilidad de diversos servicios, como el agua, la electricidad, etc. Indique cuáles podrían ser 4 variables de interés para un estudio de este tipo. Clasifique esas variables e indique sus niveles de medición.
10. Una empresa ha estado recibiendo quejas y devoluciones de sus clientes, dado que el producto tenía algún defecto de fabricación. El ingeniero a cargo del proceso de producción decide que debe realizar algunas mejoras y que debe hacer un estudio estadístico para obtener información al respecto. ¿Cuál es la unidad estadística en este caso? ¿Es la población finita o infinita? ¿Cuáles podrían ser tres variables del estudio (clasifíquelas según su naturaleza)? ¿Deberá emplearse la población o una muestra para realizar el estudio?
11. Una empresa electrónica toma una muestra de cuatro tipos de componentes electrónicos para determinar su tiempo de funcionamiento hasta que fallen. Con respecto a este problema, ¿cuál es la unidad de estudio? Señale dos variables de interés y clasifíquelas como cualitativas o cuantitativas.
12. La presencia de manganeso en la gasolina puede significar un problema, pues genera ciertos daños a los vehículos, a la salud de las personas y al ambiente. Ante la duda de altas concentraciones de esta sustancia en la gasolina vendida en el país se toman muestras en los distintos planteles de la compañía refinadora de petróleo para que vaya a ser estudiadas en un laboratorio. Cuatro personas discuten sobre esta situación e indican lo siguiente:
 - Luis: "En este problema específico del manganeso la unidad de estudio es la compañía petrolera y la variable de estudio serían sus prácticas ambientales".
 - Ana: "En mi opinión, la unidad de estudio sería cada plantel, y la variable sería la concentración de manganeso en cada plantel".
 - Marcela: "Yo considero que lo que interesa estudiar es la gasolina, y la variable es la ausencia o presencia de manganeso en cada caso".
 - Roberto: "La unidad de estudio es el manganeso, y la variable es la concentración de manganeso en la gasolina".

En cada caso analice la afirmación de cada persona e indique en cuáles aspectos tiene la razón y en cuáles no.

13. Una institución financiera desea conocer con detalle el comportamiento de sus clientes actuales (que han realizado al menos una transacción al mes durante los últimos 12 meses), y entonces se va a realizar un estudio estadístico que permita obtener datos sobre aspectos como sus hábitos de consumo y ahorro, los servicios financieros que más emplean, utilización de servicios de la competencia y qué tan satisfechos se sienten con la oferta actual de la compañía, para así contar con la información necesaria que les permita lograr un crecimiento importante en el mercado. Los aspectos anteriores desea clasificarlos según el sexo del cliente, su nivel de ingresos, la edad, el estado civil, su grado académico, la actividad a la que se dedica y el volumen de operaciones que maneja en la institución.

Con base en esta información defina:

- a. ¿Cuál es la unidad estadística?
 - b. ¿Cuál es la población?
 - c. Con respecto a las variables bajo estudio, señale:
 - Dos variables cualitativas
 - Dos variables cuantitativas
 - Una variable cuantitativa discreta
 - Dos variables cuantitativas continuas
 - Dos variables cualitativas que se midan en escala nominal
 - Dos variables cualitativas que se midan en escala ordinal
 - Dos variables cuantitativas que se midan en escala de razón
 - d. ¿Qué sería más conveniente, realizar el estudio entre toda la población o emplear una muestra? Explique.
14. En una empresa se desea efectuar un estudio de clima organizacional. El departamento de recursos humanos está a cargo y debe definir aspectos como la población bajo estudio, el tipo de instrumento a utilizar, si se debe emplear una muestra o no, y las variables que se van a incluir en el estudio (indique al menos 6 variables, clasifíquelas e indique el nivel de medición de cada una).
15. Una empresa productora de jabones para el hogar está estudiando la posibilidad de lanzar al mercado un jabón especial para niños. Desea efectuar un estudio estadístico para determinar si el citado producto tendrá la aceptación suficiente. Con base en esta información responda lo siguiente:
- a. ¿Cuál es la unidad estadística de la investigación que se llevará a cabo?
 - b. Cite tres características de la unidad estadística citada, de modo que una de ellas sea una variable continua, otra una variable discreta y otra un atributo.
 - c. Dé un ejemplo de una posible observación para cada una de las características citadas.
 - d. Mencione y explique brevemente dos de las fases del estudio estadístico que será necesario realizar en este caso.
 - e. Describa un método que podría emplear la empresa para recolectar la información.

- f. Explique las ventajas y desventajas del método citado en el inciso e.
- g. Un gerente de la empresa propone que el muestreo debería hacerse en un barrio cercano a la fábrica, ya que esas personas conocen el producto, conocen la empresa y además se minimizarían los costos. ¿Cuál es su opinión al respecto: en favor o en desacuerdo? Explique por qué. Si está en desacuerdo ¿Qué tipo de muestreo propondría y por qué?
16. Un contador está realizando una auditoría. Necesita analizar los cheques emitidos por una compañía para verificar la exactitud de sus registros contables. ¿Cuál es la unidad estadística en este caso? ¿Cuáles podrían ser cuatro variables del estudio (clasifíquelas según su naturaleza e indique su nivel de medición)? ¿Deberá emplearse la población o una muestra para realizar el estudio?
17. Un economista está tratando de determinar la relación que existe entre el ingreso familiar y los gastos en alimentación de las familias del país. Determine qué tipo de muestreo podría emplearse (aleatorio o no aleatorio), e indique las ventajas y desventajas que podría tener cada tipo de muestreo. Además, señale cuáles podrían ser algunas fuentes de información que se podrían emplear. Indique si son fuentes primarias o secundarias.
18. La cámara de empresas turísticas local está efectuando un estudio sobre las expectativas de estos empresarios. Se aplica un cuestionario entre una muestra de empresas relacionadas con el turismo en el país. En el cuestionario se preguntaron a los gerentes de las empresas aspectos como el grado de satisfacción con la rentabilidad de su negocio, sus planes de contratación de personal para los próximos meses, su percepción de la demanda del servicio que presta, el tipo de negocio (hotel, restaurante, operador de tours, etc.). a. Una persona comenta que en este estudio la unidad estadística es el gerente de cada empresa, pero otra persona le contradice y afirma que la unidad estadística es la empresa. ¿Cuál de los dos cree usted que tiene la razón? b. Anteriormente se mencionaron 4 variables del estudio. Clasifíquelas como cualitativas o cuantitativas e indique su escala de medición.
19. Una educadora desea conocer la efectividad de un nuevo método de enseñanza de la lecto-escritura. Decide que debe realizar un estudio estadístico para comparar un método tradicional con el método nuevo. ¿Cuál es la unidad estadística en este caso? ¿Cuáles podrían ser cuatro variables del estudio (clasifíquelas según su naturaleza e indique su nivel de medición)? ¿Deberá emplearse la población o una muestra para realizar el estudio? ¿Cuál método debería emplear para recopilar la información? ¿Cuál podría ser un sesgo que se cometía en este investigación y qué debería hacer la educadora para prevenirlo?

20. El fenómeno de la drogadicción entre estudiantes de secundaria es un problema grave. Un psicólogo está interesado en conocer el papel que juegan las familias, por lo que desea investigar más sobre el asunto, por lo que desea conocer cómo podría relacionarse este problema con aspectos específicos como si la madre trabaja, si los jóvenes viven tanto con su padre como su madre, o solo uno de ellos u otro familiar, el nivel de escolaridad de los padres, la cantidad de tiempo que los padres dedican a atender a sus hijos, el número de miembros del hogar y el ingreso neto total del hogar. Clasifique las características anteriores como cualitativas o cuantitativas. Las que sean cuantitativas indique si son discretas o continuas.
21. Según un estudio realizado por la Escuela de Negocios de la Universidad de Edimburgo, en el Reino Unido, conforme más amigos tiene una persona en la red social Facebook, mayor podría ser el nivel de estrés al que se vea sometido. Los investigadores realizaron el estudio en más de 300 personas que utilizan esta red social a diario. Uno de los hallazgos fue que los niveles de ansiedad en las personas aumentan cuando su jefe o sus padres les envían una solicitud de amistad, principalmente cuando el usuario presenta en Facebook una imagen de sí mismo que podría no ser aceptable para algunos de sus esos amigos, como cuando dicen malas palabras o publican fotografías fumando, tomando licor, etc. Con base en esta situación identifique: a. La unidad estadística en el estudio. b. Tres variables del estudio. c. Clasifique las variables como cualitativas o cuantitativas. d. Indique el nivel de medición de las variables. e. ¿Por qué es necesario emplear una muestra en un estudio de este tipo? f. ¿Cuál es el instrumento de medición empleado?
22. Una trabajadora social requiere analizar las condiciones en que se encuentran las personas que están albergadas en hogares de ancianos de la ciudad capital. Para formular su investigación debe determinar cuál es la unidad estadística y su población, definir las variables de su estudio (una variable cualitativa que se mida en escala nominal, una variable cualitativa que se mida en escala ordinal, una variable cuantitativa discreta y una variable cuantitativa continua), y, finalmente, decidir si realiza su estudio visitando toda la población o si emplea una muestra.
23. Un politólogo está interesado en conocer más sobre el fenómeno del abstencionismo en las elecciones presidenciales del país. Desea realizar un estudio por muestreo en el que se pregunte a las personas su edad (en años), la zona en que reside (urbana o rural), si votó en la última elección (si o no), la razón por la cual votó o por la que no votó (se da una lista de opciones en cada caso), su nivel de escolaridad (primaria incompleta, primaria completa, secundaria incompleta, etc.). Clasifique las variables anteriores según su naturaleza e indique el nivel de medición de cada una.
24. Una empresa desea monitorear la publicidad que están haciendo las empresas de telecomunicaciones del país. ¿Cuáles podrían ser cuatro variables del estudio (clasifíquelas según su naturaleza e indique el nivel de medición de cada una)? ¿Cuál método debería emplear para recopilar la información? ¿Cuál podría ser un sesgo que se cometía en este investigación y qué debería hacer la educadora para prevenirlo?

25. Una compañía está preocupada por su imagen ante clientes, proveedores, acreedores y la comunidad donde opera la empresa. Antes de implementar algún tipo de medida en este sentido desea conocer más información sobre la imagen que tiene ante cada uno de estos grupos. En una reunión en que se espera definir detalles sobre la investigación que se debe realizar, un ejecutivo expresa que el estudio tendrá una sola población y que estará compuesta por todas aquellas personas que sean sus clientes, miembros de alguna empresa proveedora o acreedora, o que resida en la comunidad aledaña a la empresa. Otro ejecutivo considera que en realidad son 4 poblaciones distintas, y que no todas están compuestas por personas. ¿Cuál de los dos considera usted que tiene razón?

Examen del capítulo:

En cada caso seleccione la opción que mejor contesta cada pregunta (las respuestas a los ejercicios se encuentran en la página de internet del texto: www.auladeeconomia.com/raeep.html):

1. La estadística es un conjunto de _____ aplicadas a la recolección, descripción y análisis de datos, los cuales constituyen evidencia numérica para la toma de decisiones en condiciones de _____. La opción que mejor completa la frase anterior:
 - (a) sistemas; incertidumbre
 - (b) métodos y teorías; incertidumbre
 - (c) métodos y teorías; certidumbre
 - (d) métodos y teorías; riesgo
2. Una _____ es una parte representativa de la población que se selecciona para ser estudiada ya que la población es demasiado grande para ser estudiada en su totalidad. La opción que mejor completa la frase anterior es:
 - (a) Característica
 - (b) Muestra
 - (c) Observación
 - (d) Población
3. Considere el siguiente concepto: "unidad de interés en el campo bajo estudio, sobre la cual recae la observación y de la cual se derivan los datos para el análisis". Esto corresponde al concepto de:
 - (a) Unidad estadística
 - (b) Característica
 - (c) Muestra
 - (d) Población
4. De las siguientes, no es una razón para trabajar con muestras en vez de la población es:
 - (a) La población se destruye al estudiarla
 - (b) El costo de estudiar la población es muy alto
 - (c) La población es muy grande
 - (d) Ninguna de las anteriores

5. Con respecto a la variable "estatura" es falso que:
- (a) Se mide en una escala de razón
 - (b) Es una variable cuantitativa discreta, pues la gente siempre la da como un número entero
 - (c) No se puede medir en una escala ordinal, o sea, como grande, mediano, pequeño
 - (d) No es una característica de la unidad estadística
6. Un ingeniero debe estimar si las varillas de construcción que la compañía ha comprado satisfacen los requerimientos establecidos en cuanto al diámetro de las mismas. Para ello se formula lo siguiente:
- A. La unidad estadística es el diámetro de las varillas, pues es lo que le interesa saber.
 - B. Dado que se han comprado miles de varillas, lo mejor será tomar una muestra de al menos la mitad de las varillas para tener una muestra representativa.

De las anteriores, con toda certeza, son correctas:

- (a) Ambas afirmaciones
- (b) Solo la afirmación A
- (c) Solo la afirmación B
- (d) Ninguna de las afirmaciones

7. Considere las dos siguientes afirmaciones:

- A. En algunos casos es necesario emplear una muestra porque la población se destruiría al estudiarla.
- B. La principal razón para estudiar una muestra en vez de la población es reducir los costos.

Con respecto a las dos afirmaciones anteriores, es correcto que:

- (a) Ambas son verdaderas
- (b) Solo A es verdadera
- (c) Solo B es verdadera
- (d) Ambas son falsas

8. Un ingeniero requiere determinar si los tiempos que duran los operarios en realizar una actividad se ajustan a los parámetros establecidos por la compañía. Para ello se formula lo siguiente:

- A. La unidad estadística es el tiempo promedio, pues es lo que le interesa saber.
- B. Dado que se han contratado cientos de operarios, lo mejor será tomar una muestra de al menos el 80% de los operarios para tener una muestra representativa.

De las anteriores, son correctas con toda certeza:

- (a) Ambas afirmaciones
- (b) Solo la afirmación A
- (c) Solo la afirmación B
- (d) Ninguna de las afirmaciones

9. Un ingeniero requiere determinar si los tiempos que duran los operarios en realizar una actividad se ajustan a los parámetros establecidos por la compañía. La característica "tiempo de realización de la actividad" es una variable que se mide en una escala:

- (a) De razón (b) Ordinal (c) De intervalo (d) Nominal

10. Un ingeniero requiere determinar si los tiempos que duran los operarios en realizar una actividad se ajustan a los parámetros establecidos por la compañía. Si la característica "grado académico del operario" se evalúa como "Primaria incompleta, primaria completa, secundaria incompleta, secundaria completa", entonces la variable se mide en una escala:

- (a) De razón (b) Ordinal (c) De intervalo (d) Nominal

11. El gerente de un centro de llamadas desea evaluar el desempeño del sistema y para ello decide basarse en los tiempos de espera de los clientes para ser atendidos (medido en segundos) y el grado de satisfacción que los clientes manifiesten al recibir el servicio (valorado como bueno, regular o malo). La semana anterior tomó una muestra de 12 llamadas por día de lunes a miércoles. Los siguientes son los tiempos de las muestras tomadas de lunes a miércoles:

Número de muestra (tiempo en segundos)												
1	2	3	4	5	6	7	8	9	10	11	12	
13	15	15	16	16	16	16	17	17	19	11	21	
12	15	15	16	16	15	16	18	15	20	30	40	
30	40	45	45	50	50	45	35	60	50	80	100	

Las preguntas de la 11 a la 17 se basan en la información anterior.

Una variable medida en escala nominal puede ser:

- (a) Tiempo de espera (b) Nombre del agente de servicio que atendió
(c) Grado de satisfacción del cliente (d) Número de llamadas hechas por el cliente

12. Con base en la información de la pregunta 11, una variable medida en escala ordinal puede ser:

- (a) Tiempo de espera (b) Nombre del agente de servicio que atendió
(c) Grado de satisfacción del cliente (d) Número de llamadas hechas por el cliente

Con respecto a las dos afirmaciones anteriores, es correcto con toda certeza que:

- (a) Ambas son verdaderas (b) Solo A es verdadera
(c) Solo B es verdadera (d) Ambas son falsas

este capítulo se analizan las principales formas de presentar la información estadística en un formato textual. Se describen los cuadros estadísticos y los gráficos más utilizados para representar datos.

CAPÍTULO 2

PRESENTACIÓN DE LA INFORMACIÓN ESTADÍSTICA

OBJETIVOS

Al concluir el capítulo, será capaz de:

- Identificar las formas principales de presentar la información estadística.
- Presentar apropiadamente la información estadística en un formato textual.
- Elaborar cuadros estadísticos.
- Construir gráficos adecuados según el tipo de datos.

Este capítulo se divide en tres secciones principales: 1) Introducción a la presentación de la información estadística, 2) Presentación de la información estadística en formato textual, y 3) Presentación de la información estadística en forma visual.

La introducción a la presentación de la información estadística se centra en definir qué es la estadística y qué tipos de datos se manejan. Se explica que la estadística es una disciplina que se ocupa de la recolección, organización, análisis y interpretación de datos. Los datos se clasifican en cuantitativos y cualitativos. Los cuantitativos son datos numéricos que se miden o contabilizan, mientras que los cualitativos son datos que describen características, propiedades o atributos de los individuos o unidades de observación. Se menciona que la estadística es una ciencia empírica que se basa en la observación y la experimentación. Los datos se recopilan a través de encuestas, encuestas, observaciones y experimentos. La estadística se aplica en numerosas disciplinas, como la economía, la medicina, la ingeniería, la psicología y la sociología.

2.1 Presentación de los datos

En una investigación estadística, una vez que se tiene la información reco-
pilada, por ejemplo cuestionarios ya contestados, es necesario ordenar los datos y
clasificarlos adecuadamente (lo que se conoce como procesamiento de la informa-
ción) para luego presentar estos datos en forma resumida, lo cual puede hacerse
en forma textual (o sea, dentro de un texto) o empleando cuadros y gráficos.

Presentación de los datos

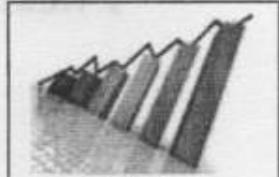
Formas de presentación de la información



TEXTUAL:
Los datos van descritos
dentro del texto.



TABULAR:
Los datos se presentan
usando tablas o cuadros.



GRÁFICA
Los datos se presentan
usando gráficas.

Material audiovisual

En la página de internet de este texto podrá en-
contrar una presentación de diapositivas que ex-
pone este tema y es un complemento a este texto.



**Presentación
de los datos**
¿Cómo se pueden
presentar los datos
adecuadamente?

Solicite su usuario y contraseña escribiendo al co-
rreo electrónico info@auladeeconomia.com.

Para ello deberá indicar el número de su factura
de compra del texto.

2.2 Presentación textual

En este método de presentación de la información se emplean pa-
labras y cifras combinadas en el texto, con el fin de informar los datos obtenidos.
Observemos el siguiente ejemplo:

De acuerdo con la encuesta aplicada, el 30% de los entrevistados estudia
en una universidad privada, mientras que el restante 70% estudia en una
universidad pública. A su vez, el 40% de estos estudiantes de universida-
des privadas estudia alguna carrera del área de los negocios, un 30% una
carrera de ingeniería, un 15% estudian carreras de ciencias sociales, un
10% carreras de educación, mientras que solo el 5% de los estudiantes de
universidades privadas entrevistados estudia una carrera relacionada con
las ciencias de la salud.

Este método, desde la perspectiva del lector, puede ser más "engoroso", pues la presentación de muchos datos puede ser muy extensa. Para el redactor, este método tiene la ventaja de que puede explicar los datos y las conclusiones a la que llega, y se emplea principalmente cuando la cantidad de datos es poca.

2.3 Cuadros estadísticos

Características de los cuadros

La presentación tabular, o sea, en forma de cuadros o tablas, se utiliza principalmente cuando la cantidad de datos que se quiere presentar es grande. Por ejemplo, si se tienen los datos de población de todos los países de América Latina para tres años distintos. Esta gran cantidad de datos presentados de forma textual producirían un texto extenso y poco agradable, mientras que a través de un cuadro va a ser fácil comparar unos países con otros o comparar unos años con otros.

CUADRO ESTADÍSTICO: Es una forma de presentación de la información estadística que utiliza una filas y columnas en las que se incluyen los datos que se desea presentar.

Al construir las tablas es posible dar énfasis a ciertas cifras y analizar con detalle los componentes de un total, por lo que los cuadros son sumamente útiles y muy utilizados en los informes estadísticos.

2.4 Construcción de cuadros

Las partes esenciales de un cuadro o tabla bien diseñado son:

- **Título:** se ubica en la parte superior central del cuadro, debe estar redactado con claridad, y responder a cuatro preguntas básicas:
 - ¿Dónde? Indicar el país, región, empresa, institución, etc. donde sucedió el evento con el que se relacionan los datos del cuadro.
 - ¿Qué? Indicar cuál variable o característica es a la que corresponden los datos del cuadro.
 - ¿Cuándo? Esto es mostrar la fecha o periodo al que corresponden las cifras del cuadro.
 - ¿Cómo? Debe indicarse con claridad la forma en que vienen expresados los datos, por ejemplo si son cifras en unidades monetarias, en millones, si son cifras relativas, etc. Para esto a veces se emplea una nota introductoria, generalmente entre paréntesis, para indicar estos aspectos que son importantes para comprender adecuadamente los datos mostrados.

- **Columna matriz:** esta es la primera columna a la izquierda del cuadro. En esta columna se indica cuáles son las partidas o componentes presentados horizontalmente en el cuadro, por ejemplo, países, empresas, productos, etc. Dependiendo de tipo de dato la columna matriz va a tener un orden particular, por ejemplo:
 - Alfabético: Se usa, por ejemplo, para ordenar países.
 - Cronológico: Cuando la columna matriz contiene años, meses, etc.
 - Por magnitud: Ciertos datos se ordenan de mayor a menor, o viceversa. Por ejemplo, cuando se muestra la cantidad de personas según su estado civil.
 - Numérico: Cuando los datos de la columna matriz corresponden a una variable cuantitativa, como pesos, distancias, etc.
 - Orden usual: Algunas variables tienen un orden "acostumbrado". Por ejemplo, generalmente las provincias o estados de un país tienen un orden usual.
- **Encabezados:** muestran cómo se presentan verticalmente los datos en el cuadro. Por ejemplo, por año, por sexo, etc. En ocasiones puede ser compuestos, por ejemplo, si los datos aparecen por sexo, y a su vez en términos absolutos y relativos.
- **Cuerpo del cuadro:** es propiamente donde aparecen los datos del cuadro.
- **Notas al pie:** son opcionales y se usan cuando es necesario aclarar algún dato. Por ejemplo, cuando una cifra es una estimación. Generalmente tiene un tamaño de letra más pequeño que el resto del cuadro.
- **Fuente:** cuando los datos del cuadro no han sido elaborados por el investigador, entonces, por honestidad intelectual, debe indicarse de dónde procede la información. En este sentido se debe indicar: nombre del autor original de los datos, año de publicación, nombre de la obra, editorial, país, páginas donde se encuentran los datos. Generalmente tiene un tamaño de letra más pequeño que el resto del cuadro.

Número de cuadro

**Título del cuadro: ¿dónde? ¿qué?
¿cuándo? ¿cómo?**

Encabezados

Columna matriz

Cuerpo del cuadro

Notas al pie:

Fuente:

Un detalle importante es que siempre debe indicarse la unidad de medida en que están expresadas las cifras, ya sea como nota introductoria en el título, en el encabezado de las columnas o en la clasificación correspondiente en la columna matriz. Igualmente, para facilitar el análisis pueden incluirse en el cuerpo del cuadro índices, porcentajes, promedios u otros datos útiles.

Además normalmente no se emplea el rayado vertical, salvo que sea necesario para separar las columnas. El rayado horizontal solo se usa en los encabezados y en la última fila, o en la fila de totales. El cuadro debe tener proporciones adecuadas, o sea, que no sea excesivamente ancho o demasiado alto.

Ejemplos:

Cuadro #1
Centroamérica: matrícula inicial en los primeros niveles,
según nivel educativo, por países.
1997
(Datos en miles)

Nivel	País	Preescolar	Primaria	Secundaria	Total
	Costa Rica	71.1	525.3	220.2	816.5
	El Salvador	179.0	1191.1	152.5	1522.6
	Guatemala	221.8	1539.5	317.4	2078.6
	Honduras	86.1	1055.0	271.1	1412.1
	Nicaragua	133.1	777.9	270.5	1811.5
	Panamá	45.3	377.9	223.2	646.4
	Total	736	5466.6	1451.7	7657.7

Fuente: Informe Estado de la Región en Desarrollo Humano Sostenible. Proyecto Estado de la Región. San José, C.R.: 1999, pág. 169.

Cuadro #2
MERCOSUR: Tasa de desempleo abierto trimestral, 2007^{1/}
(Como porcentaje de la PEA)

Países	Trimestre 1	Trimestre 2	Trimestre 3	Trimestre 4
Argentina ^{2/}	9.8	8.5	8.1	7.5
Brasil ^{3/}	9.8	10.0	9.3	8.1
Uruguay ^{2/}	10.4	10.1	9.4	8.6
Venezuela ^{4/}	10.3	8.4	8.5	6.7

Notas: ^{1/}No se obtuvo el dato para Paraguay. ^{2/}Tasa de desempleo abierto en zonas urbanas. ^{3/}Incluye solo seis áreas metropolitanas. ^{4/}Total nacional.

Fuente: CEPAL. CEPALSTAT – Bases de datos. Recuperado el 2 de octubre de 2012 de <http://web-sie.eclac.cl/sisgen/ConsultaIntegrada.asp?idAplicacion=6&idTema=359&idioma=e>

Ejercicio de revisión

Con base en los datos proporcionados elabore un cuadro estadístico completo que incluya todas las partes de un cuadro:

Consideré la siguiente información que se obtuvo del estudio "Comportamiento Clínico y Epidemiológico de las Infecciones Nosocomiales en la Unidad de Cuidados Intensivos Neonatales del Hospital Dr. Oscar Danilo Rosales Argüello".

"Al momento de su egreso fallecieron 61.4% de los recién nacidos. De ellos el 67.4% fue por enterobacter y 16.2 por pseudomonas aeruginosa. Un 31.4% de los recién nacidos fue dado de alta y abandonaron el centro hospitalario un 7.1% en muy malas condiciones". El proyecto fue elaborado por la Dra. Juana María Membreño Sequeira en el período comprendido de octubre 2002 a enero 2004 y fue publicado en <http://www.minsa.gob.ni/enfermeria/PDF/327.pdf> en marzo de 2004.

Ver solución de este ejercicio en la página de internet de este texto:
www.auladeeconomia.com/raeep.html

2.5 Representación gráfica

Características de los gráficos

Los gráficos son de las principales formas de presentar información estadística, ya que tienden a ser más atractivos y fáciles de comprender.

GRÁFICA: Conjunto de figuras geométricas que se emplean para representar un conjunto de datos estadísticos.

La principal ventaja de los gráficos es que permiten llamar la atención más que las otras formas de presentación de datos, además de que tienden a ser más fáciles y rápidos de comprender que un cuadro y facilitan las comparaciones. La principal desventaja es que no se puede presentar tanto detalle en la información como si se puede hacer en un cuadro.

2.6 Construcción de gráficos

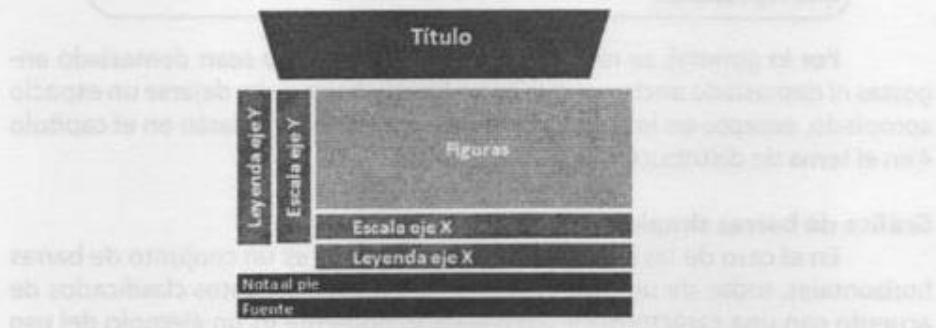
Un gráfico de cualquier tipo tiene las siguientes partes, y éstas siguen las mismas consideraciones hechas para el caso de los cuadros:

- **Título:** se ubica en la parte superior central del gráfico, debe estar redactado con claridad, señalar qué datos se presentan en el cuadro, el lugar donde ocurrió el evento al que se refieren los datos, la fecha o periodo al que corresponden y la forma en que vienen expresados los mismos. Muchas

veces el título incluye una nota introductoria, para explicar algún aspecto importante relacionado con las cifras presentadas. En resumen se puede decir que un buen título de un gráfico debe responder a las preguntas: ¿qué?, ¿dónde?, ¿cuándo? Y ¿cómo?, con su respectiva nota introductoria si es necesario.

- **Escalas y leyendas:** las leyendas indican el concepto o variable que se representa en cada coordenada y las unidades de medida empleadas.
- **Las figuras geométricas,** las cuales representan los datos.
- **Notas al pie:** es opcional y se emplea cuando es necesario aclarar algunas cifras del cuadro.
- **Fuente:** cuando el autor del gráfico no sea el productor original de los datos presentados en la gráfica.

Partes principales en un gráfico cualquiera



Apoyo audiovisual y uso de la tecnología

En la página de internet www.auladeeconomia.com/raeep.html podrá encontrar una presentación con cuatro pasos esenciales para construir un gráfico. Adicionalmente puede encontrar algunos videos explicativos sobre la construcción de gráficos en Excel y otros programas como Minitab.

¿Cuáles son las claves para construir un buen gráfico?

Vamos a ver un proceso sencillo de CUATRO etapas



Solicite su usuario y contraseña escribiendo al correo electrónico info@auladeeconomia.com. Para ello deberá indicar el número de su factura de compra de texto.

2.7 Clases de gráficos

Gráficos de barras

Este tipo de gráficos se emplean principalmente para hacer comparaciones, ya sea entre distintas partidas, como países, productos, etc., o a lo largo del tiempo. Se usan barras verticales para representar series de tiempo (series cronológicas) o datos cuantitativos (en el caso de distribuciones de frecuencia). Las barras horizontales se emplean cuando los datos están clasificados de acuerdo con una característica cualitativa, como lo son datos según país, según servicio o producto, departamento, etc.

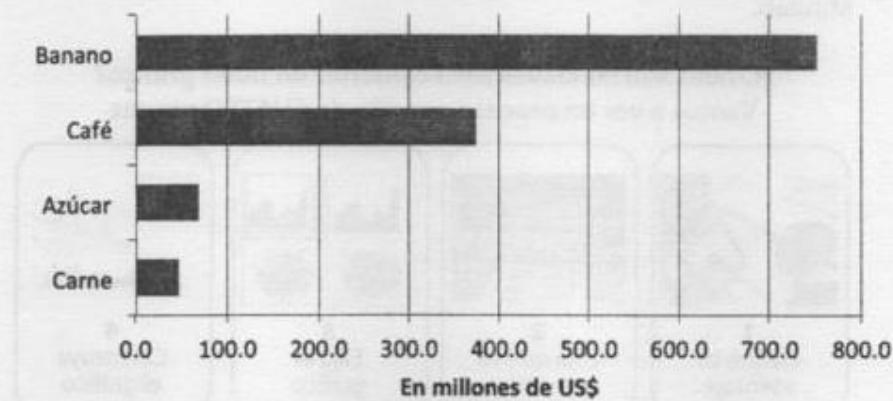
GRÁFICO DE BARRAS: Representación gráfica en que emplea un conjunto de rectángulos colocados paralelamente de forma que la extensión de cada uno es proporcional a la magnitud que representa.

Por lo general, se recomienda que las barras no sean demasiado angostas ni demasiado anchas, y que entre barra y barra debe dejarse un espacio apropiado, excepto en los histogramas, los cuales se estudiarán en el capítulo 4 en el tema de distribuciones de frecuencias.

Gráfico de barras simples

En el caso de las barras horizontales simples es un conjunto de barras horizontales, todas de un mismo ancho, que presentan datos clasificados de acuerdo con una característica cualitativa. El siguiente es un ejemplo del uso de las barras horizontales en datos clasificados geográficamente:

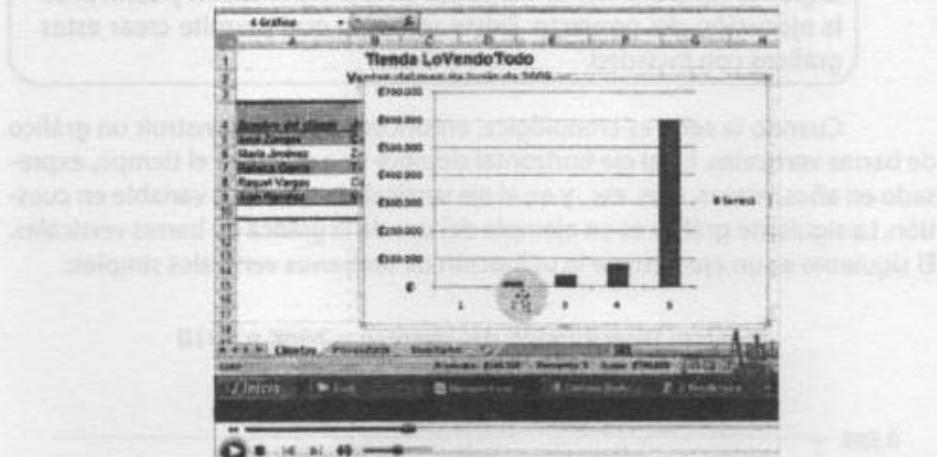
**Costa Rica: Valor de las exportaciones tradicionales por tipo de producto, 2011
(Valor FOB en millones de US\$)**



Fuente: Banco Central de Costa Rica (2012). Exportaciones FOB por tipo de producto. Recuperado el 12 de octubre de 2012, <http://indicadoreseconomicos.bccr.fi.cr/indicadoreseconomicos/Cuadros/frmVerCatCuadro.aspx?CodCuadro=479&Idioma=1&FecInicial=1997/01/31&FecFinal=2012/12/31&Filtro=0&Jump=1>

Material audiovisual

En la página de internet de este texto podrá encontrar videos que exponen cómo crear gráficos en Excel y en Minitab.



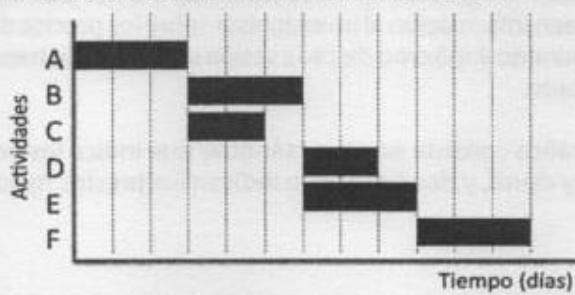
Solicite su usuario y contraseña escribiendo al correo electrónico info@auladeeconomia.com. Para ello deberá indicar el número de su factura de compra de texto.

Aplicación

Gráficas de Gantt

Un proyecto está compuesto por una serie de actividades secuenciales, o sea, que es necesario concluir unas actividades para poder iniciar otras. Por ejemplo, al construir una casa, es necesario haber terminado las paredes para poder construir los techos, y mientras los techos no estén terminados no se va realizar la instalación eléctrica. Además, cada actividad tiene una duración estimada.

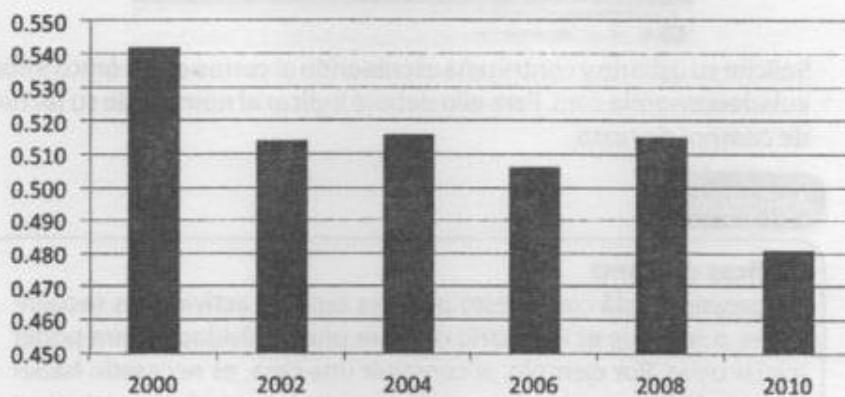
La gráfica de Gantt es un tipo de gráfica de barras que se ha convertido en una herramienta que ayuda al gestor de un proyecto a visualizar la secuencia de actividades a lo largo del tiempo. Por ejemplo, la siguiente gráfica correspondería a un proyecto compuesto por 6 actividades. La actividad A es la primera y las actividades B y C pueden iniciar cuando la actividad A esté concluida.



El encargado del proyecto podría agregar las fechas de inicio y terminación de las actividades del proyecto en el eje horizontal, y así la gráfica sería una herramienta útil en la programación y control de la ejecución del proyecto. Existe software que permite crear estas gráficas con facilidad.

Cuando la serie es cronológica, entonces se puede construir un gráfico de barras verticales. En el eje horizontal siempre va a aparecer el tiempo, expresado en años, meses, días, etc., y en el eje vertical el valor de la variable en cuestión. La siguiente gráfica es un ejemplo del uso de la gráfica de barras verticales. El siguiente es un ejemplo de la utilización de las barras verticales simples:

**México: Distribución del ingreso – 2000 a 2010
(Índice de concentración de Gini)**



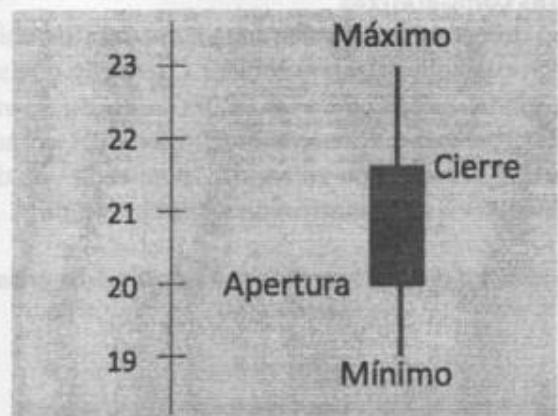
Fuente: CEPAL. CEPALSTAT, Estadísticas e indicadores sociales. Recuperado el 3 de octubre de 2012 de <http://websie.ediac.cl/sisgen/ConsultarIntegrada.asp?idAplicacion=1&idTema=363&idioma=e>

Aplicación

Gráficas de velas

En los mercados bursátiles, de divisas (Forex) y de otros activos se emplean las gráficas de velas (candelas o velas japonesas), las cuales proveen información al inversionista sobre los precios de apertura, cierre, mínimo y máximo de cada sesión o periodo de transacciones en el mercado.

La gráfica consiste en un rectángulo que indica los precios de apertura y cierre, y dos líneas que indican los precios máximo y mínimo.



Por ejemplo, suponga que la gráfica siguiente representa el precio de una acción de una empresa, cuyo precio de apertura en una determinada sesión de la bolsa de valores fue de \$20, por eso la base del rectángulo inicia al nivel de \$20 con respecto al eje vertical. Durante esa sesión el precio de la acción bajó hasta \$19, por eso sale una línea que llega hasta ese valor. También, durante esa sesión el precio llegó a subir hasta \$23, como lo indica la línea en la parte superior del rectángulo. Al final de la sesión, el precio cerró en \$21,50 como lo señala el lado superior del rectángulo.

Estas gráficas son elaboradas empleando sistemas informáticos. Cuando el precio sube, según el sistema que genera la gráfica, la vela es, generalmente, de color verde o azul, pero cuando el precio baja, lo usual es que sea roja. Observe el siguiente gráfico que muestra un conjunto de vela a lo largo de varios días y cómo ayudan a observar la tendencia en los precios de los activos transados.

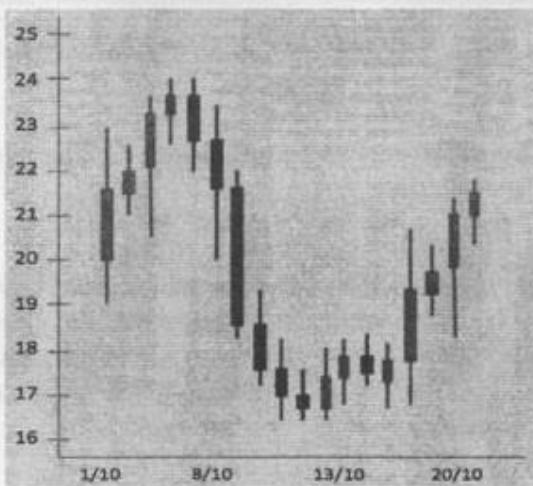
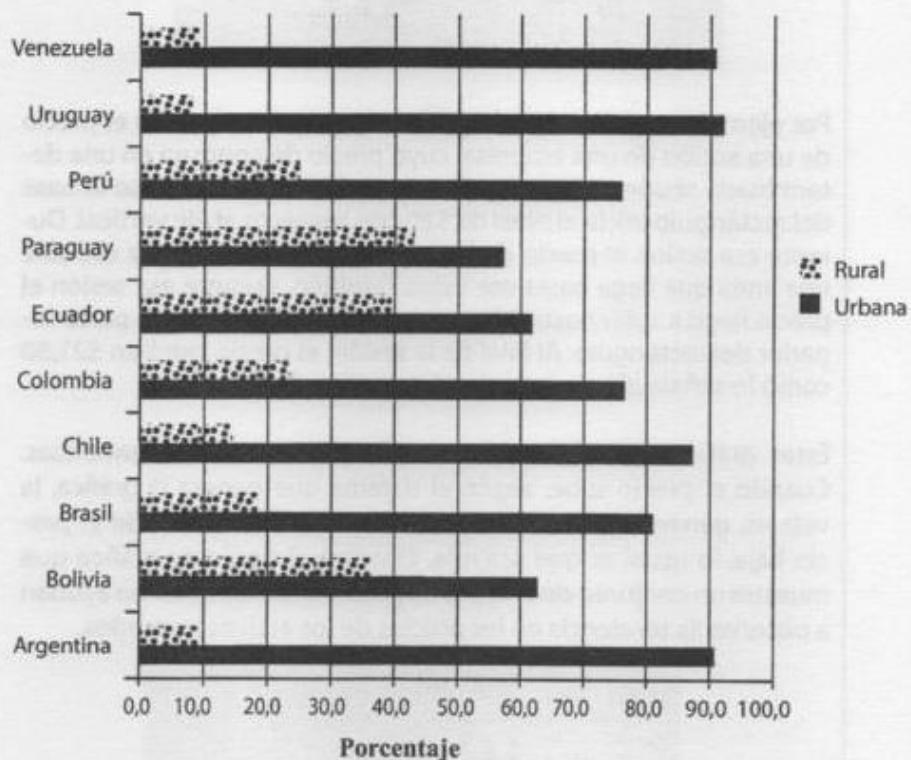


Gráfico de barras comparativas

Este tipo de barras se emplean para comparar (de ahí su nombre) varias series, ya sean cualitativas o cronológicas. Cuando lo que se desea es comparar varias categorías cada una de las cuales se divide en varias partes, entonces se pueden usar barras horizontales comparativas, como en el siguiente gráfico, en el cual se compara de país a país, pero cada una de las categorías (los países) se dividen en población urbana y población rural.

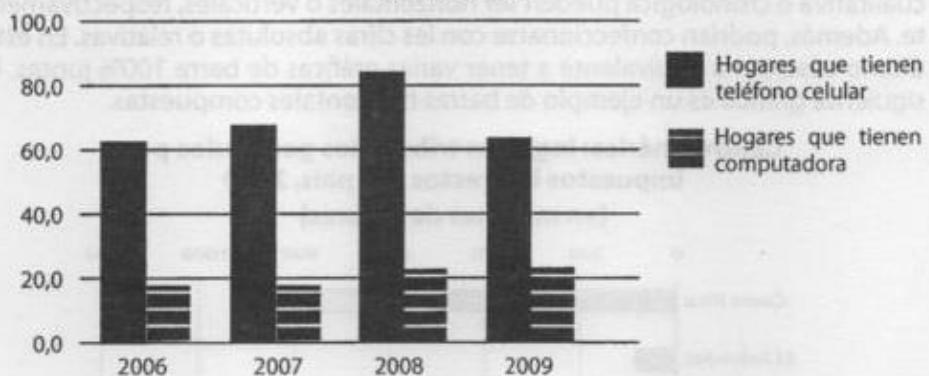
América del Sur: Porcentaje de población urbana y rural, por país, 2000



Fuente: Base de datos DEPUALC, 2009. CELADE-División de población de la CEPAL.
www.cepal.org/celade/depualc

Cuando la finalidad es comparar varias series cronológicas, entonces se pueden usar las barras verticales comparativas. Por ejemplo, en el siguiente gráfico se muestra la evolución del porcentaje de hogares con acceso a servicios de telefonía celular y el porcentaje de hogares con computadora en Colombia en el periodo que va del año 2006 al 2009.

**Colombia: Porcentaje de hogares que tienen teléfono celular móvil
y porcentaje de hogares que tienen computadora,
2006 – 2009**

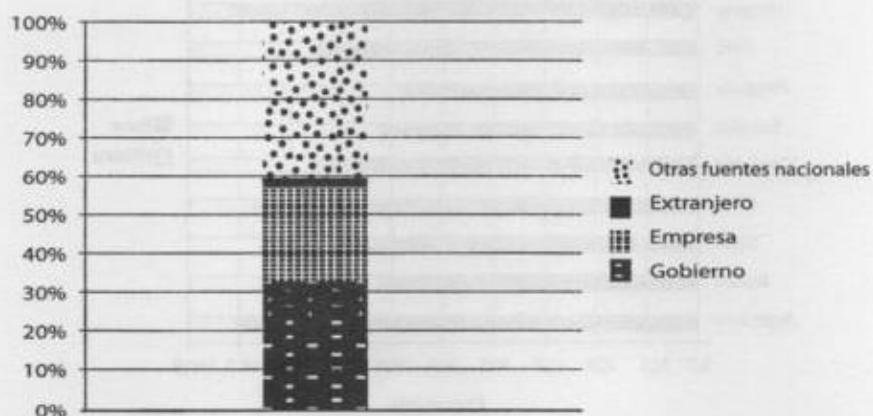


Fuente: CEPAL (2012). CEPAL STAT Indicadores de tecnologías de información: Acceso a TIC en los hogares. Recuperado el 9 de octubre de 2012 de <http://websie.eclac.cl/sisgen/ConsultaIntegrada.asp?idAplicacion=16&idioma=e>

Gráfico de barra 100 %

Cuando interesa mostrar las partes en que se divide un todo o la estructura que compone un todo (igual que la gráfica circular o de pastel), se puede usar la barra 100%. Para construir este tipo de gráfico las partes se presentan, presentando abajo la mayor y de ahí hasta la menor (o de izquierda a derecha, si se hace horizontal), de modo que cada parte se proporcional a la cifra que representa, la cual se muestra como porcentaje del total. Si ese total se divide en muchas partes o si hay partes muy pequeñas, entonces no se recomienda el uso de este tipo de gráfico, excepto que sea posible fusionar algunas de esas partes, generalmente en una categoría denominada "otros". El siguiente es un ejemplo de la aplicación de este tipo de gráfica:

**Argentina: Gasto en Investigación y Desarrollo (I+D)
por sector de financiamiento, 2004**

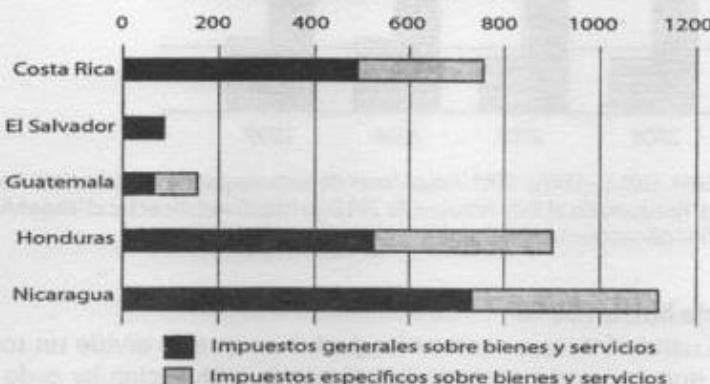


Fuente: CEPAL (2012). Info-Data: Indicadores de composición de gasto en I+D. Recuperado el 26 de setiembre de 2012 de http://www.eclac.cl/iyd/website/default_002.asp

Gráfico de barras compuestas

Este tipo de barras se emplea para representar la composición o estructura de las partes que componen una serie de totales. Según si la serie es cualitativa o cronológica pueden ser horizontales o verticales, respectivamente. Además, podrían confeccionarse con las cifras absolutas o relativas. En este último caso, sería equivalente a tener varias gráficas de barra 100% juntas. El siguiente gráfico es un ejemplo de barras horizontales compuestas.

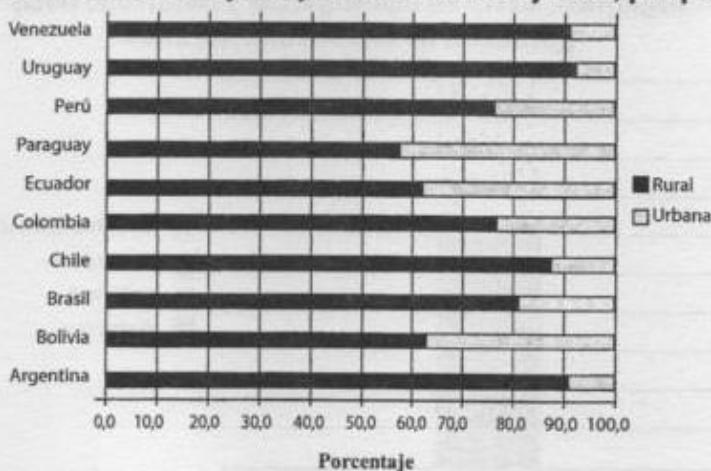
**Centroamérica: Ingresos tributarios generados por impuestos indirectos por país, 2010
(en millones de dólares)**



Fuente: CEPAL (2012). Info-Data: Indicadores fiscales. Recuperado el 10 de octubre de 2012 de <http://websie.eclac.cl/sisgen/ConsultalIntegrada.asp?idAplicacion=6&idTema=140&idioma=e>

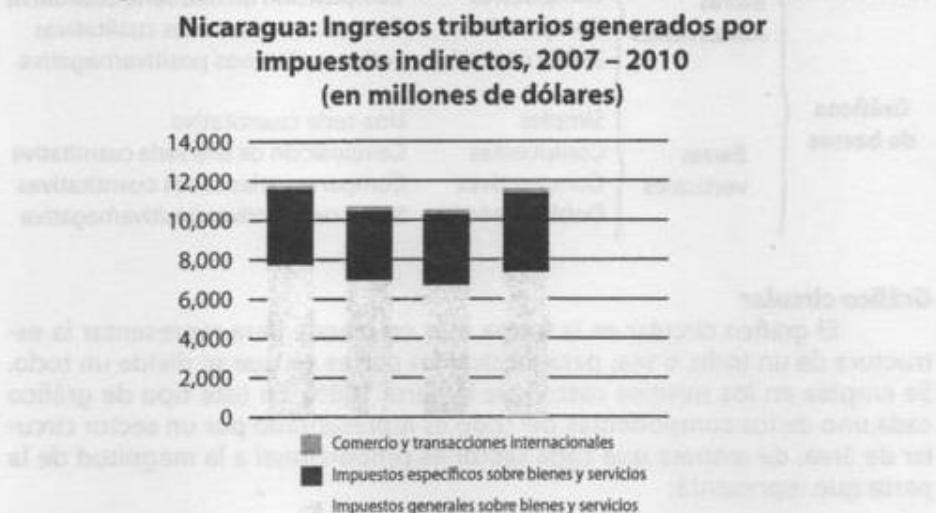
La siguiente gráfica es un ejemplo de una gráfica de barras horizontales compuestas, pero en este caso los datos están expresados en términos relativos, por lo cual la longitud de todas las barras es la misma, es decir, todas llegan al 100%. En la gráfica cada barra muestra cómo se compone la población de cada país entre población urbana y rural para cada uno de los países de América del Sur.

América del Sur: Porcentaje de población urbana y rural, por país, 2000



Fuente: Base de datos DEPUALC, 2009. CELADE-División de población de la CEPAL. www.cepal.org/celade/depualc

Las barras compuestas también pueden ser verticales, esto cuando la serie sea cronológica, como en el ejemplo siguiente, el cual muestra cómo se componen los ingresos por impuestos indirectos de Nicaragua en el periodo 2007 a 2010.

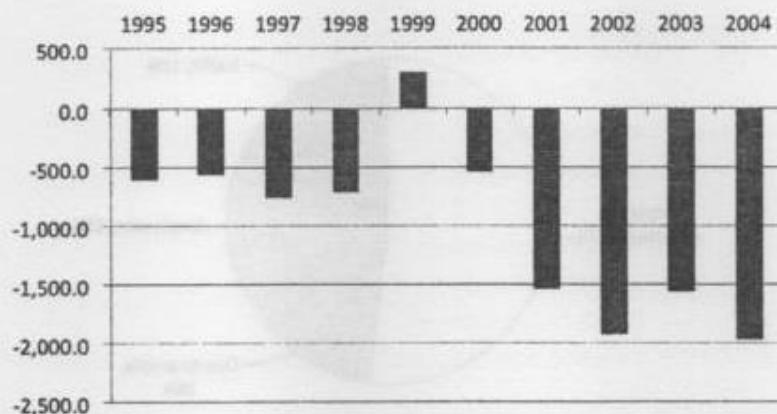


Fuente: CEPAL (2012). Info–Data: Indicadores fiscales. Recuperado el 10 de octubre de 2012 de <http://websie.eclac.cl/sisgen/ConsultaIntegrada.asp?idAplicacion=6&idTema=140&idioma=e>

Gráfico de barras de doble dirección

Este gráfico se utiliza para representar las variaciones, o sea, los aumentos y las disminuciones que una serie podría presentar. Por ejemplo, la gráfica siguiente muestra los saldos de la balanza comercial de Costa Rica para varios años. Cuando la serie es cualitativa, entonces estas barras son horizontales, y cuando es cronológica, entonces son verticales.

**Costa Rica: Saldo de la balanza comercial (en millones de US\$)
1995 – 2004**



Fuente: Banco Central de Costa Rica (2012). Balance comercial. Recuperado el 12 de octubre de 2012 de <http://indicadoreseconomicos.bccr.fi.cr/Indicadoreseconomicos/Cuadros/frmVerCatCuadro.aspx?CodCuadro=448&idioma=1&FecInicial=1995/01/31&FecFinal=2005/12/31&Jump=1>

Resumen sobre el uso de gráficos de barras

Gráficos de barras	TIPOS DE GRÁFICOS DE BARRAS	
	Barras horizontales	Barras verticales
	Simples	Una serie cualitativa
	Compuestas	Composición de una serie cualitativa
	Comparativas	Comparar varias series cualitativas
	Doble dirección	Series cualitativas positiva/negativa
	Simples	Una serie cuantitativa
	Compuestas	Composición de una serie cuantitativa
	Comparativas	Comparar varias series cuantitativas
	Doble dirección	Series cuantitativas positiva/negativa

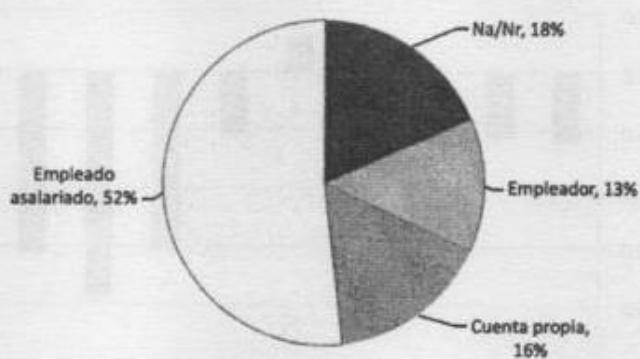
Gráfico circular

El gráfico circular es la forma más empleada para representar la estructura de un todo, o sea, para mostrar las partes en que se divide un todo. Se emplea en los mismos casos que la barra 100%. En este tipo de gráfico cada uno de los componentes del todo es representado por un sector circular de área, de manera que cada sector es proporcional a la magnitud de la parte que representa.

GRÁFICO CIRCULAR: Es un gráfico en el cual las partes de un todo son representadas de manera proporcional como sectores de un círculo.

Se recomienda ordenar los sectores de acuerdo con su magnitud y no es conveniente utilizarlo cuando el número de partes es grande, o bien, cuando hay partes muy pequeñas.

Costa Rica: Distribución de los hogares con acceso a internet según categoría ocupacional principal, 2009



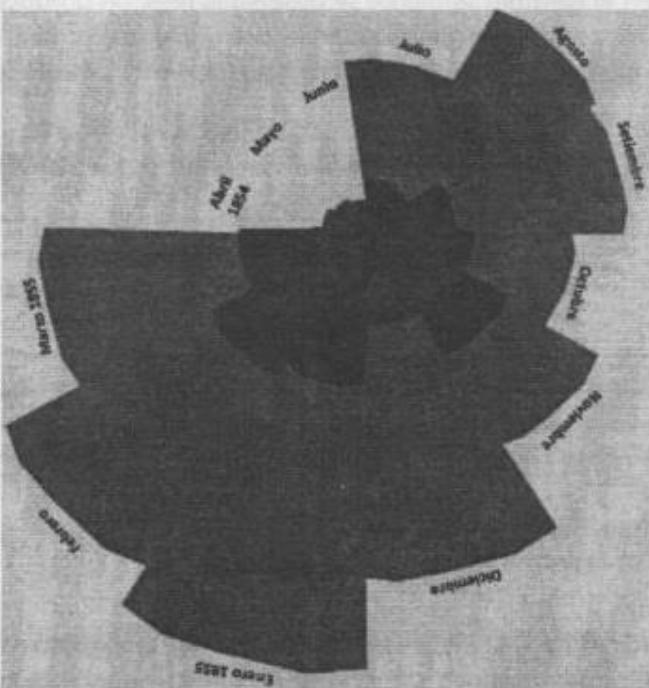
Fuente: Encuesta de Hogares de Propósitos Múltiples, 2009.
Recuperado el 26 de setiembre de 2012 de <http://www.eclac.cl/tic/flash/>

Aplicación

Diagrama de área polar

Florence Nightingale fue precursora de la práctica de la enfermería, pero también hizo algunas contribuciones importantes a la estadística. Durante la Guerra de Crimea trabajó como enfermera en un hospital de la campaña. Ella observó que las condiciones higiénicas del hospital eran muy pobres y que en determinados momentos se daban más muertes por enfermedades infecciosas que por la guerra misma. Entonces creó un tipo de gráfico que mostrara el número de muertes mensuales según su causa, llamado gráfico de área polar.

El gráfico de área polar consiste en un círculo que ha sido dividido en sectores, cada uno del mismo ángulo. Cada uno de estos sectores tiene un radio diferente, y corresponde a un mes determinado y que muestra en colores distintos el tipo de causa de muerte. El tamaño del sector es proporcional al número de muertes.



En la gráfica las zonas azules simbolizan el número de muertes por enfermedades infecciosas, las rojas, muertes por heridas y las negras, muertes por otras causas.

Nightingale introdujo una serie de mejoras en las condiciones higiénicas que ayudaron a reducir el número de muertes por enfermedades infecciosas.

Gráficas lineales

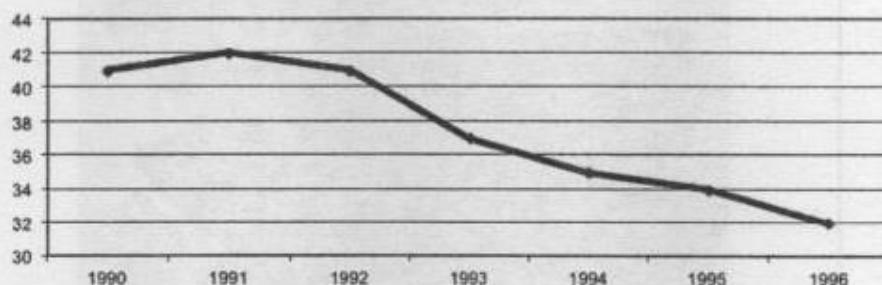
Los gráficos lineales se emplean, principalmente, para representar series cronológicas de datos, o sea, valores anuales, mensuales, etc.

GRÁFICA LINEAL: Es una representación de datos en coordenadas cartesianas, generalmente numéricos, mediante líneas, para observar su comportamiento, frecuentemente a lo largo del tiempo.

El gráfico lineal se emplea en los mismos casos que la gráfica de barras verticales, pero, generalmente se prefiere el gráfico lineal cuando el número de períodos es grande, ya que la gráfica de barras quedaría con muchas barras muy delgadas, y entonces no sería atractiva. El gráfico lineal es especialmente útil para mostrar la tendencia de una variable a lo largo del tiempo, o sea, presentar cómo se han comportado en un plazo determinado y cuál ha sido su crecimiento o decrecimiento "absoluto".

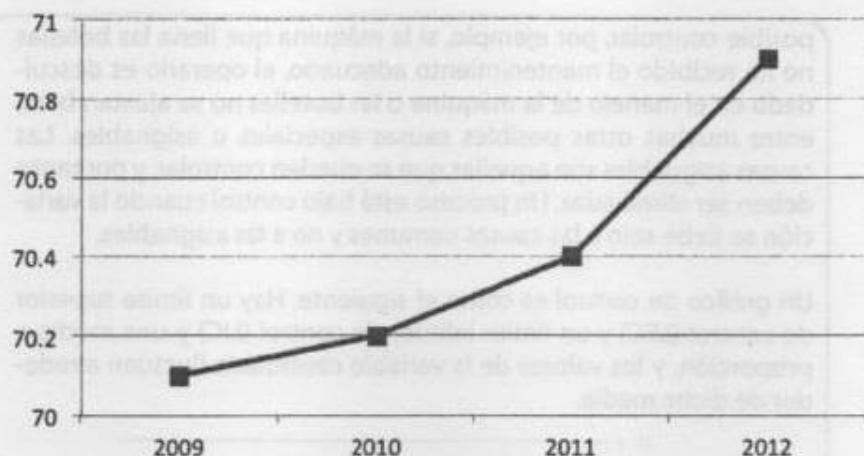
En el ejemplo siguiente se observa la relación producción / consumo interno en la refinación de hidrocarburos en Guatemala, en el periodo 1990 a 1996. Se observa claramente una tendencia a la baja a partir de 1991.

**Guatemala: Refinación de Hidrocarburos:
relación producción/consumo interno. 1990-1996**



Fuente: Informe Estado de la Región en Desarrollo Humano Sostenible.
Proyecto Estado de la Región. San José, C.R.: 1999.

Al construir gráficos es necesario ser muy cuidadoso con el eje vertical, ya que fácilmente podría construirse un gráfico engañoso, como en el caso siguiente, en el cual pareciera que la variable tiene un fuerte crecimiento, sin embargo su variación es mínima. Esto se da porque no se ha incluido el cero en el eje vertical.



Para evitar estos problemas con la escala es importante considerar los siguientes aspectos:

- En el eje vertical el cero es necesario, pues de otro modo la impresión que causa podría estar distorsionada.
- La escala empleada debe facilitar la comparación, por lo que se recomienda que sea sencilla.
- Cuando se grafican datos de gran magnitud, la curva podría quedar muy arriba en el gráfico, entonces es difícil comenzar la escala en cero, por lo que en estos casos podría interrumpirse la escala y no mostrar el cero. Cuando se representan números índices, tampoco es necesario incluir el cero (pero sí es importante la base del índice).

Aplicación

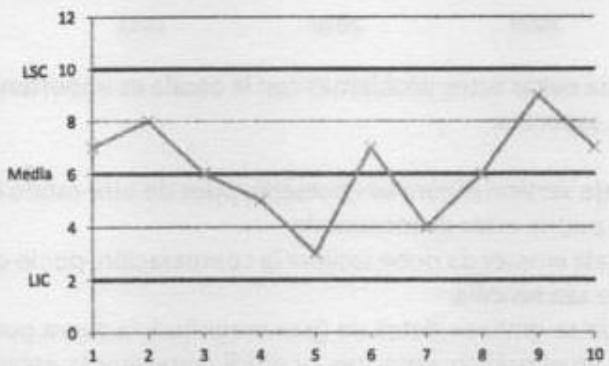
Gráficas de control

Las empresas preocupadas por el mejoramiento continua realizan un control estadístico de la calidad. Una de las herramientas que se emplean en esa gestión de la calidad es la elaboración de gráficas de control. Estas gráficas son un tipo de gráfica lineal que permiten llevar estadísticas de un proceso a lo largo del tiempo y del análisis de dichas gráficas se puede determinar si el proceso está bajo control o si está fuera de control.

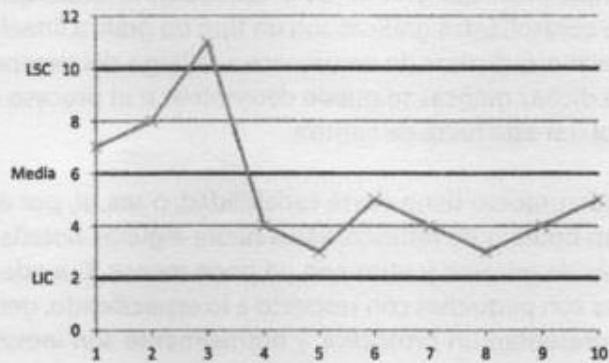
Todo proceso tiene cierta variabilidad, o sea, si, por ejemplo, se llenan botellas de refresco, van a haber algunas botellas con un poco más de refresco y otras con un poco menos. Cuando esas variaciones son pequeñas con respecto a lo especificado, generalmente no representan un problema, y normalmente son inevitables, y se les denomina causas comunes. Pero existen otras variaciones que si es

possible controlar, por ejemplo, si la máquina que llena las botellas no ha recibido el mantenimiento adecuado, el operario es descuidado en el manejo de la máquina o las botellas no se ajustan bien, entre muchas otras posibles causas especiales o asignables. Las causas asignables son aquellas que se pueden controlar, y por tanto deben ser eliminadas. Un proceso está bajo control cuando la variación se debe solo a las causas comunes y no a las asignables.

Un gráfico de control es como el siguiente. Hay un límite superior de control (LSC) y un límite inferior de control (LIC) y una media o proporción, y los valores de la variable controlada fluctúan alrededor de dicha media.



En términos del gráfico, un proceso está bajo control cuando todos los puntos se encuentran dentro de los límites de los límites de control (LIC y LSC) y además, cuando los puntos no muestran patrones no aleatorios, aun cuando estén los puntos dentro de los límites de control. Por ejemplo, en la siguiente gráfica, el tercer punto está fuera de los límites de control, y luego hay otra situación fuera de control, pues todos los puntos siguientes están debajo de la media, lo cual indicaría que las variaciones no son aleatorias, y por tanto ese patrón indicaría que no solo las causas comunes están provocando variación en el proceso.

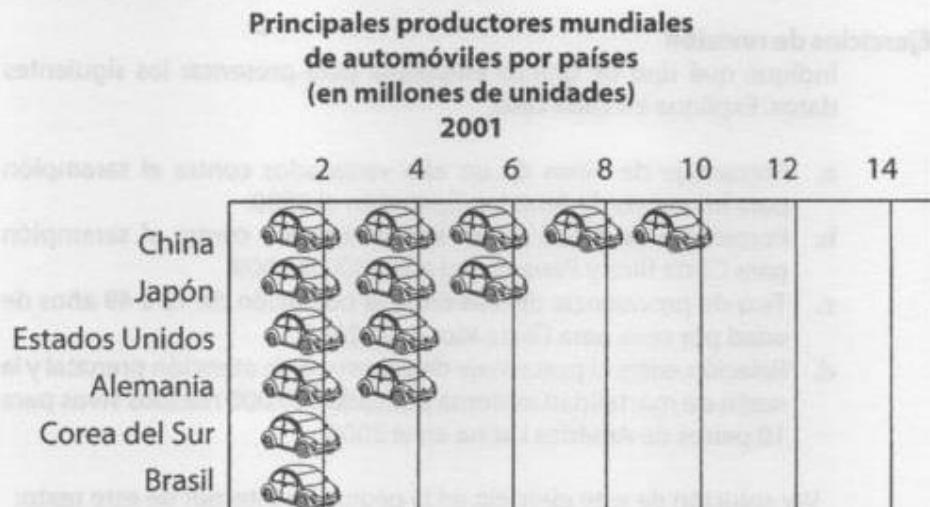


Pictogramas

Los pictogramas son similares a los gráficos de barras, pero sustituyen las barras por varias figuras pequeñas del mismo tamaño. Se utilizan, generalmente, con el objeto de realizar comparaciones. Generalmente las figuras empleadas se relacionan con la variable graficada. Por ejemplo, si el dato graficado es producción de automóvil, podrían dibujarse pequeños automóviles. Mientras más automóviles aparezcan, es porque se da una mayor producción.

PICTOGRAMA: Es un diagrama que utiliza algún tipo de imagen o símbolo para mostrar los datos a fin de lograr una rápida comprensión de los mismos.

El siguiente es un ejemplo que se relaciona con los principales productores mundiales de automóviles, por eso en el pictograma, en vez de barras, se dibujan pequeñas figuras de automóviles.



Fuente: PricewaterhouseCoopers

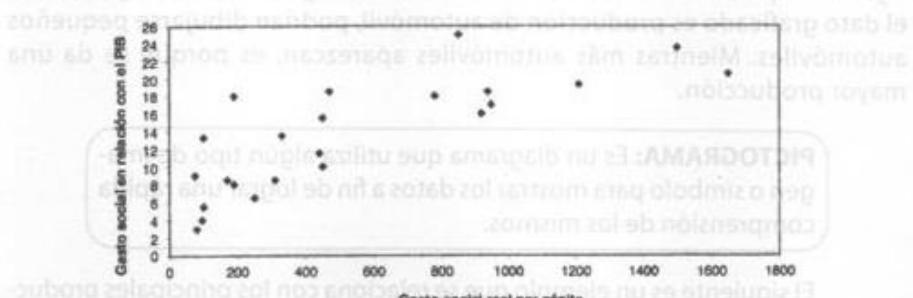
Diagrama de dispersión

En muchos casos se desea establecer la relación que existe entre dos variables de la misma unidad de estudio. Por ejemplo, el ingreso y el consumo de las personas, la cantidad de tabaco fumado y la incidencia del cáncer, etc.

DIAGRAMA DE DISPERSIÓN: Es un tipo de diagrama que utiliza coordenadas cartesianas para mostrar los valores de dos variables para un conjunto de datos, de manera que los datos se presentan como un conjunto de puntos, donde cada uno de ellos muestra el valor de una variable en el eje horizontal y el valor de la otra variable en el eje vertical.

Estos diagramas son base de muchos análisis en los que se puede establecer la correlación entre variables.

América Latina: gasto público social real en relación con el PIB, según gasto público social real per cápita. 1990, 2001



Fuente: CEPAL, 2004.

Ejercicios de revisión

Indique qué tipo de gráfico emplearía para presentar los siguientes datos. Explique en cada caso:

- Porcentaje de niños de un año vacunados contra el sarampión para los países de América Central en el 2009.
- Porcentaje de niños de un año vacunados contra el sarampión para Costa Rica y Panamá del año 2000 al 2009.
- Tasa de prevalencia del VIH entre la población de 15 a 49 años de edad por sexo para Costa Rica en el 2009.
- Relación entre el porcentaje de cobertura de atención prenatal y la razón de mortalidad materna por cada 100.000 nacidos vivos para 10 países de América Latina en el 2006.

Ver solución de este ejercicio en la página de internet de este texto:
www.auladeeconomia.com/raeep.html

Apoyo audiovisual y uso de la tecnología

En la página de internet www.auladeeconomia.com/raeep.html podrá encontrar una presentación de diapositivas que expone este tema y es una parte importante de este texto. Esta presentación presenta el tema en forma visual, pues emplea fotografías, esquemas u otros recursos visuales.

Adicionalmente puede encontrar algunos videos explicativos.

Solicite su usuario y contraseña escribiendo al correo electrónico info@auladeeconomia.com. Para ello deberá indicar el número de su factura de compra de texto.

2.8 Ejercicios

Ejercicios de desarrollo

Conteste cada una de las preguntas que a continuación se formulan (las respuestas a los ejercicios se encuentran en la página de internet del texto: www.auladeeconomia.com/raeep.html).

1. A continuación se presenta una lista de variables obtenidas en distintas investigaciones, en cada caso indique cuál tipo de gráfico debería efectuarse para representar en forma adecuada los datos:
 - a. Tiempo medio de espera de los pacientes de un hospital en el servicio de consulta externa obtenido mes a mes durante un año.
 - b. Prevalencia de enfermedades cardiovasculares por sexo y según rango de edad.
 - c. Porcentaje de ratas que presentaron interacciones de un nuevo medicamento con otras sustancias según cada tipo de sustancia.
 - d. Número de horas promedio de uso continuo de computadoras de los niños de 6 a 12 años según su sexo.
 - e. Población estimada de ranas en peligro de extinción en las zonas montañosas del país para los últimos 5 años.
 - f. Distribución porcentual de la población del país según niveles de peso (delgadez, normal, obesidad) en el país durante el último año.
 - g. Número de viviendas del país según su estado y por zona (urbana o rural).
 - h. Número de artículos defectuosos por día en el proceso de producción de una empresa para el último mes.
 - i. Tiempo de fusión de determinados componentes electrónicos según distintas sobrecargas.
 - j. Tipos de medidas de seguridad informática que las empresas financieras implementaron en el último semestre según tipo de entidad (banco, cooperativa, mutual, etc.).
 - k. Proporción de empresas que están dispuestas a implementar un nuevo sistema operativo en el transcurso de los próximos doce meses según tipo de empresa (pequeña, mediana, grande).
 - l. Rendimiento sobre la inversión promedio de las empresas que cotizan en la Bolsa de Valores de Nueva York según su sector de actividad económica (industria, servicios, etc.) para los últimos 10 años.
 - m. Porcentaje de habitantes del país que poseen un teléfono "inteligente" según su sexo.
 - n. Distribución porcentual de los distintos niveles de satisfacción de los empleados de una empresa por departamento de la empresa (finanzas, ventas, etc.).
 - o. Monto promedio mensual de las cuentas por cobrar de una empresa para los últimos seis meses.

- p. Distribución porcentual de la percepción del impacto de un nuevo paquete de impuestos según sector productivo del país.
- q. Número de turistas que visitan el país debido a la campaña efectuada por el gobierno para promocionar al país como destino turístico según país de origen del turista.
- r. Porcentaje de estudiantes de secundaria del país que emplean con regularidad internet para efectuar sus tareas según rango de edad.
- s. Número de personas que aplicaron para el puesto de gerente financiero según su tipo de personalidad de acuerdo con un test aplicado a los candidatos.
- t. Distribución porcentual de la opinión de los jefes del hogar sobre el efecto que tienen las redes sociales en internet en las relaciones familiares según rango de edad del jefe del hogar.
- u. Proporción niños dados en adopción según la evaluación de su estado general de salud efectuada en el último semestres.
- v. Número de adultos según su opinión sobre el efecto de las manifestaciones estudiantiles en la imagen del gobierno.
- w. Composición porcentual de la opinión de los periodistas del país según su grado de percepción del respeto a la libertad de prensa que se da en el país según grado académico.
- x. Volumen mensual de ventas de una compañía según la exposición mensual de los consumidores a los comerciales pautados por la misma empresa para los últimos 12 meses.

2. Considere la siguiente información:

"En un estudio sobre la prevalencia del asma en varios países de América Latina, en el grupo etario de 6 a 7 años, medida por el porcentaje de niños con diagnóstico médico de la enfermedad con la presencia de los principales síntomas respiratorios sugestivos de asma. En Brasil esa tasa fue de 11,3%, en Chile de 12,5%, en Costa Rica de 26,9% y en Panamá de 19,3%".

El proyecto fue elaborado por Dr. Manuel Soto-Martínez y Dr. Manuel Soto-Quirós PhD, publicado en 2004 en la Revista Médica del Hospital Nacional de Niños Dr. Carlos Sáenz Herrera.

Construya un cuadro estadístico que muestre la tasa de prevalencia del asma mencionada en los cuatro países citados. El cuadro debe contener todos los elementos estudiados.

3. Una farmacia comunitaria vende, además de distintos fármacos, otras líneas de productos, como productos de belleza, productos de cuidado personal, etc. Se ha reunido información de las ventas de los últimos 3 años según los distintos tipos de productos. En cada caso indique el tipo de gráfico que se podría construir para representar la información del modo adecuado:
- Ventas mensuales totales para los últimos 36 meses por tipo de producto.
 - Importancia relativa anual de las distintas líneas de producto para los últimos 3 años.
 - Comparación de los montos de ventas totales del último año por tipo de producto y según forma de pago (efectivo o tarjeta).
4. En un periódico se publicaron los siguientes datos sobre las ventas de teléfonos celulares según su marca: "En el 2011 del 6% de las ventas eran de teléfonos marca Samsung, pero para el 2012 su participación alcanzó el 18%. En ese mismo periodo, la participación de mercado de Nokia cayó del 69% al 43%. Por otro lado, Apple se mantiene en ambos periodos en un 15% del total. Sony ha perdido un poco de participación, pasando del 16% al 12%".
- Construya una tabla que muestre los datos presentados anteriormente.
 - Construya una gráfica que muestre los datos anteriores.
 - ¿Cuál de los dos modos de presentar la información estadística (cuadro o gráfica) preferiría usted en este caso?
5. La presencia de manganeso en la gasolina puede significar un problema, pues genera ciertos daños a los vehículos, a la salud de las personas y al ambiente. Ante la duda de altas concentraciones de esta sustancia en la gasolina vendida en el país se toman muestras en los distintos planteles de la compañía refinadora de petróleo para que vaya a ser estudiadas en un laboratorio. Los siguientes son datos recopilados para los 4 planteles que posee la refinadora y corresponden a 3 mediciones realizadas a lo largo del año (los datos se refieren a la concentración promedio obtenida en cada ocasión en mg/L de gasolina):

Plantel	Febrero	Mayo	Agosto
A	5	7	4
B	1,2	3,5	0,7
C	0,4	0,3	0,5
D	18	20	15

Indique cuál tipo de gráfico construiría en cada caso si se desea:

- a. Mostrar los niveles de manganeso en el plantel A en las tres mediciones efectuadas.
 - b. Comparar los niveles de manganeso para los cuatro planteles en la medición de agosto.
 - c. Comparar el comportamiento de los niveles de manganeso a lo largo de las tres mediciones para los planteles B y C.
6. Un contador está realizando una auditoría de los cheques extendidos por una compañía. En cada caso indique el tipo de gráfico adecuado para presentar los datos:
- a. Número de cheques extendidos por mes durante el último año.
 - b. Monto promedio de los cheques extendidos cada mes durante el último año.
 - c. Número de cheques extendidos durante el último mes según tipo de gasto.
7. Con base en la información que se le brinda a continuación, elabore un gráfico de barras horizontales, un gráfico de barras verticales, un gráfico lineal y un gráfico circular. Escoja el gráfico apropiado en cada caso.
- a. El índice de precios al consumidor (IPC) para Imaginalandia ha tenido la siguiente evolución: fue 1215,36 en 1999; 1369,87 en 2000; 1502,56 en 2001; 1672,14 en 2002; 1802,35 en 2003; 1963,87 en 2004; 2135,98 en 2005; 2218,78 en 2006. Muestre gráficamente la tendencia que ha tenido el IPC.
 - b. Los siguientes de datos corresponden a las tasas de desempleo en zonas rural y urbana en Imaginalandia de 2001 a 2005. Elabore una gráfica que compare ambas tasas por año.

Zona	2001	2002	2003	2004	2005
Rural	15,6	15,5	15,3	14,9	14,9
Urbana	11,3	11,1	10,8	10,5	10,3

- c. En 2006 había en Imaginalandia un déficit de 155659 Km. de carreteras, el 61,3% era en las zonas costeras y el resto en el interior del país. Muestre gráficamente la distribución del déficit según la zona.
- d. En 2005 el total de las exportaciones de Imaginalandia se compuso del modo siguiente (en millones de US\$):

Tipo de actividad	Valor de las exportaciones
Pecuario y pesca	262,3
Agrícola	578,4
Textiles	674,4
Industria alimentaria	323,0

- e. Con base en la información del último punto de la pregunta anterior, relacionado con las exportaciones, elabore un cuadro estadístico, con todos sus elementos, que muestre las cifras absolutas y también las relativas.
8. Cada trimestre la cámara de empresas turísticas local efectúa un estudio sobre las expectativas de los empresarios afiliados. Se aplica un cuestionario entre una muestra de empresas relacionadas con el turismo en el país. En el cuestionario se pregunta a los gerentes de las empresas aspectos como el grado de satisfacción con la rentabilidad de su negocio, sus planes de contratación de personal para los próximos meses, su percepción de la demanda del servicio que presta, el tipo de negocio (hotel, restaurante, operador de tours, etc.).
- a. Indique qué tipo de gráfico emplearía en cada caso:
 - a.1. Número de empresarios según el grado de satisfacción con la rentabilidad de su negocio según los resultados de este año.
 - a.2. Distribución porcentual para este año de los empresarios de acuerdo con sus planes de contratación de personal para los próximos meses.
 - a.3. Percepción de la demanda del servicio que presta según tipo de negocio para los últimos tres años.
 - b. Construya una gráfica que muestre los datos siguientes:
 - b.1. Un 7% de los empresarios consideró la rentabilidad de su empresa como muy buena, un 19% como buena, un 39% como regular, un 25% como mala y solo un 10% la consideró muy mala.
 - b.2. Los resultados de expectativa con respecto al empleo fue que un 59% afirmó que mantendrá su planilla igual en el 2013, un 20% dijo que la aumentará y un 21% anunció que la bajará. El año pasado estos datos fueron 48%, 36% y 16% respectivamente.
 - b.3. En lo referente a la demanda, en el caso de los hoteles un 70% consideró que en el 2013 la demanda será menor que en el 2012, un 20% creen que seguirá igual y el resto piensan que aumentará. Entre los restaurantes, un 40% piensan que la demanda disminuirá, un 45% consideran que seguirá igual y el resto creen que aumentará. Con respecto a las empresas relacionadas con el transporte de turistas, un 50% consideran que la demanda disminuirá, un 40% que seguirá igual y el resto estiman que bajará.

9. Con base en la información que se le brinda a continuación, elabore un gráfico de barras horizontales, un gráfico de barras verticales, un gráfico lineal y un gráfico circular. Escoja el gráfico apropiado en cada caso.
- El índice de desarrollo humano (IDH) para un país llamado ha tenido la siguiente evolución durante la década pasada: fue 0,916 en 1990; 0,876 en 1991; 0,842 en 1992; 0,852 en 1993; 0,848 en 1994; 0,883 en 1995; 0,884 en 1996; 0,889 en 1997. Muestre gráficamente la tendencia que ha tenido el IDH.
 - Los siguientes de datos corresponden a las tasas de repitencia en primaria y secundaria en Costa Rica de 1993 a 1997. Elabore una gráfica que compare ambas tasas por año.
- | Año | 1993 | 1994 | 1995 | 1996 | 1997 |
|------------|------|------|------|------|------|
| Primaria | 8,6 | 9,3 | 9,8 | 10,1 | 9,1 |
| Secundaria | 30,5 | 33,6 | 37,4 | 35,8 | 32,4 |

- En 1994 habían en la República de Santa Marta un déficit de 58740 viviendas, el 31,3% de ellas era en área urbana y el 68,7% en áreas rurales. Muestre gráficamente la distribución del déficit de viviendas según la zona.
 - Con base en la información del punto c de la pregunta anterior, relacionado con la vivienda en la República de Santa Marta, elabore un cuadro estadístico, con todos sus elementos, que muestre las cifras absolutas y relativas descritas.
10. La Escuela de Negocios de la Universidad de Edimburgo, en el Reino Unido, realizó un estudio sobre el uso que las personas realizan en la red social Facebook. Según esta investigación la mayor parte de los usuarios de Facebook tienen amigos que proceden de sus diferentes círculos sociales. El 97% de los usuarios afirma que conoce presencialmente a sus amigos de Facebook y no por Internet. Un 81% indica tener relación con primos y otros familiares lejanos. El 80% señala tener como amigos a sus hermanos, un 69% dice que tiene relación con los amigos de sus amigos y el 65% se relaciona con sus colegas. Construya un gráfico apropiado para representar toda la información anterior.

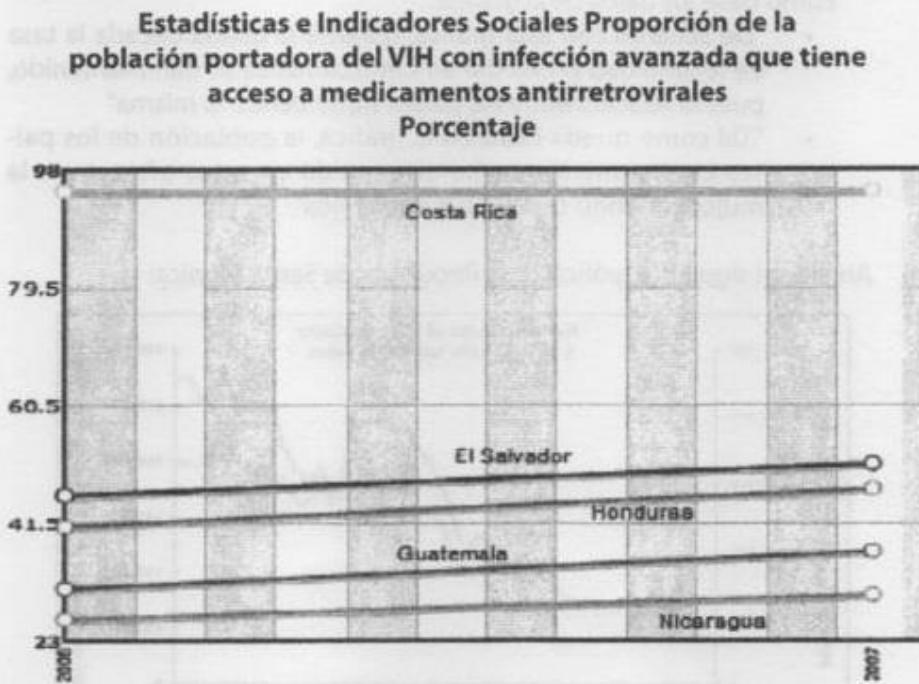
11. Con base en la siguiente tabla efectúe lo que a continuación se le solicita:

Población total del País X
Por provincias
A Octubre de 1998

Provincia	Población	%
Provincia A	450.000	
Provincia B	300.000	
Provincia C	100.000	
Total		

- Complete la tabla.
- Elabore una gráfica que muestre la población por provincias.
- Elabore una gráfica que muestre la distribución porcentual de la población por provincias.

12. Analice el siguiente gráfico:



Fuente: CEPAL

- Identifique al menos 2 errores en el gráfico anterior. Explique cada uno.
- Exponga una conclusión a la que usted podría llegar al estudiar el gráfico anterior. Explique.

13. Analice el siguiente gráfico:



- Identifique al menos 3 errores en el gráfico anterior. Explique cada uno.
- Con base en los datos una persona efectuó los siguientes comentarios, analice cada frase e indique si es verdadera o falsa, tomando como base los datos de la gráfica.
 - "De acuerdo con esta gráfica, durante la última década la tasa de fecundidad promedio en Centroamérica se han mantenido, pues la relación entre los países sigue siendo la misma"
 - "Tal como queda claro en la gráfica, la población de los países centroamericanos ha disminuido en estos años, pues las mujeres tienen cada vez menos hijos"

14. Analice el siguiente gráfico de la República de Santa Mónica:



- Identifique al menos 3 errores en el gráfico anterior. Explique cada uno.
- Con base en los datos el Ministro de Salud de Santa Mónica efectuó, a inicios de 2004, los siguientes comentarios, analice cada frase e indique si es verdadera o falsa, tomando como base los datos de la gráfica.

- "De acuerdo con esta gráfica, desde 1950, excepto algunos períodos, la tendencia del gasto per cápita en salud en nuestro país es creciente, especialmente durante la última década"
- "Tal como queda claro en la gráfica, el gasto per cápita en salud excede al producto interno bruto per cápita (PIB per cápita) desde hace ya muchos años"
- "Hasta antes de la crisis de finales de la década de los 70's e inicios de los 80's, el gasto per cápita en salud crecía más aceleradamente que el PIB per cápita, tal como se muestra claramente en la gráfica"

Examen del capítulo

En cada caso seleccione la opción que mejor contesta cada pregunta (las respuestas a los ejercicios se encuentran en la página de internet del texto: www.auladeeconomia.com/raep.html).

1. Si se quiere representar la composición de un todo, el tipo de gráfico adecuado para representar esos datos es:
 - (a) Gráfica de barras horizontales
 - (b) Gráfica de barras verticales
 - (c) Gráfica circular
 - (d) Pictograma
2. Si se tiene una serie cronológica, el tipo de gráfico adecuado para representarla es:
 - (a) Gráfica de barras horizontales
 - (b) Gráfica de barras verticales
 - (c) Histograma
 - (d) Diagrama de dispersión
3. Si se tiene una serie cualitativa, el tipo de gráfico adecuado para representarla es:
 - (a) Gráfico circular
 - (b) Gráfico lineal
 - (c) Gráfico de barras verticales
 - (d) Gráfico de barras horizontales
4. ¿Qué tipo de gráfico utilizaría para representar el porcentaje de niños de un año vacunados contra el sarampión para los países de América Central en el 2009?
 - (a) Gráfico circular
 - (b) Gráfico lineal
 - (c) Gráfico de barras verticales
 - (d) Gráfico de barras horizontales

5. ¿Qué tipo de gráfico utilizaría para representar el porcentaje de niños de un año vacunados contra el sarampión para Costa Rica y Panamá del año 2000 al 2009?
- (a) Gráfico de barras horizontales comparativas
 - (b) Gráfico de barras horizontales compuestas
 - (c) Gráfico de barras verticales comparativas
 - (d) Gráfico de barras verticales compuestas
6. ¿Qué tipo de gráfico utilizaría para representar la distribución porcentual del número de personas afectadas por el VIH entre la población de 15 a 49 años de edad para Costa Rica en el 2009?
- (a) Gráfico circular
 - (b) Gráfico lineal
 - (c) Gráfico de barras verticales
 - (d) Gráfico de barras horizontales
7. ¿Qué tipo de gráfico utilizaría para representar la relación entre el porcentaje de cobertura de atención prenatal y la razón de mortalidad materna por cada 100.000 nacidos vivos para 10 países de América Latina en el 2006?
- (a) Gráfico circular
 - (b) Diagrama de dispersión
 - (c) Pictograma
 - (d) Gráfico de barras horizontales
8. Si usted va a representar las exportaciones anuales de un país en el periodo 2009 – 2011, ¿cuál tipo de gráfico es más apropiado?
- (a) Gráfico circular
 - (b) Gráfico lineal
 - (c) Gráfico de barras verticales de doble dirección
 - (d) Gráfico de barras horizontales
9. Si usted va a representar las exportaciones anuales de un país en el 2011 clasificadas por tipo de producto, ¿cuál tipo de gráfico es más apropiado?
- (a) Gráfico circular
 - (b) Gráfico lineal
 - (c) Gráfico de barras verticales de doble dirección
 - (d) Gráfico de barras horizontales
10. Si usted va a representar las exportaciones anuales de un país en el periodo 2009 – 2011 por tipo de producto, ¿cuál tipo de gráfico es más apropiado?
- (a) Gráfico de barras horizontales comparativas
 - (b) Gráfico de barras horizontales compuestas
 - (c) Gráfico de barras verticales comparativas
 - (d) Gráfica de dispersión

11. ¿Qué tipo de gráfico utilizaría para representar el monto de las ventas (en dólares) de una empresa por tipo de producto para el año 2012?
- (a) Gráfico circular
 - (b) Gráfico lineal
 - (c) Gráfico de barras verticales
 - (d) Gráfico de barras horizontales
12. ¿Qué tipo de gráfico utilizaría para representar el monto de las ventas (en dólares) de una empresa por año del 2007 al 2012?
- (a) Gráfico circular
 - (b) Barra 100%
 - (c) Gráfico de barras verticales
 - (d) Gráfico de barras horizontales
13. ¿Qué tipo de gráfico utilizaría para representar la distribución porcentual de las ventas de una empresa por tipo de producto para el año 2012?
- (a) Gráfico circular
 - (b) Gráfico lineal
 - (c) Gráfico de barras verticales
 - (d) Gráfico de barras horizontales
14. ¿Qué tipo de gráfico utilizaría para representar la distribución porcentual de las ventas de una empresa por tipo de producto y según tipo de cliente para el año 2012?
- (a) Gráfico de barras horizontales comparativas
 - (b) Barra 100%
 - (c) Gráfico de barras verticales compuestas
 - (d) Gráfico de barras horizontales compuestas
15. ¿Qué tipo de gráfico utilizaría para comparar el monto de las ventas de una empresa (en miles \$) por tipo de producto y según tipo de cliente para el año 2012?
- (a) Gráfico de barras horizontales comparativas
 - (b) Barra 100%
 - (c) Gráfico de barras verticales compuestas
 - (d) Gráfico de barras horizontales compuestas
16. ¿Qué tipo de gráfico utilizaría para comparar el monto de las ventas de una empresa (en miles \$) por tipo de producto para el periodo 2007 al 2012?
- (a) Gráfico de barras horizontales comparativas
 - (b) Gráfico de barras verticales comparativas
 - (c) Gráfico de barras verticales compuestas
 - (d) Gráfico de barras horizontales compuestas

17. El gerente de un centro de llamadas desea evaluar el desempeño de los agentes de servicio y para ello decide basarse en los tiempos de espera de los clientes para ser atendidos (medido en segundos) y el grado de satisfacción que los clientes manifiesten al recibir el servicio (valorado como bueno, regular o malo). La semana anterior tomó una muestra de 12 llamadas por día de lunes a miércoles y 20 llamadas el jueves y el viernes. Los siguientes son los tiempos de las muestras tomadas de lunes a miércoles. El lunes y el martes se tomaron tiempos de llamadas atendidas y el miércoles solo de llamadas no atendidas:

Día	Número de muestra (tiempo en segundos)											
	1	2	3	4	5	6	7	8	9	10	11	12
Lunes	13	15	15	16	16	16	16	17	17	19	11	21
martes	12	15	15	16	16	15	16	18	15	20	30	40
Miércoles	30	40	45	45	50	50	45	35	60	50	80	100

El tipo de gráfico apropiado para representar la distribución porcentual de las llamadas según el agente que contestó es:

- (a) Gráfico lineal
(b) Gráfico de barras verticales
(c) Gráficos de barras horizontales
(d) Gráfico circular
18. Con base en los datos de la pregunta 17, el tipo de gráfico apropiado para representar el número de llamadas recibidas por mes durante los últimos 12 meses es:
- (a) Gráfico lineal
(b) Diagrama de dispersión
(c) Gráficos de barras horizontales
(d) Barra 100%

CAPÍTULO 3

CAPÍTULO 3

ANÁLISIS DESCRIPTIVO DE LA INFORMACIÓN ESTADÍSTICA

OBJETIVOS

Al concluir el capítulo, será capaz de:

- Reconocer la importancia y utilidad de las medidas de posición central.
 - Calcular e interpretar las principales medidas de posición en datos no agrupados.
 - Calcular e interpretar las principales medidas de variabilidad en datos no agrupados.

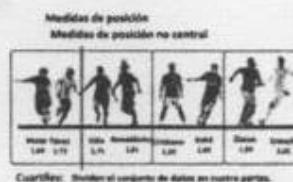
3.1 Medidas de posición

Las medidas de posición corresponden a estadísticos que se emplean para describir un cierto conjunto de datos, ya sea que correspondan a la población o a la muestra. De las medidas más conocidas está la media aritmética, o promedio, la cual es muy empleada para describir el comportamiento de los datos. Ahora bien, a parte del promedio existen muchas otras medidas, de las cuales se estudiarán a continuación algunas de las más importantes.

Se recomienda que antes de iniciar la lectura de este tema se estudie el apéndice de este capítulo, el cual ayudará a la comprensión de las fórmulas y los conceptos expuestos.

Material audiovisual

En la página de internet de este texto podrá encontrar una presentación y un video de diapositivas que exponen tanto los conceptos como los cálculos de este tema.



Solicite su usuario y contraseña escribiendo al correo electrónico info@auladeeconomia.com. Para ello deberá indicar el número de su factura de compra de texto.

3.2 Medidas de posición central

Cuando se desea describir un conjunto de datos resulta útil emplear medidas de posición central, tal como el promedio. Por ejemplo, suponga que en un aula hay un grupo de jóvenes de 15 a 18 años. Si se calcula la edad promedio, y da 16 años, una persona que no se encuentre en dicho sitio puede saber con facilidad qué tipo de personas hay en ese lugar (con respecto a su edad solamente), es decir, sabría que son adolescentes. Es por eso que se dice que un valor medio es, en cierto sentido, un valor representativo de un conjunto de datos.

MEDIDAS DE POSICIÓN CENTRAL: Son valores representativos de un conjunto de datos, de modo que cuando el conjunto de datos se ordena ascendente, estos valores tienden a localizarse hacia el centro del conjunto.

Las medidas de posición central más conocidas son la media aritmética o promedio, la mediana y la moda.

El objetivo básico de calcular estas medidas es poder reducir los datos, sustituyéndolos por algunas pocas medidas que los describan. Por supuesto que cada medida posee sus ciertas ventajas y desventajas que luego se mencionarán.

Material audiovisual

Antes de iniciar este tema recomendamos ver el video sobre notación de sumatoria en la página de internet de este texto.

Solicite su usuario y contraseña escribiendo al correo electrónico info@auladeeconomia.com. Para ello deberá indicar el número de su factura de compra de texto.

Series de datos

- Conjunto de valores:

12, 23, 18, 6, 21, 14, 12, 32

- Calcule las siguientes sumatorias:

$$\sum_{i=1}^8 x_i = 12 + 23 + 18 + 6 + 21 + 14 + 12 + 32 = 138$$

$$\sum_{i=3}^4 x_i = 18 + 6 + 21 + 14 = 59$$

La media aritmética

La media aritmética, también llamada media o promedio, se denota por m , cuando corresponde a una población y por \bar{x} cuando se refiere a una muestra. Se trata del valor medio de todos los valores que toma la variable estadística de una serie de datos. La media aritmética es la medida de posición de más importancia en estadística y la de más uso, principalmente por su fácil cálculo e interpretación.

MEDIA ARITMÉTICA: La media aritmética de un conjunto de datos es la suma de los datos entre el número de datos.

Para una serie de n valores de una variable $x_1, x_2, x_3, \dots, x_n$, es el cociente de dividir la sumatoria de todos los valores que toma la variable x_i , entre el número total de datos. Esto se puede expresar como una fórmula de la siguiente manera.

Media aritmética poblacional: Se representa con la letra griega m y se calcula como:

$$\mu = \frac{\sum_{i=1}^n X_i}{N}$$

Media aritmética muestral: Se representa como *equis barra* y se calcula como:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

Ejemplo

Suponga que se tienen los siguientes datos correspondientes a las ventas mensuales que ha realizado un vendedor durante los últimos siete meses (en millones de dólares):

20, 33, 42, 40, 19, 23, 28

Calcule la media aritmética.

Solución

El cálculo de la media sería:

$$\bar{x} = \frac{20 + 33 + 42 + 40 + 19 + 23 + 28}{7} = 29.29$$

Según ese resultado, sus ventas mensuales promedio son de 29.29 millones de dólares.

Ejercicio de revisión

Con base en el siguiente conjunto de datos, calcule la media aritmética:

40, 50, 65, 85, 75, 90, 60, 60, 60, 75

Ver solución de este ejercicio en la página de internet de este texto:
www.auladeeconomia.com/raeep.html

Uso de Excel y Minitab para el cálculo de la media aritmética

Utilice Excel y Minitab para resolver el ejercicio: Suponga que se tienen los siguientes datos correspondientes a las ventas mensuales que ha realizado un vendedor durante los últimos siete meses (en millones de dólares):

20, 33, 42, 40, 19, 23, 28

Calcule la media aritmética.

Solución

En **Excel**, se introducen los datos, bien sea en una fila o una columna. En este caso los datos se encuentran en el rango de celdas de A1 hasta A7:

	A	B
1	20	
2	33	
3	42	
4	40	
5	19	
6	23	
7	28	

Luego se elige una celda para determinar el resultado, por ejemplo la celda B8, y en ella se ingresa la siguiente función de Excel:

=PROMEDIO(A1:A7)

Luego se presiona Enter (o Intro) y se obtiene el resultado de 29,29, aproximadamente:

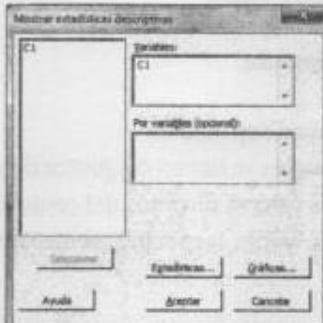
	A	B
1	20	Media =
2	33	29,2857143
3	42	
4	40	
5	19	
6	23	
7	28	
8	Media =	29,2857143

Estadísticas descriptivas

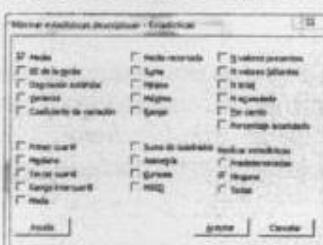
En **Minitab**, se introducen los datos en una columna, por ejemplo la columna C1:

	C1	C2
1	20	
2	33	
3	42	
4	40	
5	19	
6	23	
7	28	

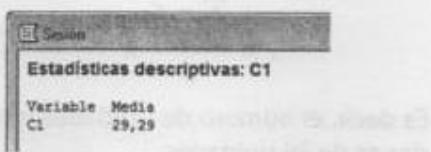
Luego se da clic en el menú Estadísticas, se elige Estadística básica y ahí selecciona Mostrar estadísticas descriptivas. Ahí completa el cuadro de diálogo seleccionando la variable, que en este caso se encuentra en la columna C1:



Luego en el botón Estadísticas selecciona la Media:



Luego de dar clic en Aceptar en cada cuadro, se obtiene el resultado en la ventana Sesión:



Material audiovisual

En la página de internet de este texto podrá encontrar videos que exponen el uso de Excel y Minitab para calcular estas medidas descriptivas.

Solicite su usuario y contraseña escribiendo al correo electrónico info@auladeconomia.com. Para ello deberá indicar el número de su factura de compra de texto.



Características principales de la media aritmética:

- Un conjunto de datos posee una media única, o sea, que un conjunto de datos no posee dos medias distintas simultáneamente.
- El valor de la media se obtiene a través de la suma de todos y cada uno de los valores de la serie de datos, por lo que depende de cada uno de ellos.
- Los valores extremos afectan fuertemente a la media aritmética, lo cual es su principal desventaja.
- La media resume o representa a un conjunto de datos a través de un solo valor.

La media ponderada

a. Caso de valores repetidos

Muchas veces se tienen conjuntos de datos con observaciones que se repiten varias veces. Si los valores distintos del conjunto son $x_1, x_2, x_3, \dots, x_k$, los cuales aparecen $f_1, f_2, f_3, \dots, f_k$ veces, respectivamente, entonces la media ponderada será:

$$\bar{x} = \frac{\sum_{i=1}^k x_i f_i}{n}$$

Ejemplo

Suponga que una empresa posee quince vendedores de un determinado producto. Cuatro de los vendedores lograron vender 50 unidades, 6 vendieron 40 unidades, tres vendieron 35 unidades y 2 vendieron 20 unidades. ¿Cuál es el número de unidades promedio de cada vendedor?

Solución

Dado que existen valores repetidos, entonces se aplica la fórmula:

$$\bar{x} = \frac{\sum_{i=1}^k x_i f_i}{n} = \frac{4 \cdot 50 + 6 \cdot 40 + 3 \cdot 35 + 2 \cdot 20}{15} = 39$$

Es decir, el número de unidades promedio vendidas por cada vendedor es de 39 unidades.

Ejercicio de revisión

En un muelle hay 20 contenedores que pesan 15 toneladas cada uno, 25 que pesan 20 toneladas cada uno y 10 que pesan 25 toneladas cada uno. ¿Cuál es el peso promedio de los contenedores?

Ver solución de este ejercicio en la página de internet de este texto:
www.auladeeconomia.com/raeep.html

b. Caso de valores con diferente importancia

En ocasiones es necesario calcular la media de una serie de valores $x_1, x_2, x_3, \dots, x_k$, pero estos valores tienen diferente importancia o peso relativo. Por ejemplo, cuando se calcula la nota final de un curso puede tener más peso la nota de un examen que la nota de una tarea. En estos casos es necesario determinar un factor de ponderación para cada observación $w_1, w_2, w_3, \dots, w_k$ respectivamente, el cual indica la importancia o peso relativo que se otorga a cada valor. De ese modo la media sería:

$$\bar{x} = \frac{\sum_{i=1}^k x_i w_i}{\sum_{i=1}^k w_i}$$

Ejemplo

Una empresa obtiene distintos márgenes de utilidad según los diferentes productos que vende. Suponiendo que vende 3 productos diferentes A, B y C, de acuerdo con los siguientes datos:

Producto	Margen de utilidad	Volumen de ventas (en millones de dólares)
A	20%	200
B	30%	100
C	40%	60
Total		\$360

¿Cuál es el margen de utilidad promedio?

Solución

Para responder a esta pregunta es necesario calcular la media ponderada, ya que el volumen de ventas de cada producto es distinto, y eso afecta al promedio. El cálculo debe ser el siguiente:

$$\bar{x} = \frac{\sum_{i=1}^k x_i w_i}{\sum_{i=1}^k w_i} = \frac{20\% \cdot 200 + 30\% \cdot 100 + 40\% \cdot 60}{360} = 26.11\%$$

El margen de utilidad promedio es de 26.11%. Obsérvese que los pesos (w_i) corresponde a las ventas de cada producto, y entonces se divide entre el total de ventas.

Ejercicio de revisión

En un curso universitario se realizan tres exámenes. El segundo examen tiene un valor que es el doble del primero y el tercer examen tiene un valor que es el triple del segundo. Si un estudiante obtiene una nota de 8 en el primer examen, un 9 en el segundo y un 6 en el tercero (todas estas notas están en una escala de 0 a 10), calcule su calificación promedio.

Ver solución de este ejercicio en la página de internet de este texto:
www.auladeeconomia.com/raeep.html

La media geométrica

En muchos casos no es apropiado determinar la media aritmética, por ejemplo, cuando se necesita calcular la tasa de crecimiento promedio del índice de precios al consumidor. En estas situaciones se debe calcular la media geométrica.

MEDIA GEOMÉTRICA: Es la raíz enésima del producto de las n observaciones que componen un conjunto de datos.

Algebraicamente se expresa: $Mg = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n}$

La media geométrica se emplea cuando se desea calcular el promedio de razones, tasas de variación promedio, tasas de crecimiento promedio, etc.

Ejemplo

El precio de un cierto producto se incrementó un 5,5% durante 1999, un 7,4% durante el 2000, un 3,7% en el 2001, un 9,85% en el 2002 y un 10% en el 2003. ¿Cuál ha sido el incremento promedio en el precio de ese producto?

Solución

Para responder a la pregunta conviene ordenar la información de la manera siguiente:

Año	Incremento porcentual	En forma decimal
1999	5,50%	1,055
2000	7,40%	1,074
2001	3,70%	1,037
2002	9,85%	1,0985
2003	10,00%	1,10

Aplicando la fórmula de la media geométrica:

$$Mg = \sqrt[5]{x_1 \cdot x_2 \cdot \dots \cdot x_5} = \sqrt[5]{1,055 \cdot 1,074 \cdot 1,037 \cdot 1,0985 \cdot 1,10} = 1,0726$$

Esto quiere decir que el incremento promedio del precio es de 7,26%.

Ejercicio de revisión

Un país tuvo una tasa de inflación de 5% durante el año 2009, un 4% en 2010, un 6% en 2011 y 3% en 2012. ¿Cuál es la tasa de inflación promedio en estos 4 años?

Ver solución de este ejercicio en la página de internet de este texto:
www.auladeeconomia.com/raeep.html

Uso de Excel y Minitab para calcular la media geométrica

Ejemplo

Utilice Excel y Minitab para calcular la media geométrica del conjunto de datos siguiente:

1,055 - 1,074 - 1,037 - 1,0985 - 1,10

Solución

En **Excel**, primero se introducen los datos en una fila o columna, por ejemplo, en la columna A, en el rango de celdas de A1 hasta A5:

	A	B
1	1,055	
2	1,074	
3	1,037	
4	1,0985	
5	1,1	

Luego, en la celda en la cual se desea el resultado, se introduce la siguiente función:

=MEDIA.GEOM(A1:A5)

Y así se obtiene el resultado de 1,0726.

En **Minitab**, primero se introducen los datos en una columna, por ejemplo, en la columna C1:

	C1	C2
1	1,0550	
2	1,0740	
3	1,0370	
4	1,0985	
5	1,1000	

Luego, en el menú Calc, se selecciona Calculadora y se completa el cuadro de diálogo:



Es necesario indicar dónde se desea almacenar el resultado, en este caso en la columna C2. Luego en el campo Expresión se indica la siguiente función:

GMEAN(C1)

Y así se obtiene el resultado de 1,0726 en la hoja de trabajo de Minitab, y no en la ventana Sesión.

Cuando los valores $x_1, x_2, x_3, \dots, x_k$, se repiten $f_1, f_2, f_3, \dots, f_k$ veces respectivamente, entonces la media geométrica sería:

$$Mg = \sqrt[k]{x_1^{f_1} \cdot x_2^{f_2} \cdot \dots \cdot x_k^{f_k}}$$

La moda

Cuando se está interesado en tener una cierta idea de dónde se concentra la mayor parte de los datos de una serie de valores, entonces resulta útil la moda. La moda, o modo, es simplemente la medida de posición que indica el valor que más se repite en el conjunto de datos, o sea, el valor que se presenta con más frecuencia en una serie de datos. La moda se designa como M_o .

MODA: La moda es el valor que más se repite en el conjunto de datos.

La moda se puede obtener por una simple observación de los datos, es decir, se determina con gran facilidad.

Un conjunto de datos podría no tener moda, si ningún valor aparece más de una vez, pero también puede tener una o varias modas. Cuando tiene una moda, se dice que es unimodal. Si el conjunto de datos posee dos modas, y entonces se dice que el conjunto es bimodal. También es posible que el conjunto tenga más de dos modas, y entonces se le llama multimodal.

Ejemplo

Con base en los siguientes conjuntos de datos, obtenga la moda:

Conjunto 1:

12, 14, 14, 15, 18, 18, 18, 22, 25

Conjunto 2:

12, 14, 14, 14, 15, 18, 18, 18, 22, 25

Conjunto 3:

12, 14, 15, 18, 22, 25

Conjunto 4:

12, 14, 14, 15, 18, 18, 18, 22, 2500

Solución

- El dato que más veces aparece es el 18, por tanto la moda es 18.
- El dato que más veces aparece es el 14 y el 18, por tanto la moda es 14 y 18. Este es un conjunto bimodal.
- No tiene moda.
- El dato que más veces aparece es el 18, por tanto la moda es 18. Observe que el valor extremo 2500 no afectó el resultado, pues el conjunto 1 y el 4 son iguales excepto por ese valor.

Características principales de la moda:

- A diferencia de la media aritmética, el valor de la moda no se afecta por la presencia de valores extremos.
- Cuando el conjunto de datos es pequeño o no ofrece una cierta tendencia central, entonces la moda podría ser un dato de poca utilidad.
- La moda se puede determinar en datos cualitativos y no solo cuantitativos, y en cualquier escala de medición (nominal, ordinal, de intervalo o de razón).
- Cuando se calcula en datos agrupados (distribución de frecuencias), no se obtiene con exactitud, sino solo una aproximación.

Ejercicio de revisión

Con base en el siguiente conjunto de datos:

40, 50, 65, 85, 75, 90, 60, 60, 60, 75

Calcule la moda o modo.

Ver solución de este ejercicio en la página de internet de este texto:
www.auladeeconomia.com/raeep.html

La mediana

La **mediana**, denotada Med , es una medida de posición que tiene diversas aplicaciones. Para calcularla es necesario ordenar el conjunto de datos en forma ascendente, pues la mediana divide el conjunto de datos en dos partes, de manera que haya igual cantidad de valores de un lado que de otro.

MEDIANA: Es una medida de posición que divide a la serie de valores en dos partes iguales, un cincuenta por ciento que es mayor o igual a la mediana y otro cincuenta por ciento que es menor o igual que ella.

Para encontrar la mediana en una serie de datos no agrupados, lo primero que se hace es ordenar los datos en una forma creciente y luego se ubica la posición que esta ocupa en esa serie de datos; para ello hay que determinar si la serie de datos es par o impar.

Si el número N es el número de datos de la serie, entonces la posición de la mediana será:

$$P_{Med} = \frac{N+1}{2}$$

Luego el número que se obtiene indica el lugar o posición que ocupa la mediana en la serie de valores, y entonces la mediana será el número que ocupe el lugar de la posición encontrada.

Cuando el número N de datos es una cantidad par, se requerirá obtener el punto medio para poder obtener la mediana.

Ejemplo

Sean los siguientes datos, 5, 12, 7, 8, 10, 6, y 9, los años de servicios de un grupo de trabajadores. ¿Cuál es la mediana?

Solución

Lo primero que se hace es ordenar los datos en forma creciente o decreciente:

$$5, 6, 7, 8, 9, 10, 12$$

Dado que se tienen 7 datos, una cantidad impar de datos, se aplica la fórmula:

$$P_{Med} = \frac{N+1}{2} = \frac{7+1}{2} = 4$$

Ese resultado indica que la mediana será el cuarto dato de la serie, es decir, la mediana será 8, $Med = 8$.

Ejemplo

Sean los siguientes datos, 5, 12, 7, 8, 11, 10, 6, y 9, los años de servicios de un grupo de trabajadores. ¿Cuál es la mediana?

Solución

Lo primero que se hace es ordenar los datos en forma creciente o decreciente:

5, 6, 7, 8, 9, 10, 11, 12

Dado que se tienen 8 datos, una cantidad par de datos, se aplica la fórmula:

$$P_{Med} = \frac{N+1}{2} = \frac{8+1}{2} = 4.5$$

Ese resultado indica que la mediana estará entre el cuarto y el quinto dato de la serie, y por tanto será necesario calcular el punto medio entre 8 y 9, es decir, la mediana será $(8+9)/2$, $Med = 8.5$.

Ejercicios de revisión

Con base en los siguientes conjuntos de datos:

Conjunto 1: 40, 50, 65, 85, 75, 90, 60, 60, 60, 75

Conjunto 2: 85, 110, 125, 130, 90, 100, 140

Calcule la mediana en cada caso.

Ver solución de este ejercicio en la página de internet de este texto:
www.auladeeconomia.com/raep.html

Características principales de la mediana:

- Al igual que la moda, la mediana no se afecta por la presencia de valores extremos en la serie de datos.
- Cuando se calcula en datos agrupados (distribución de frecuencias), no se obtiene con exactitud, sino solo una aproximación.
- La suma de los valores absolutos de las desviaciones de los datos individuales con respecto a la mediana siempre es mínima, lo cual la hace útil para resolver ciertos problemas.

Interpretación y uso de las medidas de posición central

Con respecto a las medidas de posición central estudiadas es posible realizar las siguientes observaciones:

- En grupos pequeños la moda puede ser una medida muy inestable, ya que al cambiar alguno de los datos, la moda podría variar drásticamente.

- La moda y la mediana puede ser consideradas medidas más robustas que la media aritmética, ya que no se ven afectas por los valores extremos, mientras que la media sí. Cuando hay valores extremos, generalmente es mejor usar la mediana que la media.
- La media aritmética es la medida de posición central que es más estable de muestra a muestra.
- Cuando los datos se han obtenido aleatoriamente, la media aritmética tiende a ser el mejor descriptor del conjunto de valores original.

En ciertos conjuntos de datos la media, la mediana y la moda tienden a ser iguales o muy similares. Esto ocurre en la medida que la distribución de los datos sea simétrica, en cuyo caso serían las tres medidas iguales.

SIMETRÍA: Una distribución es simétrica si existe la misma cantidad de valores a la derecha que a la izquierda de la media y, por tanto, hay igual cantidad de desviaciones positivas como negativas.

La existencia de valores extremos en un solo sentido, provocando una asimetría, afecta principalmente al promedio. Entonces, el promedio será mayor que la mediana y la mediana mayor que la moda.

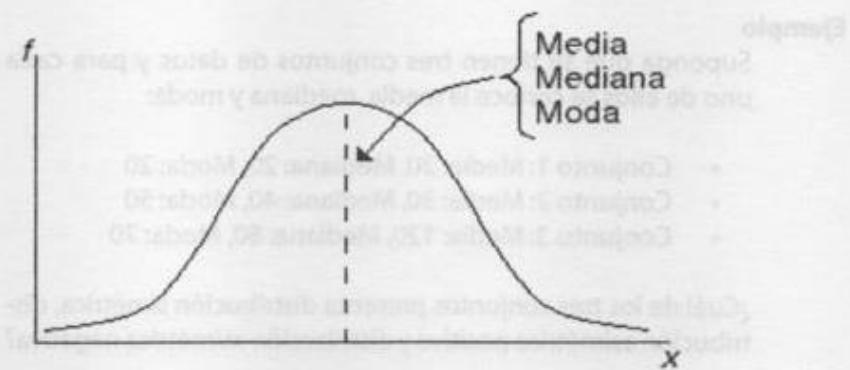
ASIMETRÍA POSITIVA: Una distribución es asimétrica positiva, o a la derecha, si la cola derecha con respecto a la media es más larga que la de la izquierda.

Esta asimetría puede ser positiva, cuando los valores extremos son mayores a la media, o negativa, cuando sean menores. Entonces, el promedio será menor que la mediana y la mediana menor que la moda.

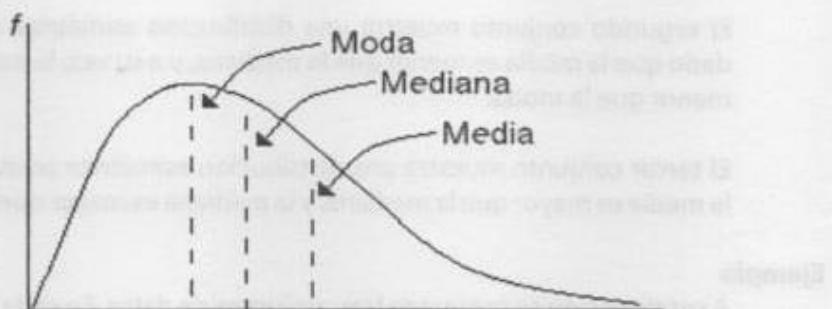
ASIMETRÍA NEGATIVA: Una distribución es asimétrica negativa, o a la izquierda, si la cola con respecto a la media es más larga que la de la derecha.

Existen medidas sobre el grado de simetría o asimetría de una distribución, las cuales no se consideran en este texto, sin embargo, es posible mostrar esas simetrías o asimetrías en forma gráfica, y la relación que sea con la moda, la mediana y la media del modo siguiente.

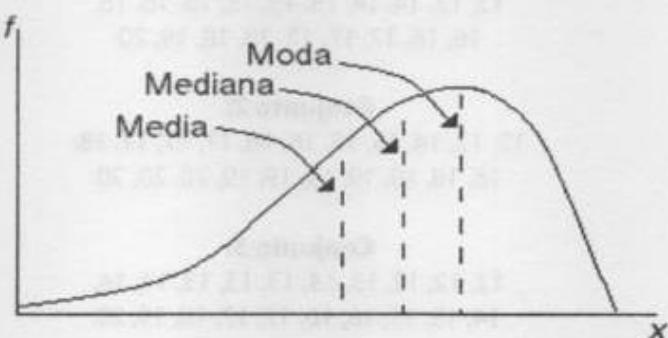
En una distribución simétrica, la media, la moda y la mediana serán iguales. En una distribución asimétrica positiva, la media será mayor que la mediana, y a su vez, la mediana será mayor que la moda. En una distribución asimétrica negativa, la media será menor que la mediana, y a su vez, la mediana será menor que la moda.



Simétrica



Asimétrica positiva



Asimétrica negativa

Ejemplo

Suponga que se tienen tres conjuntos de datos y para cada uno de ellos se conoce la media, mediana y moda:

- Conjunto 1: Media: 20, Mediana: 20, Moda: 20
- Conjunto 2: Media: 30, Mediana: 40, Moda: 50
- Conjunto 3: Media: 120, Mediana: 80, Moda: 70

¿Cuál de los tres conjuntos presenta distribución simétrica, distribución asimétrica positiva y distribución asimétrica negativa?

Solución

El primer conjunto presenta una distribución simétrica, pues la media, la moda y la mediana son todas iguales.

El segundo conjunto muestra una distribución asimétrica negativa, dado que la media es menor que la mediana, y a su vez, la mediana es menor que la moda.

El tercer conjunto muestra una distribución asimétrica positiva, pues la media es mayor que la mediana, y la mediana es mayor que la moda.

Ejemplo

A continuación se presentan tres conjuntos de datos. En cada caso, calcule la media aritmética, la mediana y la moda del siguiente conjunto de datos:

Conjunto 1:

12, 13, 14, 14, 15, 15, 15, 16, 16, 16,
16, 16, 17, 17, 17, 18, 18, 19, 20

Conjunto 2:

12, 13, 14, 15, 15, 16, 16, 17, 17, 17, 18,
18, 18, 19, 19, 19, 19, 19, 20, 20, 20

Conjunto 3:

12, 12, 12, 13, 13, 13, 13, 13, 14, 14,
14, 15, 15, 16, 16, 17, 17, 18, 19, 20

Además en cada caso, construya una gráfica (histograma) para representar a este conjunto de datos.

Observe la gráfica y la relación entre la media, la mediana y la moda. ¿Qué puede decirse de la simetría o asimetría de cada conjunto?

Solución

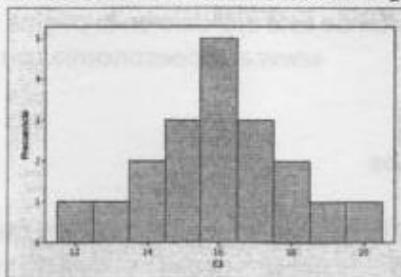
Conjunto 1:

Media = 16

Mediana = 16

Moda = 16

La media, la mediana y la moda son iguales, lo que indica una distribución simétrica, lo cual se observa claramente en la gráfica siguiente.



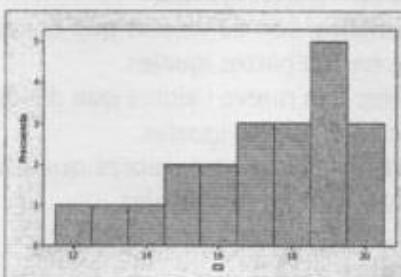
Conjunto 2:

Media = 17,19

Mediana = 18

Moda = 19

La media es menor que la mediana, y a su vez la mediana es menor que la moda, lo que indica una distribución asimétrica negativa, lo cual se observa claramente en la gráfica siguiente.



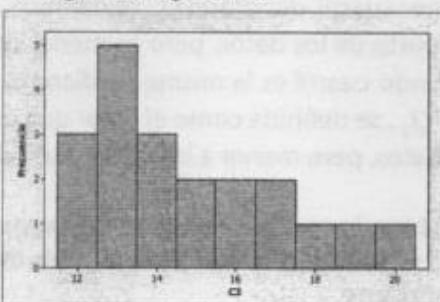
Conjunto 3:

Media = 14,8

Mediana = 14

Moda = 13

La media es mayor que la mediana, y a su vez la mediana es mayor que la moda, lo que indica una distribución asimétrica positiva, lo cual se observa claramente en la gráfica siguiente.



Ejercicio de revisión

Con base en el siguiente conjunto de datos:

40, 50, 65, 85, 75, 90, 60, 60, 60, 75

Determine qué tipo de simetría o asimetría se presenta.

Ver solución de este ejercicio en la página de internet de este texto:

www.auladeeconomia.com/raeep.html

3.3 Los cuantílos

Los cuantílos son medidas de posición que dividen al conjunto ordenado de datos en una determinada cantidad de partes iguales, de tal manera que superan a una cierta proporción de las observaciones, pero son superados por la proporción complementaria.

Las principales de estas medidas son:

- **Cuartiles:** Son tres valores que dividen al conjunto ordenado de datos en 4 partes iguales.
- **Percentiles:** Son 99 valores que dividen al conjunto ordenado de datos en 100 partes iguales.
- **Deciles:** Son nueve valores que dividen al conjunto ordenado de datos en 10 partes iguales.
- **Quintiles:** Son cuatro valores que dividen al conjunto ordenado de datos en 5 partes iguales.

Cuartiles o cuartilos

Cada serie de datos posee tres cuartilos, los cuales dividen al conjunto de datos en cuatro partes iguales.

CUARTIL: Los cuartiles son los tres valores que dividen a un conjunto ordenado de datos en cuatro partes iguales.

El primer cuartil, denotado Q_1 , se definiría como aquel valor que supera a la cuarta parte de los datos, pero es menor que las restantes tres cuartas partes. El segundo cuartil es la misma mediana. A su vez, el tercer cuartil, denotado como Q_3 , se definiría como el valor que es mayor que las tres cuartas partes de los datos, pero menor a la cuarta parte de ellos.

Ahora bien, los cuartiles pueden ser expresados como percentiles 25 y 75, y por tanto ser calculados usando la fórmula de los percentiles, tal como se expone más adelante.

Deciles

Cada serie de datos posee nueve deciles, los cuales dividen al conjunto de datos en diez partes iguales.

DECIL: Los deciles son los nueve valores que dividen a un conjunto ordenado de datos en diez partes iguales.

El primer decil, denotado D_1 , se definiría como aquel valor que supera a la décima parte de los datos, pero es menor que las restantes nueve décimas partes. El segundo decil es aquel valor que supera a dos décimas partes de los datos, pero es menor que las restantes ocho décimas partes. Así sucesivamente se pueden definir los demás deciles.

Ahora bien, los deciles pueden ser expresados como percentiles 10, 20, ..., y 90, y por tanto ser calculados usando la fórmula de los percentiles, tal como se expone más adelante.

Quintiles

Cada serie de datos posee cuatro quintiles, los cuales dividen al conjunto de datos en cinco partes iguales.

QUINTIL: Los quintiles son los cuatro valores que dividen a un conjunto ordenado de datos en cinco partes iguales.

El primer quintil, denotado K_1 , se definiría como aquel valor que supera a la quinta parte de los datos, pero es menor que las restantes cuatro quintas partes. El segundo quintil es aquel valor que supera a dos quintas partes de los datos, pero es menor que las restantes tres quintas partes. Así sucesivamente se pueden definir los demás quintiles.

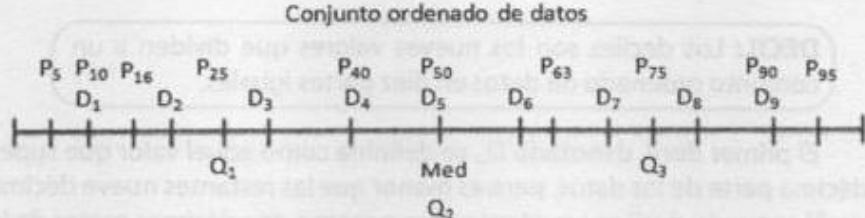
Ahora bien, los quintiles pueden ser expresados como percentiles 20, 40, 60 y 80, y por tanto ser calculados usando la fórmula de los percentiles, tal como se expone más adelante.

Percentiles

Una serie de datos tiene 99 percentiles, los cuales dividen a un conjunto de datos en 100 partes iguales. Los percentiles son un concepto similar al de los cuartiles. Por ejemplo, el percentil 57, denotado por P_{57} , corresponderá a aquel valor tal que es superior al 57% de las observaciones y menor al restante 43%.

PERCENTIL Los percentiles son los 99 valores que dividen a un conjunto ordenado de datos en 100 partes iguales.

Como se mencionó anteriormente, los cuartiles, deciles y quintiles pueden ser expresados en términos de percentiles, como se ilustra a continuación:



Relaciones entre cuantílos

Cuartil Percentil

Cuartil	Percentil
Q_1	P_{25}
Q_3	P_{75}

$$\text{Cuartil} \times 25 = \text{Percentil}$$

Decil Percentil

Decil	Percentil
D_1	P_{10}
D_2	P_{20}
D_3	P_{30}
D_4	P_{40}
D_5	P_{50}
D_6	P_{60}
D_7	P_{70}
D_8	P_{80}
D_9	P_{90}

$$\text{Decil} \times 10 = \text{Percentil}$$

Quintil Percentil

Quintil	Percentil
K_1	P_{20}
K_2	P_{40}
K_3	P_{60}
K_4	P_{80}

$$\text{Quintil} \times 20 = \text{Percentil}$$

Recuerde que para calcular mediana, cuartiles, percentiles, quintiles y deciles, los datos deben ordenarse ascendenteamente.

Para datos sin agrupar, en un conjunto de datos ordenados de acuerdo con su magnitud, el percentil m , denotado P_m , será el, $\left[\frac{m}{100}(n+1)\right]$ término.

Ejemplo

El número de unidades de un cierto producto vendidas por 10 vendedores el mes pasado son:

120, 100, 20, 70, 100, 140, 120, 150, 100, 40

Determine el primer cuartil, el tercer cuartil, el decil 4 y el 80º percentil

Solución

Primer cuartil:

Paso 1: El primer paso es ordenar la serie de datos:

20, 40, 70, 100, 100, 100, 120, 120, 140, 150

Paso 2: El primer cuartil equivale al percentil 25, por lo que $m = 25$ y se tienen 10 datos, por lo que $n = 10$. Entonces se sustituye en la fórmula:

$$P_m = \left[\frac{m}{100} (n + 1) \right] = \left[\frac{25}{100} (10 + 1) \right] = 2,75$$

Paso 3: La fórmula anterior no da el valor del percentil, sino que da la posición del percentil 25. Hay que buscar el dato en la posición 2,75. Como no se tiene un valor en la posición 2,75, quiere decir que el valor del percentil va a estar entre el segundo valor y el tercero, entonces se realiza una interpolación. Esto es, se toma el segundo dato en la serie ordenada, que es 40, y se le suma el producto de la parte decimal del resultado de la fórmula, que es 0,75 por la diferencia entre el segundo y el tercer dato, que es $70 - 40 = 30$. O sea, el percentil equivale a:

$$Q_1 = P_{25} = 40 + 0,75 * 30 = 62,5$$

Tercer cuartil:

Paso 1: El primer paso es ordenar la serie de datos:

20, 40, 70, 100, 100, 100, 120, 120, 140, 150

Paso 2: El tercer cuartil equivale al percentil 75, por lo que $m = 75$ y se tienen 10 datos, por lo que $n = 10$. Entonces se sustituye en la fórmula:

$$P_m = \left[\frac{m}{100} (n + 1) \right] = \left[\frac{75}{100} (10 + 1) \right] = 8,25$$

Paso 3: La fórmula anterior no da el valor del percentil, sino que da la posición del percentil 75. Hay que buscar el dato en la posición 8,25. Como no se tiene un valor en la posición 8,25, quiere decir que el valor del percentil va a estar entre el octavo valor y el noveno, entonces se realiza una interpolación. Esto es, se toma el octavo dato en la serie ordenada, que es 120, y se le suma el producto de la parte decimal del resultado de la fórmula, que es 0,25 por la diferencia entre el octavo y el noveno dato, que es $140 - 120 = 20$. O sea, el percentil equivale a:

$$Q_3 = P_{75} = 120 + 0,25 * 20 = 125$$

Decil 4:

Paso 1: El primer paso es ordenar la serie de datos:

20, 40, 70, 100, 100, 100, 120, 120, 140, 150

Paso 2: El decil 4 equivale al percentil 40, por lo que $m = 40$ y se tienen 10 datos, por lo que $n = 10$. Entonces se sustituye en la fórmula:

$$P_m = \left[\frac{m}{100} (n + 1) \right] = \left[\frac{40}{100} (10 + 1) \right] = 4,4$$

Paso 3: La fórmula anterior no da el valor del percentil, sino que da la posición del percentil 40. Hay que buscar el dato en la posición 4,4. Como no se tiene un valor en la posición 4,4, quiere decir que el valor del percentil va a estar entre el cuarto valor y el quinto, entonces se realiza una interpolación. Esto es, se toma el cuarto dato en la serie ordenada, que es 100, y se le suma el producto de la parte decimal del resultado de la fórmula, que es 0,4 por la diferencia entre el cuarto y el quinto dato, que es $100 - 100 = 0$. O sea, el percentil equivale a:

$$D_4 = P_{40} = 100 + 0,4 * 0 = 100$$

Percentil 80:

Paso 1: El primer paso es ordenar la serie de datos:

20, 40, 70, 100, 100, 100, 120, 120, 140, 150

Paso 2: Se desea calcular el percentil 80, por lo que $m = 80$ y se tienen 10 datos, por lo que $n = 10$. Entonces se sustituye en la fórmula:

$$P_m = \left[\frac{n}{100} (n+1) \right] = \left[\frac{80}{100} (10+1) \right] = 8,8$$

Paso 3: La fórmula anterior no da el valor del percentil, sino que da la posición del percentil 80. Hay que buscar el dato en la posición 8,8. Como no se tiene un valor en la posición 8,8, quiere decir que el valor del percentil va a estar entre el octavo valor y el noveno, entonces se realiza una interpolación. Esto es, se toma el octavo dato en la serie ordenada, que es 120, y se le suma el producto de la parte decimal del resultado de la fórmula, que es 0,8 por la diferencia entre el octavo y el noveno dato, que es $140 - 120 = 20$. O sea, el percentil equivale a:

$$P_{80} = 120 + 0,8 * 20 = 136$$

Ejercicio de revisión

Con base en el siguiente conjunto de datos:

40, 50, 65, 85, 75, 90, 60, 60, 60, 75

Calcule el primer cuartil, el tercer cuartil, el decil 4, el quintil 3 y el percentil 65.

Ver solución de este ejercicio en la página de internet de este texto:

www.auladeeconomia.com/raeep.html

Uso de Excel y Minitab para el cálculo de percentiles

Ejemplo

Utilice Excel y Minitab para resolver el ejercicio: El número de unidades de un cierto producto vendidas por 10 vendedores el mes pasado son:

20, 40, 70, 100, 100, 100, 120, 120, 140, 150

Determine el primer cuartil, el tercer cuartil y el 80º percentil.

Solución

En **Excel**, primero se introducen los datos en una fila o columna, por ejemplo, en la columna A, en el rango de celdas de A1 hasta A10:

	A	B
1	20	
2	40	
3	70	
4	100	
5	100	
6	100	
7	120	
8	120	
9	140	
10	150	

Para obtener el primer cuartil, en la celda en la cual se desea el resultado, se introduce la siguiente función:

=CUARTIL(A1:A10;1)

En la función, dentro del paréntesis, primero se indica el rango de datos, y luego (generalmente separado por punto y coma) se indica el número del cuartil, que en este caso es 1. Y así se obtiene el resultado de 77,5. Este resultado cambia con respecto al anterior, calculado manualmente, porque se está empleando un algoritmo distinto para el cálculo del cuartil. Esto mismo sucederá en los cálculos siguientes.

Para obtener el tercer cuartil, en la celda en la cual se desea el resultado, se introduce la siguiente función:

=CUARTIL(A1:A10;3)

En la función, dentro del paréntesis, primero se indica el rango de datos, y luego (generalmente separado por punto y coma) se indica el número del cuartil, que en este caso es 3. Y así se obtiene el resultado de 120.

Para obtener el percentil 80, en la celda en la cual se desea el resultado, se introduce la siguiente función:

=PERCENTIL(A1:A10;0,80)

En la función, dentro del paréntesis, primero se indica el rango de datos, y luego (generalmente separado por punto y coma) se indica el número del percentil, pero indicado en forma decimal, que en este caso es 0,80. Y así se obtiene el resultado de 124.

En **Minitab**, primero se introducen los datos en una columna, por ejemplo, en la columna C1:

	C1	C2
1	20	
2	40	
3	70	
4	100	
5	100	
6	100	
7	120	
8	120	
9	140	
10	150	

Luego se da clic en el menú Estadísticas, se elige Estadística básica y ahí selecciona Mostrar estadísticas descriptivas. Ahí completa el cuadro de diálogo seleccionando la variable, que en este caso se encuentra en la columna C1:

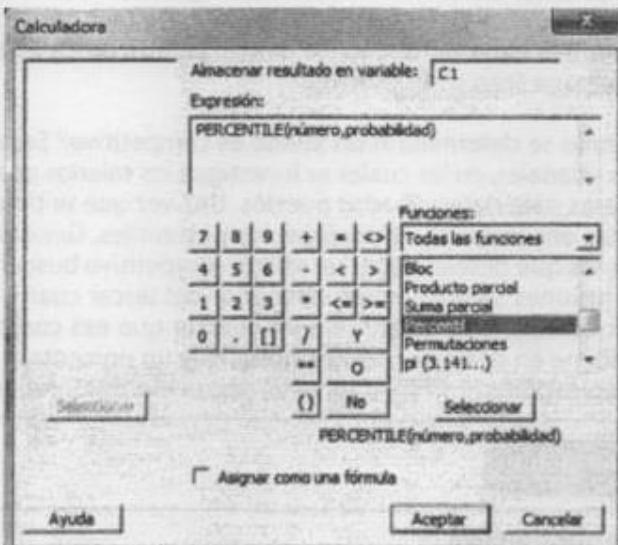


Luego en el botón Estadísticas selecciona primer cuartil y tercer cuartil. Luego de dar clic en Aceptar en cada cuadro, se obtiene el resultado en la ventana Sesión:

Sesión		
Estadísticas descriptivas: C1		
Variable	Q1	Q3
C1	62,5	125,0

Este resultado cambia con respecto al anterior, calculado manualmente y al obtenido en Excel, porque se está empleando un algoritmo distinto para el cálculo del cuartil.

Para obtener el percentil se da clic en el menú Calc y se selecciona Calculadora. Ahí se completa el cuadro de diálogo siguiente:



Se debe indicar en cuál columna se almacenará el resultado, por ejemplo en la columna C2. Luego en expresión debe seleccionarse la función:

PERCENTILE(número,probabilidad)

En esta función número corresponde a la columna que almacena los datos, en este caso C1, y probabilidad es el número del percentil expresado en forma decimal, que sería 0,80:

PERCENTILE(C1,0,80)

Observe que los datos de entrada de la función se separan por medio de un punto. Luego se da clic en Aceptar y el resultado se obtiene en la hoja de trabajo, no en la sesión. Según Minitab el percentil 80 es 136.

Aplicación

Administración de salarios

Muchas empresas emplean distintas técnicas estadísticas para diseñar sus políticas salariales, y, en este sentido, los percentiles son bastante utilizados.

El objetivo de una política salarial es poder dar una remuneración a los empleados de manera que se incentive el buen desempeño y se puede retener a los mejores colaboradores en la empresa, a la vez que no se excedan los costos que sean sostenibles para la empresa. Para esto es necesario diseñar un sistema que satisfaga criterios de equidad, a lo interno (para que puestos del mismo nivel tengan los mismos salarios, por ejemplo) y a lo externo (de manera que los salarios sean competitivos, o sea, que no sean muy bajos con respecto a lo que pagan otras empresas). Con relación a este último punto, se puede decir que a una empresa no le interesa pagar los salarios más altos de su industria, pues eso elevaría mucho los costos. Lo que busca es que los salarios que paga sean competitivos.

¿Cómo se determina si un salario es competitivo? Se realizan encuestas salariales, en las cuales se investigan los salarios que pagan las empresas para determinados puestos. Una vez que se tienen los datos de estas encuestas se determinan los percentiles. Generalmente las empresas que desean pagar un salario competitivo buscan que sus remuneraciones se encuentren alrededor del tercer cuartil o del percentil 80. Un salario así le indica al trabajador que esa compensación no la obtiene en cualquier otra empresa, que un porcentaje muy alto de los empleadores (un 75% o un 80%) pagan montos inferiores.

Aplicación

Antropometría

Cuando nace un niño, el médico o la enfermera mide su talla, la circunferencia cefálica, entre otras dimensiones. Conforme crecemos en algún momento también nos sometemos a algunas mediciones similares relacionadas con nuestro cuerpo. Cuando usamos un automóvil, los ingenieros emplearon datos sobre las medidas de los seres humanos para determinar las dimensiones de distintas partes del vehículo, como los asientos, la distancia entre el asiento y el volante, entre otros. De igual manera en el diseño de vestuarios, máquinas, edificios y muchos otros artículos. Así, la antropometría ha llegado a ser importantísima en muchas áreas.

La antropometría es la ciencia que se encarga del estudio de las dimensiones del cuerpo humano. Para ello se realizan estudios estadísticos con la finalidad de poder determinar las medidas promedio y diferentes percentiles. Es importante saber si la talla de un niño se encuentra

dentro de los parámetros considerados como normales, los cuales se definen empleando, generalmente los percentiles 5 y 95.

La antropometría ha desarrollado una serie de procedimientos para realizar las mediciones de forma adecuada, además de cuestionarse sobre cuáles son las mediciones que vale la pena realizar, o sea, aquellas que vayan a ser de utilidad en el ciencias biomédicas, la ingeniería, la arquitectura, etc. Así, se estudian medidas estructurales, tales como la cabeza, las extremidades y el tronco, y otras son medidas funcionales, como lo son movimientos específicos realizados por el cuerpo de las personas.

Gráfica de caja, rango intercuartil y desviación cuartil

La gráfica de caja es un gráfico bastante útil para explorar y visualizar el comportamiento de un conjunto de datos. Esta gráfica está compuesta por un rectángulo, llamado caja, y por dos líneas llamadas brazos o bigotes, por eso también se le conoce a esta gráfica como diagrama de caja y bigotes.

La caja de la gráfica se construye con base en el primer cuartil Q_1 , y el tercer cuartil Q_3 , los cuales definen los límites de la caja. La caja tiene una línea que la cruza que corresponde a la mediana. Los bigotes salen de la caja y llegan hasta el valor mínimo y máximo, o bien, dos valores a y b que se trazan hasta 1,5 veces la altura de la caja, o sea, que el bigote inferior llega hasta $a = Q_1 - 1,5(Q_3 - Q_1)$ y el bigote superior llega hasta $b = Q_3 + 1,5(Q_3 - Q_1)$. Otros autores trazan los bigotes hasta dos veces el rango intercuartil ($Q_3 - Q_1$), o sea, que el bigote inferior llega hasta $Q_1 - 2(Q_3 - Q_1)$ y el bigote superior llega hasta $Q_3 + 2(Q_3 - Q_1)$. También hay autores que trazan simplemente los bigotes hasta el mínimo y el máximo. En este texto se empleará la primera forma, ya que así lo construye el software Minitab. Los valores que quedan fuera de esos límites establecidos por los bigotes se consideran valores extremos o atípicos.

La gráfica de caja es útil para valorar la existencia de valores extremos o atípicos y para analizar la simetría o asimetría de la distribución de los datos.

Ejemplo

Se tiene un conjunto de datos con respecto al cual se conoce la siguiente información:

Primer cuartil: 20

Tercer cuartil: 36

Mediana: 30

Mínimo: 8

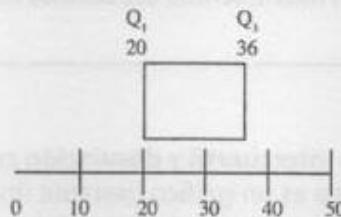
Máximo: 42

Construya la gráfica de caja.

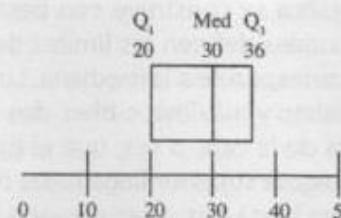
Solución

La gráfica de caja puede construir horizontal o vertical. En este caso se va a hacer horizontal, por lo que se construye un eje horizontal. Luego se realizan los siguientes pasos:

Paso 1: Determinar los cuartiles. La caja queda delimitada por el primer cuartil que es 20 y el tercer cuartil que es 36. En este caso ya están calculados, pero de otro modo habría que calcularlos, por lo que se dibuja la caja, la cual inicial en el primer cuartil y finaliza en el tercer cuartil:



Paso 2: Determinar la mediana. En este caso ya está calculada la mediana. Si no, se calcula. Entonces se traza la línea que representa la mediana, la cual es 30.



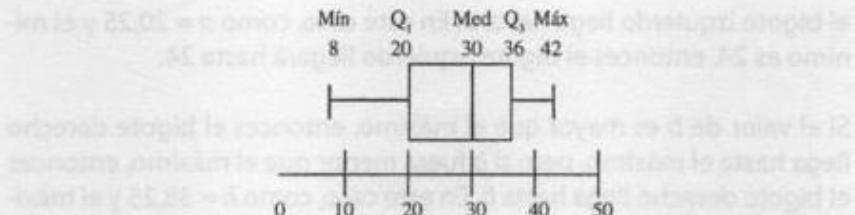
Paso 3: Determinación de los bigotes. Se calculan los valores a y b :

$$a = Q_1 - 1,5(Q_3 - Q_1) = 20 - 1,5(36 - 20) = -4$$
$$b = Q_3 + 1,5(Q_3 - Q_1) = 36 + 1,5(36 - 20) = 60$$

Si el valor de a es menor que el mínimo, entonces el bigote izquierdo llega hasta el mínimo, pero si a fuera mayor que el mínimo, entonces el bigote izquierdo llega hasta a . En este caso, como $a = -4$ y el mínimo es 8, entonces el bigote izquierdo llegará hasta 8.

Si el valor de b es mayor que el máximo, entonces el bigote derecho llega hasta el máximo, pero si b fuera menor que el máximo, entonces el bigote derecho llega hasta b . En este caso, como $b = 60$ y el máximo es 42 entonces el bigote derecho llegará hasta 42.

Finalmente se traza el brazo o bigote izquierdo, el cual parte de la caja hasta el punto mínimo, que es 8, y se traza el brazo o bigote derecho, el cual parte de la caja hasta el punto máximo, que es 42. No hay valores atípicos en este caso.



Ejemplo

Se tiene el siguiente conjunto de datos:

24, 25, 26, 28, 29, 30, 30, 30, 31, 31, 32, 35, 40

Construya la gráfica de caja usando Minitab.

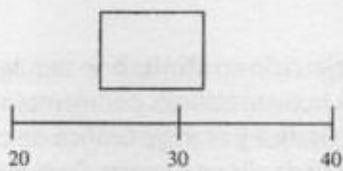
Solución

Paso 1: Determinar los cuartiles. Se calculan los dos cuartiles:

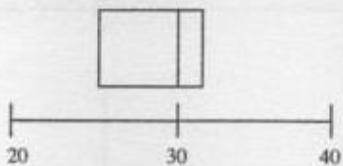
$$Q_1 = 27$$

$$Q_3 = 31,5$$

La caja queda delimitada por el primer cuartil que es 27 y el tercer cuartil que es 31,5, por lo que se dibuja la caja, la cual inicial en el primer cuartil y finaliza en el tercer cuartil:



Paso 2: Determinar la mediana. Se calcula la mediana, la cual es 30. Entonces se traza la línea que representa la mediana, la cual es 30.



Paso 3: Determinación de los bigotes. Se calculan los valores a y b :

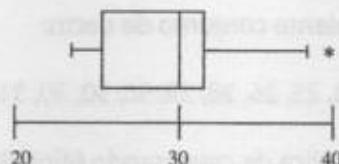
$$a = Q_1 - 1,5(Q_3 - Q_1) = 27 - 1,5(31,5 - 27) = 20,25$$

$$b = Q_3 + 1,5(Q_3 - Q_1) = 31,5 + 1,5(31,5 - 27) = 38,25$$

Si el valor de a es menor que el mínimo, entonces el bigote izquierdo llega hasta el mínimo, pero si a fuera mayor que el mínimo, entonces

el bigote izquierdo llega hasta a . En este caso, como $a = 20,25$ y el mínimo es 24, entonces el bigote izquierdo llegará hasta 24.

Si el valor de b es mayor que el máximo, entonces el bigote derecho llega hasta el máximo, pero si b fuera menor que el máximo, entonces el bigote derecho llega hasta b . En este caso, como $b = 38,25$ y el máximo es 40, entonces el bigote derecho llegará hasta 38,25 y el valor de 40 se marcará con un asterisco, pues se considera como valor atípico.



Uso de Minitab para construir una gráfica de caja

Ejemplo

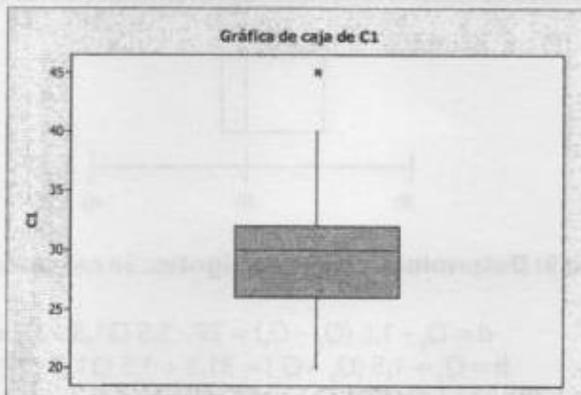
Se tiene el siguiente conjunto de datos:

20, 24, 25, 26, 28, 29, 30, 30, 30, 31, 31, 32, 35, 40, 45

Construya la gráfica de caja usando Minitab.

Solución

Para realizar este ejercicio en **Minitab** se requiere introducir los datos en una columna de la hoja de trabajo, por ejemplo en la columna C1. Luego se da clic al menú **Grafica** y se elige **Gráfica de caja**. En el cuadro se escoge **Una Y Simple**, y se da clic en **Aceptar**. En el cuadro de diálogo que aparece se selecciona la columna que contiene los datos, que en este caso es la columna C1, y se da clic en **Aceptar**. Se obtiene la gráfica siguiente:



Como se observa, Minitab hace la gráfica vertical y no horizontal como se expuso en el ejemplo anterior, sin embargo representa los mismos datos.

La gráfica de caja puede facilitar la identificación de valores atípicos o valores extremos, es decir, observaciones inusualmente grandes o pequeñas con respecto a los demás datos del conjunto de valores. Los paquetes de cómputo como Minitab trazan los valores atípicos como asteriscos, los cuales se van a observar ubicados más allá de los bigotes.

Ejercicios de revisión

Con base en el siguiente conjunto de datos:

40, 50, 65, 85, 75, 90, 60, 60, 75

Construya la gráfica de caja.

Ver solución de este ejercicio en la página de internet de este texto:
www.auladeeconomia.com/raeep.html

Ejemplo

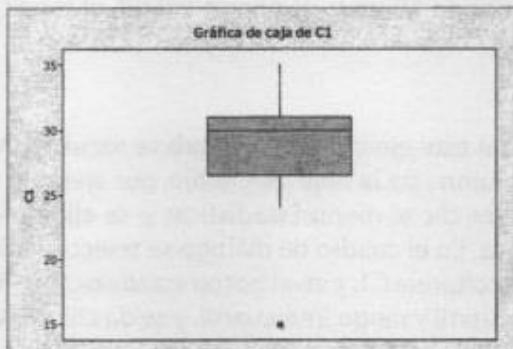
Se tiene el siguiente conjunto de datos:

15, 24, 26, 28, 29, 30, 30, 30, 31, 31, 32, 35

Construya la gráfica de caja usando Minitab e identifique la presencia de valores atípicos.

Solución

Para realizar este ejercicio en **Minitab** se requiere introducir los datos en una columna de la hoja de trabajo, por ejemplo en la columna C1. Luego se da clic al menú **Grafica** y se elige **Gráfica de caja**. En el cuadro se escoge **Una Y Simple**, y se da clic en **Aceptar**. En el cuadro de diálogo que aparece se selecciona la columna que contiene los datos, que en este caso es la columna C1, y se da clic en **Aceptar**. Se obtiene la gráfica siguiente:



Minitab ha dibujado la gráfica, pero ha colocado en la parte inferior un asterisco, el cual representa un valor atípico, o sea, un valor muy grande o muy pequeño con respecto a los demás datos del conjunto.

La gráfica de caja también es útil porque muestra, en cierto grado, la dispersión del conjunto de datos, pues se puede observar la diferencia entre el valor máximo y el mínimo, que se conoce como el **rango, recorrido o amplitud** del conjunto de datos. También, en este mismo sentido, muestra el rango intercuartil o rango intercuartílico (*RIC*), el cual corresponde a la diferencia entre el tercer cuartil y el primer cuartil:

$$RIC = Q_3 - Q_1$$

El **rango intercuartil** indica el rango en el cual se concentra la mitad de los datos alrededor de su mediana, por lo que un rango intercuartil pequeño indicaría que ese 50% de los datos se aleja poco de su mediana, mientras que un rango intercuartil elevado indicaría que esos valores que se hallan entre el primer cuartil y el tercer cuartil fluctúan en un rango grande. En la gráfica de caja, el rango intercuartil se observa en la longitud de la caja (o su altura, si la caja es vertical).

Cuando se divide el rango intercuartil entre 2, se obtiene la **desviación cuartil**, Q , que es una medida de la dispersión estadística:

$$Q = RIC / 2$$

O bien, se expresa como:

$$Q = (Q_3 - Q_1) / 2$$

Ejemplo

Se tiene el siguiente conjunto de datos:

15, 24, 26, 28, 29, 30, 30, 30, 31, 31, 32, 35

Calcule, usando Minitab, el primer cuartil, el tercer cuartil y el rango intercuartil.

Solución

Para realizar este ejercicio en **Minitab** se requiere introducir los datos en una columna de la hoja de trabajo, por ejemplo en la columna C1. Luego se da clic al menú Estadísticas y se elige Mostrar estadísticas descriptivas. En el cuadro de diálogo se selecciona la variable, en este caso en la columna C1, y en el botón estadísticas se marca primer cuartil, tercer cuartil y rango intercuartil, y se da clic en Aceptar.

El resultado se obtiene en la ventana Sesión, e indica que el primer cuartil es 26,5, el tercer cuartil 31, y el rango intercuartil ($RIC = IQR$) es 4,5, que es la diferencia $Q_3 - Q_1 = 31 - 26,5 = 4,5$.

Ejercicio de revisión

Con base en el siguiente conjunto de datos:

40, 50, 65, 85, 75, 90, 60, 60, 60, 75

Calcule el rango intercuartil y la desviación cuartil.

Ver solución de este ejercicio en la página de internet de este texto:
www.auladeeconomia.com/raeep.html

Adicionalmente, a través del análisis de la gráfica de caja puede observar que tan simétrica es la distribución de los datos, pues si se tiene que la mediana está exactamente en la mitad de la caja y que los bigotes izquierdo y derecho (o superior e inferior) son de igual longitud, entonces indicará una distribución simétrica. Conforme el bigote izquierdo sea más largo que el derecho (o el inferior más largo que el superior) y más cercana esté la mediana del tercer cuartil que del primero, entonces la distribución será más asimétrica positiva o sesgada hacia la derecha. Pero, en la medida que el bigote derecho sea más largo que el izquierdo (o el superior más largo que el inferior) y más cercana esté la mediana del primer cuartil que del tercero, entonces la distribución será más asimétrica negativa o sesgada hacia la izquierda.

Ejemplo

Se tiene dos conjuntos de datos, el primero corresponde a la estatura de 8 futbolistas, y el segundo corresponde a la estatura de 6 basquetbolistas:

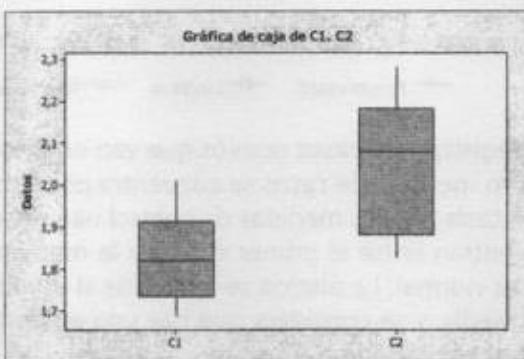
Futbolistas: 1.83, 1.73, 1.75, 1.69, 1.94, 1.83, 1.81, 2.01

Basquetbolistas: 2.01, 2.15, 1.90, 2.28, 1.83, 2.15

Utilice Minitab para elaborar una gráfica de caja para cada conjunto de datos.

Solución

Primero que todo se introducen los datos en la hoja de trabajo, cada conjunto en una columna distinta, en este caso C1 para los futbolistas y C2 para los basquetbolistas. Luego se da clic en el menú Gráfica y se elige Gráfica de caja. En el cuadro de diálogo se selecciona Múltiples Y. Despues se seleccionan las dos variables y se da clic en Aceptar.



Al comparar las dos gráficas, se observa que las estaturas de los futbolistas tienden a ser menores que las de los basquetbolistas, y que el tercer cuartil de los primeros es, apenas, un poco superior que el primer cuartil de los segundos.

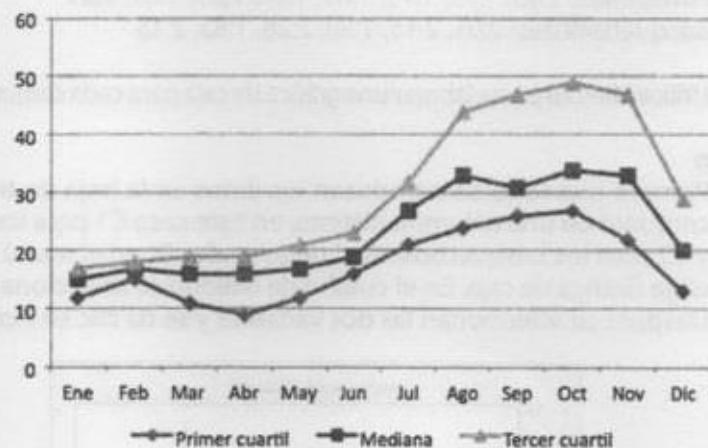
Además, las estaturas de los futbolistas tienden a ser bastante simétricas, tal vez con una ligera asimetría positiva, pues la mediana está apenas un poco abajo de la mitad de la caja y el bigote superior es más largo que el inferior. Al contrario, las estaturas de los basquetbolistas presentan una cierta asimetría negativa, pues la mediana está más arriba de la mitad de la caja, a pesar de que el bigote superior es más largo que el inferior.

Aplicación

Índice endémico

Los cuartiles, la mediana y el rango intercuartil son útiles para determinar el llamado índice endémico y poder construir una gráfica que facilite descubrir en forma oportuna un número inusual de casos de alguna enfermedad.

Para determinar este índice es necesario reunir información de los casos mensuales ocurridos en los últimos 7 años. Luego se toman los datos para cada mes y se calcula el primer cuartil, la mediana y el tercer cuartil.



Luego se registran los casos nuevos que van ocurriendo cada mes. Si el número mensual de casos se encuentra por debajo del primer cuartil, indicaría que las medidas de control han sido exitosas, pero si se encuentran entre el primer cuartil y la mediana, entonces la situación es normal. La alarma se enciende si el número de casos supera la media, y se considera que hay una epidemia si supera al valor del tercer cuartil.

Apoyo audiovisual y uso de la tecnología

En la página de internet www.auladeeconomia.com podrá encontrar una presentación de diapositivas que expone este tema y es una parte importante de este texto. Esta presentación presenta el tema en forma visual, pues emplea fotografías, esquemas u otros recursos visuales, e incluso ejercicios resueltos paso a paso.

Adicionalmente puede encontrar algunos videos explicativos.

Solicite su usuario y contraseña escribiendo al correo electrónico info@auladeeconomia.com. Para ello deberá indicar el número de su factura de compra de texto.

3.4 Medidas de variabilidad

La dispersión o variabilidad en un conjunto de datos se refiere a que los valores son distintos unos de otros. Ahora bien, la idea es establecer algún tipo de medida de esta variabilidad, que permita saber qué tan homogéneo o qué tan heterogéneo es el conjunto de datos. Un modo de definir esa medida es determinar qué tanto se alejan los datos de un valor típico, como el promedio. Así, mientras más heterogénea sea la serie de datos, habrá diferencias mayores de los datos con respecto a su media. Lo contrario sucedería en una serie más homogénea, pues los valores serían más parecidos entre sí, y entonces esas diferencias con respecto a la media serían menores. En términos generales, la variabilidad o dispersión de los datos, es el grado en que los valores de una distribución o serie numérica tienden a acercarse o alejarse alrededor de un promedio.

El problema de la variabilidad

La variabilidad es un aspecto esencial de la estadística. Si no hubiera variabilidad, no existiría el problema del muestreo, por ejemplo. Dado que las variables cambian de una unidad de estudio a otra, entonces es necesario cuantificar ese grado de variabilidad o dispersión, lo cual va a ser clave para muchos análisis estadísticos. Veamos un ejemplo del impacto de la variabilidad y de por qué no basta sólo con las medidas de posición. Dos conjuntos de datos podrían tener la misma media y ser totalmente distintos. Por ejemplo, el conjunto A contiene a los datos 5, 5, 5, 5 y 5, mientras que el conjunto B contiene a los datos 1, 1, 3, 10 y 10. Ambos conjuntos tienen una media de 5, pero en el conjunto A la variabilidad es cero, ya que todos son iguales. Pero en el conjunto B los datos son muy distintos unos de otros, es decir, la variabilidad es elevada.

Las medidas de dispersión pueden ser clasificadas en dos grandes grupos:

- **Medidas de dispersión absolutas.** Las medidas absolutas vienen expresadas en las mismas medidas de la serie de datos. Las más conocidas son: el recorrido, la desviación cuartílica, la desviación semicuartílica, la desviación media, la desviación estándar (también conocida como desviación típica) y la varianza. Aunque estas dos últimas son las más empleadas.
- **Medidas de dispersión relativas.** Las medidas de dispersión relativa son relaciones entre medidas de dispersión absolutas y medidas de tendencia central multiplicadas por 100, por lo tanto vienen expresadas como porcentaje. La medida de dispersión relativa de mayor importancia es el coeficiente de variación.

3.5 Rango o amplitud total

El rango o amplitud total es la medida de dispersión más sencilla. Simplemente es la diferencia entre los valores mayor y menor de un conjunto de datos.

RANGO O AMPLITUD: Diferencia entre el valor máximo y el valor mínimo de un conjunto de datos.

Ejemplo

Suponga que se tienen los dos siguientes conjuntos de cinco datos:

Conjunto A: 1, 2, 3, 7, 10

Conjunto B: 1, 9, 9, 10, 10

Se desea calcular el rango de este conjunto de datos.

Solución

Para el conjunto A el máximo es 10 y el mínimo es 1, por lo que su rango o amplitud será:

$$\text{Rango} = 10 - 1 = 9$$

Para el conjunto B el máximo también es 10 y el mínimo también es 1, por lo que su rango o amplitud será:

$$\text{Rango} = 10 - 1 = 9$$

En este ejemplo se ilustra qué tan limitado es el rango como medida de la variabilidad, pues en el conjunto todos los datos son muy similares entre sí, excepto uno de ellos, sin embargo el rango es igual que el del conjunto A, el cual sí presenta mayor variabilidad.

Ejercicios de revisión

Con base en el siguiente conjunto de datos:

40, 50, 65, 85, 75, 90, 60, 60, 60, 75

Calcule el rango o recorrido.

Ver solución de este ejercicio en la página de internet de este texto:
www.auladeeconomia.com/raeep.html

3.6 La varianza y la desviación estándar

Tanto la varianza, o variancia, como la desviación estándar son las medidas de dispersión más utilizadas en estadística, principalmente por ser muy estables, pues para su cálculo se utilizan todos las diferencias o desviaciones con respecto a la media aritmética de las observaciones.

VARIANZA: Medida de variabilidad que consiste en la media aritmética del cuadrado de las desviaciones de un conjunto de datos con respecto a su media.

La varianza se expresa como s^2 o σ^2 si se calcula para datos de una población o una muestra, respectivamente.

La desviación estándar se puede definir como la raíz cuadrada de la varianza, o sea, la raíz cuadrada del promedio aritmético de los cuadrados de las desviaciones de las observaciones con respecto a su media aritmética.

DESVIACIÓN ESTÁNDAR: Medida de variabilidad que consiste en la raíz cuadrada de la varianza.

La desviación estándar se designa con la letra griega minúscula s (sigma) cuando se trabaja con una población y con la letra s cuando se trabaja con una muestra.

Las fórmulas para calcular la varianza en datos sin agrupar son, para población y muestra respectivamente:

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \mu)^2}{N}} = \sqrt{\sigma^2}$$

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

Es importante destacar que cuando se hace referencia a la población el número de datos se expresa con N y cuando se refiere a la muestra el número de datos se expresa con n .

Las fórmulas para calcular la desviación estándar en datos sin agrupar son, para población y muestra respectivamente:

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \mu)^2}{N}} = \sqrt{\sigma^2}$$

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} = \sqrt{s^2}$$

Otra manera de calcular la desviación estándar es:

$$s = \sqrt{\frac{\sum_{i=1}^n x_i^2 - \left(\frac{\sum_{i=1}^n x_i}{n}\right)^2}{n-1}}$$

Ejemplo

Suponga que se tiene el siguiente conjunto de cinco datos: 1, 1, 3, 10, 10 y se desea calcular la varianza y la desviación estándar de dicha muestra.

Solución

a. Si se emplea la fórmula de la varianza para una muestra, es necesario calcular la media aritmética primero:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{1+1+3+10+10}{5} = \frac{25}{5} = 5$$

b. Luego se calcula la diferencia entre cada dato y la media, resultados que luego serán elevados al cuadrado:

x	$x - \bar{x}$	$(x - \bar{x})^2$
1	$1 - 5 = -4$	$(-4)^2 = 16$
1	$1 - 5 = -4$	$(-4)^2 = 16$
3	$3 - 5 = -2$	$(-2)^2 = 4$
10	$10 - 5 = 5$	$(5)^2 = 25$
10	$10 - 5 = 5$	$(5)^2 = 25$

$$\text{Suma: } \sum (x - \bar{x})^2 = 86$$

c. Finalmente se aplica la fórmula:

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} = \frac{86}{5-1} = 21.5$$

La varianza es 21.5. Si se desea conocer la desviación estándar, entonces lo más práctico es sacar la raíz cuadrada de la varianza:

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} = \sqrt{s^2} = \sqrt{21.5} \approx 4.64$$

La desviación estándar es aproximadamente 4.64. Esta medida mide el grado de dispersión o variabilidad de los datos alrededor de su media. Mientras más grande sea este valor, indica mayor dispersión.

Ejercicio de revisión

Con base en el siguiente conjunto de datos:

40, 50, 65, 85, 75, 90, 60, 60, 60, 75

Calcule la varianza y la desviación estándar.

Ver solución de este ejercicio en la página de internet de este texto:

www.auladeeconomia.com/raeep.html

Uso de Excel y Minitab para calcular la desviación estándar y la varianza

Ejemplo

Utilice Excel y Minitab para resolver el ejercicio: Suponga que se tiene el siguiente conjunto de cinco datos: 1, 1, 3, 10, 10 y se desea calcular la varianza y la desviación estándar de dicha muestra.

Solución

En **Excel**, primero se introducen los datos en una fila o columna, por ejemplo, en la columna A, en el rango de celdas de A1 hasta A5:

	A	B
1	1	
2	1	
3	3	
4	10	
5	10	

Para obtener la varianza, en la celda en la cual se desea el resultado, se introduce la siguiente función:

=VAR(A1:A5)

Y así se obtiene el resultado de 21,5.

Para obtener la desviación estándar, en la celda en la cual se desea el resultado, se introduce la siguiente función:

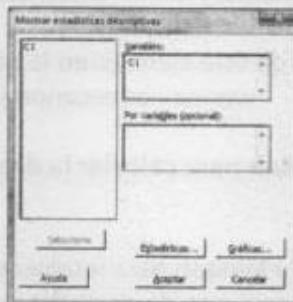
=DESVEST(A1:A5)

Y así se obtiene el resultado de 4,64.

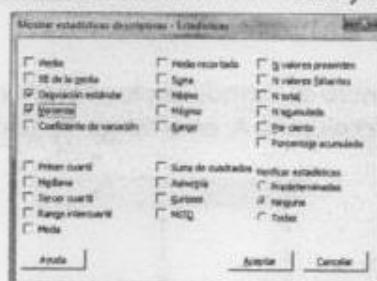
En **Minitab**, primero se introducen los datos en una columna, por ejemplo, en la columna C1:

	C1	C2
1	1	
2	1	
3	3	
4	10	
5	10	

Luego se da clic en el menú Estadísticas, se elige Estadística básica y ahí selecciona Mostrar estadísticas descriptivas. Ahí completa el cuadro de diálogo seleccionando la variable, que en este caso se encuentra en la columna C1:



Luego en el botón Estadísticas selecciona la varianza y la desviación estándar:



Luego de dar clic en Aceptar en cada cuadro, se obtiene el resultado en la ventana Sesión:

Sesión		
Estadísticas descriptivas: C1		
Variable	Desv.Est.	Varianza
C1	4,64	21,50

Material audiovisual

En la página de internet de este texto podrá encontrar videos que exponen el uso de Excel y Minitab para calcular estas medidas descriptivas.

Solicite su usuario y contraseña escribiendo al correo electrónico: info@auladeeconomia.com. Para ello deberá indicar el número de su factura de compra de texto.



Características de la varianza y la desviación estándar

- La varianza y la desviación estándar se calculan con cada todos y cada uno de los valores de una serie de datos y miden la variación alrededor de la media, pues se obtienen a partir de las diferencias de cada dato de la serie con respecto a su media aritmética.
- La desviación estándar es susceptible de operaciones algebraicas.
- La varianza y la desviación estándar siempre son cantidades positivas.

Interpretación y usos de la desviación estándar

La desviación estándar o desviación típica como medida absoluta de dispersión, es la que mejor proporciona la variación de los datos con respecto a la media aritmética, su valor se encuentra en relación directa con la dispersión de los datos, a mayor dispersión de ellos, mayor desviación típica, y a menor dispersión, menor desviación típica. Así, la desviación estándar puede emplearse en distintas áreas, por ejemplo, puede ser una medida de la calidad de un producto, pues en la medida en que haya poca variabilidad en un proceso de producción, entonces el proceso es consistente, lo que quiere decir que el producto podría satisfacer los estándares requeridos. Un proceso con alta variabilidad no puede generar productos de calidad. Por ejemplo, si se llenan botellas con un refresco, y cada botella debe tener 300 ml de líquido, una variabilidad elevada significaría que habría muchas botellas muy vacías y otras muy llenas.

También, la desviación estándar puede verse como una medida del riesgo. Si una inversión es segura, va a dar el mismo rendimiento, o casi el mismo rendimiento, en cualquier caso. Pero una inversión riesgosa podría dar grandes ganancias, pero también grandes pérdidas, por lo que su resultado es altamente variable.

La regla empírica

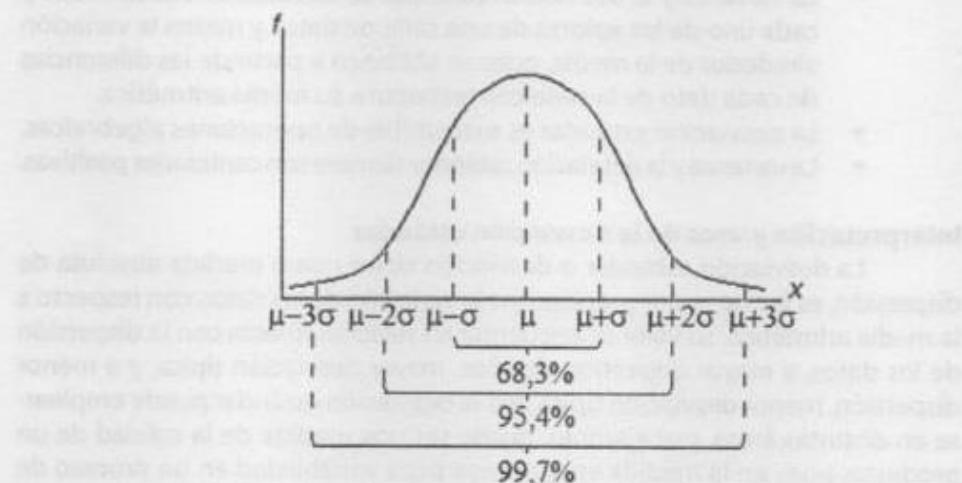
La desviación estándar es de gran utilidad en una distribución normal, ya que en dicha distribución en el intervalo determinado por $\mu \pm s$ se encuentra el 68,3% de los datos de la serie; en el intervalo determinado por $\mu \pm 2s$ se

encuentra el 95,4% de los datos y entre la $\mu \pm 3\sigma$ se encuentra la casi totalidad de los datos, es decir, el 99,7% de los datos; de donde se puede derivar una regla general de gran utilidad para la comprobación de los cálculos. Esta regla se conoce como **Regla empírica**.

Cuando la distribución es **simétrica**, se puede aplicar la **distribución normal**, y las probabilidades que se obtienen son:

- Si el intervalo está definido por la media más/menos una vez la desviación estándar, o sea, $\mu \pm \sigma$, la probabilidad es 68,3%
- Si el intervalo está definido por la media más/menos dos veces la desviación estándar, o sea, $\mu \pm 2\sigma$, la probabilidad es 95,4%
- Si el intervalo está definido por la media más/menos tres veces la desviación estándar, o sea, $\mu \pm 3\sigma$, la probabilidad es 99,7%

Gráficamente se muestra como:



Ejemplo

De acuerdo con datos de un estudio, el gasto destinado a salud en el hogar en el país tiene una media de \$600 anuales y una desviación estándar de \$30. De acuerdo con la regla empírica, ¿por lo menos qué porcentaje de los hogares tendrá un gasto destinado a salud entre \$510 y \$690?

Solución

Se tiene una media de \$600 con una desviación estándar de \$30, y el intervalo dado está entre \$510 y \$690. Para aplicar la regla empírica es necesario saber cuántas veces se ha sumado y restado la desviación estándar al promedio. Esto puede obtenerse fácilmente porque se sabe que cada límite se obtuvo a partir de $\mu \pm k\sigma$, así que, tomando el límite inferior de 510 (y por eso va con signo menor):

$$600 - k * 30 = 510$$

Ahora se despeja esa ecuación:

$$-k * 30 = 510 - 600$$

$$k = -90/-30$$

$$k = 3$$

Si se hubiera tomado el límite superior de 690 se habría obtenido el mismo resultado de $k = 3$.

Sabiendo que $k = 3$, según la regla empírica, el porcentaje de los hogares que tendrá un gasto destinado a salud entre \$510 y \$690 será aproximadamente del 99,7%.

Teorema de Chebyshev

La regla empírica dada anteriormente solo se aplica cuando la distribución es normal, pero ¿es posible generalizar esa regla cuando la distribución sea distinta? Es posible construir intervalos de modo similar, pero las probabilidades van a cambiar. Según el teorema de Chebyshev estas probabilidades van a estar dadas por:

$$1 - \frac{1}{k^2}$$

donde k es el número de veces que se suma y resta la desviación estándar, $k > 1$. Entonces, si el intervalo se construye como $\mu \pm k\sigma$, se tiene que $k = 3$, por tanto, la probabilidad será:

$$1 - \frac{1}{k^2} = 1 - \frac{1}{3^2} = 1 - \frac{1}{9} = \frac{8}{9} = 0,889$$

Es decir, que en el intervalo $\mu \pm k\sigma$, se tiene una probabilidad del 88,9%.

Ejemplo

De acuerdo con datos de un estudio, el gasto destinado a salud en el hogar en el país tiene una media de \$600 anuales y una desviación estándar de \$30. a.

De acuerdo con el teorema de Chebyshev, ¿por lo menos qué porcentaje de los hogares tendrá un gasto destinado a salud entre \$525 y \$675?

Solución

Se tiene una media de \$600 con una desviación estándar de \$30, y el intervalo dado está entre \$525 y \$675. Para aplicar el teorema de Chebyshev es necesario saber cuántas veces se ha sumado y restado

la desviación estándar al promedio. Esto puede obtenerse fácilmente porque se sabe que cada límite se obtuvo a partir de $\mu \pm k\sigma$, así que, tomando el límite inferior de 525 (y por eso va con signo menos):

$$600 - k * 30 = 525$$

Ahora se despeja esa ecuación:

$$\begin{aligned} -k * 30 &= 525 - 600 \\ k &= -75 / -30 \\ k &= 2,5 \end{aligned}$$

Si se hubiera tomado el límite superior de 675 se habría obtenido el mismo resultado de $k = 2,5$.

Sabiendo que $k = 3$, según el teorema de Chebyshev, se aplica la fórmula sustituyendo $k = 2,5$:

$$1 - \frac{1}{k^2} = 1 - \frac{1}{2,5^2} = 1 - \frac{1}{6,25} = 0,84$$

Así, el porcentaje de los hogares que tendrá un gasto destinado a salud entre \$525 y \$675 será al menos 84%.

Ejercicio de revisión

Las botellas de agua envasadas en un proceso de llenado tienen una media de 501 ml con una desviación estándar de 2 ml.

- Si no se conoce si la distribución es simétrica o asimétrica, ¿qué porcentaje de las botellas podrían tener entre 497 ml y 505 ml de agua?
- ¿Cómo cambia su respuesta anterior si se sabe que la distribución del contenido de agua en las botellas se distribuye normalmente?

Ver solución de este ejercicio en la página de internet de este texto:
www.auladeeconomia.com/raeep.html

Puntajes estandarizados

Considere la siguiente situación: una persona ha realizado una prueba de aptitud académica, que mide la capacidad del razonamiento verbal, cuantitativo y analítico, entre otras destrezas. Esta persona obtuvo una puntuación de 40 en el área de aptitud verbal y 42 puntos en la de matemáticas y razonamiento analítico. ¿Es cierto que tiene mejores destrezas para el razonamiento cuantitativo que para el verbal? Para poder responder se requiere un poco más de información. Según la institución que desarrolla la prueba, la sección de aptitud verbal tiene una media de 44 puntos con una desviación estándar de

11 puntos, mientras que la sección de razonamiento matemático tiene una media de 50 puntos con una desviación estándar de 9 puntos. Entonces, se puede emplear la desviación estándar como una especie de unidad de medida para realizar comparaciones entre datos con diferente media, e incluso con distintas unidades de medición. En una situación como esta se pueden calcular los puntajes estandarizados, que consisten en tomar un valor de una variable y restarle su media, para luego dividirlo entre su desviación estándar. Este resultado va a indicar "cuántas desviaciones estándar se aleja un dato de su media". La fórmula para calcular los puntajes estandarizados sería:

$$z = \frac{x - \mu}{\sigma}$$

En el caso de la prueba de aptitud verbal se tiene que la media (m) es 44 puntos, la desviación estándar (s) es 11 y la puntuación de la persona (x) es 40, por lo que el puntaje estandarizado será:

$$z = \frac{x - \mu}{\sigma} = \frac{40 - 44}{11} = -0,36$$

El resultado es negativo porque su puntuación está por debajo de la media e indica que su resultado está 0,36 veces la desviación estándar por debajo del promedio.

En el caso de la prueba de razonamiento matemático se tiene que la media (m) es 50 puntos, la desviación estándar (s) es 9 y la puntuación de la persona (x) es 42, por lo que el puntaje estandarizado será:

$$z = \frac{x - \mu}{\sigma} = \frac{42 - 50}{9} = -0,89$$

Puede observarse que en términos absolutos obtuvo más puntos en la sección de razonamiento matemático, pero en términos relativos, o sea, en términos de los puntajes estandarizados, su puntuación en aptitud verbal es mayor, pues se aleja menos veces la desviación estándar que en el caso de análisis cuantitativo.

Ejemplo

Se sabe que los recién nacidos varones de una ciudad tienen un peso medio de 3.450 gramos, con una desviación estándar de 75 gramos, mientras que los recién nacidos varones de una zona rural tienen un peso medio de 3.350 gramos con una desviación estándar de 100 gramos. Una madre residente de esa ciudad acaba de tener un niño con un peso de 3.475 gramos y otra madre residente de la zona rural dada acaba de tener un niño con un peso de 3.450 gramos, ¿cuál de los dos niños tiene, en términos relativos, un peso mayor?

Solución

En el caso de la ciudad se tiene que la media (m) es 3.450 gramos, la desviación estándar (s) es 75 gramos y el peso del recién nacido (x) es 3.475 gramos, por lo que el puntaje estandarizado será:

$$z = \frac{x - \mu}{\sigma} = \frac{3475 - 3450}{75} = 0,33$$

En el caso de la zona rural se tiene que la media (m) es 3.350 gramos, la desviación estándar (s) es 100 gramos y el peso del recién nacido (x) es 3.450 gramos, por lo que el puntaje estandarizado será:

$$z = \frac{x - \mu}{\sigma} = \frac{3450 - 3350}{100} = 1$$

El puntaje estandarizado para el niño de zona rural es mayor que para el niño de la ciudad, por lo que, en términos relativos, tiene un peso mayor.

Ejercicio de revisión

Suponga que el gasto promedio anual en salud de cada habitante de Argentina es de \$742 con una desviación estándar de \$250, mientras que en Chile se destinan, en promedio, \$947 en salud al año, con una desviación estándar de \$358.

Si una persona en Argentina gastó este año \$850 en salud, mientras que otra persona en Chile gastó \$1050 en salud, ¿cuál de los dos gastó más en términos relativos?

Ver solución de este ejercicio en la página de internet de este texto:
www.auladeeconomia.com/raeep.html

3.7 Medidas de variabilidad relativa: El coeficiente de variación

Las medidas de variabilidad descritas hasta ahora miden la variabilidad absoluta de un conjunto de datos. Esto es muy útil cuando se trata de analizar una sola población o una sola muestra, pero no es tan útil cuando hay que establecer comparaciones entre distintas poblaciones o muestras, especialmente si las magnitudes que se comparan son muy diferentes o están expresadas en unidades de medida distintas. Es por esto que es necesario calcular medidas de variabilidad en términos relativos, o sea, como proporciones o porcentajes.

El **coeficiente de variación** se designa con las letras CV y consiste en el cociente de la desviación estándar entre la media.

COEFICIENTE DE VARIACIÓN Es el cociente de la desviación estándar entre la media de un conjunto de datos y muestra la variabilidad relativa de un conjunto de datos.

El coeficiente de variación es la medida de variabilidad relativa más usada. Es un índice de variabilidad sin dimensiones, lo que facilita la comparación entre diferentes conjuntos de datos, aun expresados en diferentes unidades de medida.

La fórmula matemática es, para población y muestra, respectivamente:

$$CV = \frac{\sigma}{\mu} \times 100 \quad CV = \frac{s}{\bar{x}} \times 100$$

Ejemplo

En el caso del conjunto de datos anterior, se calculó una media de 5 y una desviación estándar de 4.64, calcule el coeficiente de variación.

Solución

Dado que se calculó una media de 5 y una desviación estándar de 4.64, entonces el coeficiente de variación es:

$$CV = \frac{s}{\bar{x}} \times 100 = \frac{4.64}{5} \times 100 = 92.74\%$$

Ejercicio de revisión

Con base en el siguiente conjunto de datos:

40, 50, 65, 85, 75, 90, 60, 60, 60, 75

Calcule el coeficiente de variación.

Ver solución de este ejercicio en la página de internet de este texto:
www.auladeeconomia.com/raeep.html

Uso de Excel y Minitab para calcular el coeficiente de variación

Ejemplo

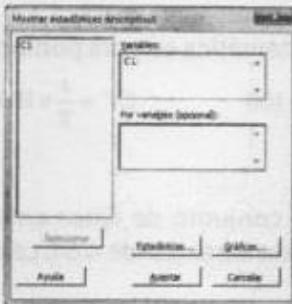
Utilice Minitab para resolver el ejercicio: Suponga que se tiene el siguiente conjunto de cinco datos: 1, 1, 3, 10, 10 y se desea calcular el coeficiente de variación de dicha muestra.

Solución

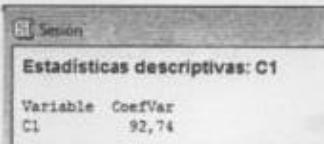
En Minitab, primero se introducen los datos en una columna, por ejemplo, en la columna C1:

	C1	C2
1	1	
2	1	
3	3	
4	10	
5	10	

Luego se da clic en el menú Estadísticas, se elige Estadística básica y ahí selecciona Mostrar estadísticas descriptivas. Ahí completa el cuadro de diálogo seleccionando la variable, que en este caso se encuentra en la columna C1:



Luego en el botón Estadísticas selecciona el coeficiente de variación. Después de dar clic en Aceptar en cada cuadro, se obtiene el resultado en la ventana Sesión:



Uso de Excel y Minitab para calcular medidas descriptivas

Ejemplo

Se tiene el siguiente conjunto de datos:

24, 24, 25, 25, 25, 25, 25, 25, 26, 26, 26, 26, 27, 27, 27, 28, 29, 30

Utilice la herramienta de análisis de datos de Excel para obtener las principales estadísticas descriptivas de este conjunto de datos.

Solución

Lo primero es introducir los datos en la hoja de Excel. Lo más conveniente es agregarlos todos en una misma columna, que en este caso va de la celda A1 hasta la celda A15.

Luego se da clic a la pestaña Datos y en la sección Análisis se elige el botón Análisis de datos. Ahora elige Estadística descriptiva. Ahora hay que completar el cuadro de diálogo.



En rango de entrada se indica el rango de datos, por lo que se seleccionan las celdas de la A1 hasta la A15. Después marca la opción Resumen de estadísticas y da clic en Aceptar.

Excel genera una serie de medidas estadísticas de uso común, como se muestra a continuación.

	A Columna 1
3	
4	Media 26,1333333
5	Error típico 0,45634852
6	Mediana 26
7	Moda 25
8	Desviación estándar 1,7674302
9	Varianza de la muestra 3,12380952
10	Curtosis 0,21741866
11	Coeficiente de asimetría 0,93233034
12	Rango 6
13	Mínimo 24
14	Maximo 30
15	Varianza 392
	Cuenta 15

Ejemplo

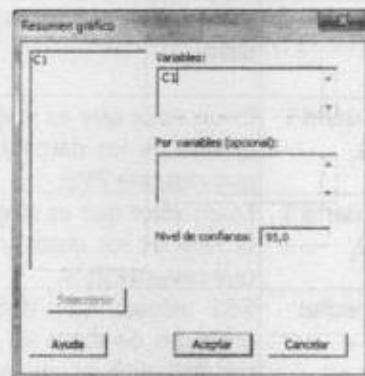
Se tiene el siguiente conjunto de datos:

24, 24, 25, 25, 25, 25, 25, 26, 26, 26, 26, 27, 27, 28, 29, 30

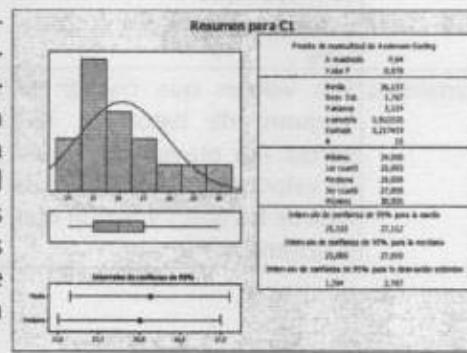
Utilice el resumen gráfico de Minitab para obtener las principales estadísticas descriptivas de este conjunto de datos.

Solución

El primer paso es digitar estos datos en una columna de la hoja de trabajo de Minitab, por ejemplo, en la columna C1. Luego se da clic en el menú Estadísticas, se selecciona Estadística básica y ahí se elige Resumen gráfico.



En el cuadro de diálogo se selecciona la variable en la columna C1 y se da clic en Aceptar. Minitab despliega una ventana con un histograma con ajuste a la curva normal y una gráfica de caja. Además un cuadro con varias medidas descriptivas y otros datos que se estudiarán más adelante en este texto.



Apoyo audiovisual y uso de la tecnología

En la página de internet www.auladeeconomia.com/raeep.html podrá encontrar una presentación de diapositivas que expone este tema y es una parte importante de este texto. Esta presentación presenta el tema en forma visual, pues emplea fotografías, esquemas u otros recursos visuales, e incluso recursos resueltos paso a paso.

Adicionalmente puede encontrar algunos videos explicativos.

Solicite su usuario y contraseña escribiendo al correo electrónico info@auladeeconomia.com. Para ello deberá indicar el número de su factura de compra de texto.

Cuadro de resumen sobre las medidas de posición y variabilidad:

Medidas de posición		
Medida	Significado	Uso
Moda	El valor que más se repite.	Se usa sobre todo en variables cualitativas para indicar lo más frecuente, aunque también se puede usar en variables cuantitativas.
Mediana	Es un valor mayor que el 50% de los datos y menor que el restante 50%.	Se usan para describir el conjunto y es muy útil cuando la distribución de los datos no es simétrica y hay valores extremos.
Media	Es el valor promedio de los datos.	Es la medida más usada para describir un conjunto de datos. Cuando hay valores extremos, se ve distorsionada.
Cuartil 1 Q_1	Es un valor que es mayor que el 25% de los datos y menor que restante 75%.	Se usa para indicar la posición relativa de un valor con respecto a los demás. Equivale al percentil 25.
Cuartil 3 Q_3	Es un valor que es mayor que el 75% de los datos y menor que restante 25%.	Se usa para indicar la posición relativa de un valor con respecto a los demás. Equivale al percentil 75.
Deciles	Son valores que dividen al conjunto de datos en 10 partes. Así, el D_6 es un valor que es mayor que un 60% de los datos y menor que el restante 40%.	Se usa para indicar la posición relativa de un valor con respecto a los demás. Cada decil equivale a un percentil. Por ejemplo $D_6 = P_{60}$.
Percentiles	Son valores que dividen al conjunto de datos en 100 partes. Por ejemplo, el P_{38} es un valor que es mayor que un 38% de los datos y menor que el restante 62%.	Se usa para indicar la posición relativa de un valor con respecto a los demás.

Medidas de variabilidad		
Medida	Significado	Uso
Rango o recorrido	Es la diferencia entre el valor máximo y el mínimo de un conjunto de datos.	Se usa para medir la variabilidad de los datos, pero es una medida deficiente.
Varianza	Mide la media de los cuadrados de las desviaciones con respecto al promedio.	Es una de las medidas más empleadas para medir la variabilidad. Mientras más alta sea, indica más variabilidad.
Desviación estándar	Mide la media de las desviaciones con respecto al promedio	Es la medida más empleada para medir la variabilidad absoluta. Mientras más alta sea, indica más variabilidad.
Coeficiente de variación	Es la desviación estándar entre el promedio.	Es la medida más empleada para medir la variabilidad relativa. Mientras más alta sea, indica más variabilidad.

3.8 Ejercicios

Ejercicios de desarrollo

Resuelva los ejercicios que a continuación se le presentan (las respuestas a los ejercicios se encuentran en la página de internet del texto: www.auladeeconomia.com/raeep.html).

1. Los siguientes son los ingresos de siete personas (en \$/mes):

300, 450, 500, 250, 600, 550, 450, 3000

- a. Calcule e interprete la media, moda y mediana.
 - b. ¿Cuál de las tres medidas describe mejor al conjunto de datos?
2. La presencia de manganeso en la gasolina puede significar un problema, pues genera ciertos daños a los vehículos, a la salud de las personas y al ambiente. Ante la duda de altas concentraciones de esta sustancia en la gasolina vendida en el país se toman muestras en los distintos planteles de la compañía refinadora de petróleo para que vaya a ser estudiadas en un laboratorio. Los siguientes son datos recopilados para un plantel de la refinadora durante el mes de octubre anterior (concentración de manganeso en mg/L de gasolina):

0,4	0,5	0,0	0,1	0,0	0,2	0,6	0,2	0,1	0,0
0,2	0,3	0,6	0,2	0,3	0,4	0,3	0,2	0,1	0,0

Con base en los datos anteriores:

- a. Calcule e interprete la media, la mediana y la moda.
- b. Indique la forma de la distribución de las concentraciones de manganeso (simétrica o asimétrica positiva o negativa).

3. Los siguientes son los pesos, en gramos, de una muestra de frascos de mermelada que se envasan en una empresa:

252, 260, 266, 248, 240, 246, 255, 260, 270, 258, 259, 260, 264, 254, 256, 262

- Calcule e interprete la media, moda y mediana.
- ¿Qué se puede concluir sobre la simetría o asimetría de los pesos?
- Si cada frasco debe contener 250 gramos, ¿qué indican los datos calculados y el tipo de simetría o asimetría que se presenta?

4. Los siguientes datos corresponden a las tasas de pobreza de un país en el transcurso de los últimos 12 años:

18,7	18,8	18,7	17,8	17,3	17,8
16,6	16,5	16,5	16,5	17,3	17,6

Con base en los datos anteriores, conteste las siguientes preguntas (incluya todos los cálculos y procedimientos necesarios):

- Calcule la tasa media de pobreza en el periodo citado e interprete el resultado.
 - Calcule la tasa mediana de pobreza en el periodo citado e interprete el resultado.
 - Calcule la tasa de pobreza más frecuente en el periodo citado e interprete el resultado.
 - ¿Qué puede decirse sobre la simetría de las tasas de pobreza de este país en el periodo dado? Explique la forma de la distribución de las tasas de pobreza basándose en los indicadores adecuados.
5. Un contador está realizando una auditoría de los cheques extendidos por una compañía. En una muestra de 15 cheques se obtuvieron los siguientes montos, en miles dólares:

8, 4, 2, 60, 6, 8, 14, 25, 20, 5, 6, 8, 2, 19, 11

- Calcula la media, mediana y moda de los montos de los cheques.
- ¿Cuál de las tres medidas calculadas en el inciso anterior describe mejor al conjunto de datos?
- ¿Qué podría afirmarse sobre la forma de la distribución de los montos de los cheques?

6. En una muestra de 12 personas con miopía elevada superior a 12,00 D se midieron algunos parámetros biométricos de sus ojos, como la longitud axial. Los resultados fueron los siguientes:

Paciente	Longitud axial (mm)
1	25
2	28
3	26
4	30
5	28
6	27
7	25
8	29
9	31
10	30
11	31
12	33
13	29
14	31
15	29
16	32
17	29

Con base en estos datos calcule las medidas de tendencia central e indique si la distribución de esta variable es simétrica o asimétrica.

7. Un instituto de idiomas ha contratado un estudio de salarios de profesores (salarios mensuales). En dicho estudio se obtuvieron los siguientes resultados:

Percentil 12 = \$600

Cuartil 1 = \$650

Mediana = \$725

Cuartil 3 = \$800

Quintil 4 = \$1000

Decil 9 = \$1200

Interprete cada uno de estos resultados.

8. Sean los siguientes datos:

45, 56, 72, 81, 43, 62, 55, 90, 49, 52, 60, 75, 66, 72, 58, 50, 88

Calcule:

- a. Cuartil 1
- b. Cuartil 3
- c. Decil 7
- d. Percentil 42
- e. Percentil 93
- f. Quintil 3

9. Sea la siguiente muestra de datos:
- 45, 56, 72, 81, 43, 62, 55, 90, 49, 52, 60, 75, 66, 72, 58, 50, 88

Calcule la varianza y la desviación estándar.

10. Un médico ha reunido los siguientes datos sobre la estancia hospitalaria en horas de una muestra de 10 neonatos en la unidad de cuidados intensivos de un hospital:

150, 70, 95, 120, 120, 72, 240, 84, 600, 48

Con base en los datos anteriores, calcule las siguientes medidas descriptivas (incluya todos los cálculos y procedimientos necesarios):

- a. Mediana
- b. Media aritmética
- c. Percentil 95
- d. Desviación estándar
- e. Interprete cada uno de los resultados anteriores.

11. Sean las siguientes muestras de datos:

Muestra 1: 45, 56, 72, 81, 43, 62, 55, 90, 49

Muestra 2: 67, 52, 60, 70, 66, 72, 58, 50, 58

¿Cuál posee mayor variabilidad relativa?

12. Sea la siguiente muestra de datos:

45, 56, 72, 81, 43, 62, 55, 90, 49, 52, 60, 75, 66, 72, 58, 50, 88

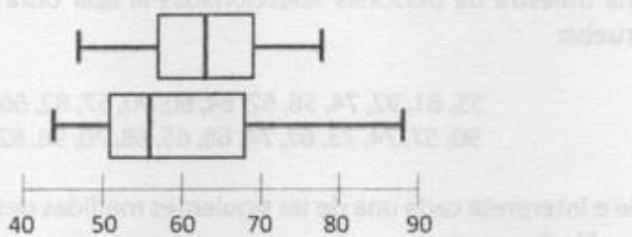
Calcule el rango intercuartil y la desviación cuartil.

13. Sea la siguiente muestra de datos:

45, 56, 60, 60, 43, 62, 55, 69, 49, 52, 60, 75, 66, 72, 58, 50, 88

Construya el diagrama de caja.

14. Analice y compare las siguientes dos diagramas de dispersión.



15. Los siguientes son los tiempos de funcionamiento, en horas, de un tipo de componente electrónico: 100, 110, 80, 95, 97, 82, 113, 128, 116, 98, 99, 105, 187, 114, 121, 104, 118, 109. Construya la gráfica de caja e interprete la. ¿Puede identificarse valores extremos?
16. Un ingeniero ha recopilado datos sobre la vida útil de 20 filtros de gasolina del mismo tipo y marca. Los datos son los siguientes (en miles de kilómetros): 12, 14, 16, 15, 10, 20, 13, 15, 16, 14, 12, 13, 11, 13, 15, 16, 13, 14, 14, 12.
- Calcule la media, mediana y moda e interprete los resultados.
 - Construya e interprete la gráfica de caja.
 - ¿Qué rango se encontraría al menos el 75% de las duraciones de los filtros de gasolina, suponiendo que no se conoce la forma de la distribución de esta variable?
17. En un país el 25% de los hogares con menos recursos reciben un ingreso mensual promedio, por persona, de \$125. En el otro extremo, el 25% de los hogares con mayores ingresos, reciben, en promedio, por persona, \$1970 por mes.
- Calcule e interprete el rango intercuartílico.
 - Calcule e interprete la desviación cuartil.
18. Sea la siguiente muestra de datos:
- 45, 56, 60, 60, 43, 62, 55, 69, 49, 52, 60, 75, 66, 72, 58, 50, 88
- Calcule el rango de variación del 68,3% de estos datos.
 - Calcule el rango en el que se encuentra el 95,4% de los datos.
 - Calcule el rango en el que se encuentra el 99,7% de los datos.

19. Dado el siguiente conjunto de datos sobre el tiempo (en minutos) que una muestra de personas seleccionadas al azar dura en completar una prueba:

55, 81, 92, 74, 58, 62, 84, 60, 90, 57, 82, 66,
90, 57, 74, 73, 67, 74, 88, 65, 68, 70, 98, 82

Calcule e interprete cada una de las siguientes medidas descriptivas:

- a. Moda
- b. Mediana
- c. Media aritmética
- d. Primer cuartil
- e. Tercer cuartil
- f. Percentil 45
- g. Decil 3
- h. Quintil 3
- i. Desviación estándar
- j. Varianza
- k. Coeficiente de variación
- l. Rango intercuartil
- m. Desviación cuartil
- n. ¿Qué se puede concluir sobre la simetría o asimetría de los datos?

20. Una enfermera ha reunido los siguientes datos sobre la cantidad de horas extra semanales laboradas por 12 enfermeras de la unidad de cuidados intensivos de un hospital:

12, 14, 18, 6, 22, 19, 20, 14, 14, 17, 0, 6

Con base en los datos anteriores, calcule las siguientes medidas descriptivas (incluya todos los cálculos y procedimientos necesarios):

- a. Mediana
- b. Media aritmética
- c. Percentil 65
- d. Desviación estándar
- e. Interprete cada uno de los resultados anteriores:

21. Dado el siguiente conjunto de datos sobre el número de turistas que se hospedan en un hotel por semana en una muestra de 22 semanas seleccionadas al azar:

25	81	92	44	58	62	34	40	90	57	42
40	57	57	57	57	62	40	65	58	50	18

Calcule e interprete cada una de las siguientes medidas descriptivas:

- a. Moda
 - b. Mediana
 - c. Media aritmética
 - d. Primer cuartil
 - e. Tercer cuartil
 - f. Percentil 45
 - g. Decil 3
 - h. Quintil 3
 - i. Desviación estándar
 - j. Varianza
 - k. Coeficiente de variación
 - l. Rango intercuartil
 - m. Desviación cuartil
 - n. ¿Qué se puede concluir sobre la simetría o asimetría de los datos?
 - o. ¿Qué % de los datos se encuentra en un rango de ± 2 veces la desviación estándar?
 - p. Construya el diagrama de caja e interprétele.
22. Un laboratorio recibe un lote de comprimidos de un fármaco cuyo principio activo es omeprazol. La especificación indica que cada uno debe contener 20 mg de dicho principio activo. En una muestra de 20 comprimidos se encontraron las siguientes cantidades de omeprazol:

19,3	18,1	17,2	16,5	21,6	19,7	16,6	14,2	19,4	19,3
18,7	17,1	11,6	16,4	15,8	16,8	17,9	16,4	19,3	12,9

Con base en los datos anteriores, calcule e interprete:

- a. Media aritmética
- b. Mediana
- c. Moda
- d. ¿Qué puede concluirse con respecto a la simetría o asimetría de los contenidos de principio activo de estos comprimidos?
- e. Construya la gráfica de caja para los datos suministrados e interprete el gráfico.
- f. Calcule la varianza. ¿Qué se concluye si se ha especificado que la variabilidad relativa de esta variable no debe exceder el 30%?

23. Una maestra ha reunido los siguientes datos sobre la talla en centímetros de niñas de una institución preescolar. Los datos corresponden a cinco grupos de la institución. Los datos se presentan a continuación:

Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5
108	99	116	115	117
119	96	117	115	115
119	99	119	114	101
101	92	112	112	112
115	99	128	115	125
117	92	126	115	119
115	94	133	110	112
119	113	125	110	117
112	128	126	119	115
110	101	124	115	119
112	96	121	115	121
103	99	131	125	108
99	126	126	117	110
119	114	119	119	105
117	108	108	108	-
115	112	121	-	-
-	125	115	-	-
-	101	-	-	-

Con base en los datos anteriores relativos a la zona indicada, calcule las siguientes medidas descriptivas (incluya todos los cálculos y procedimientos necesarios):

- Moda del grupo 1
- Mediana del grupo 2
- Media aritmética del grupo 3
- Primer cuartil del grupo 4
- Percentil 85 del grupo 5
- Desviación estándar del grupo 5
- Interprete cada uno de los resultados anteriores:
- ¿Qué puede decirse de la simetría o asimetría de la distribución de los datos en el grupo número 2?
- Con base en los resultados calculados, ¿es cierto que un niño del grupo 1 con una talla de 102 centímetros es relativamente pequeño? Explique basándose en los datos.
- ¿Cuál grupo tiene mayor variabilidad relativa: el grupo 2 o el grupo 3?
- Construya e interprete la gráfica de caja para los datos del grupo 4.

24. Suponga que se realizó un estudio en el que se comparó la efectividad de dos métodos de enseñanza A y B en una escuela rural. Luego de aplicar exámenes a los estudiantes se obtuvieron los siguientes resultados:

Método de enseñanza	A	B
Media aritmética	74,5	74,9
Desviación estándar	11,3	24,6
Mediana	72	68
Moda	71	65

Determine la validez de las siguientes afirmaciones (justifique su respuesta):

- a. Los resultados de ambos métodos son prácticamente iguales.
 - b. Con el método B casi todos los estudiantes obtuvieron la misma nota.
 - c. La distribución de las notas del método B presenta una asimetría hacia la izquierda.
 - d. El método A presenta una dispersión de las notas mayor que el método B.
25. Una investigadora ha aplicado un instrumento estandarizado para la medición de los aprendizajes en preescolares de 5 años. La prueba fue desarrollada por la investigadora y utiliza una selección de aprendizajes esperados y cubre distintos ámbitos, como la formación social, la comunicación, la relación con el medio natural, entre otros. Al final se obtiene un puntaje total luego de la aplicación de todos los ítems de la prueba. El instrumento se aplicó a una muestra de niños de 5 años y se obtuvieron los siguientes resultados:

35	69	50	41	51	49	31	27	66	78	56	46
45	47	49	33	24	55	49	51	46	47	46	47

Con base en los datos anteriores, calcule e interprete las siguientes medidas descriptivas siguientes (incluya todos los cálculos y procedimientos necesarios):

- a. Mediana
- b. Moda
- c. Media aritmética
- d. Percentil 85
- e. Cuartil 1
- f. Cuartil 3
- g. Desviación estándar
- h. Con base en los datos suministrados, utilice las medidas de posición central para determinar si la distribución de los datos es simétrica o asimétrica. Explique.
- i. Dada la respuesta anterior, ¿cuál medida preferiría en este caso para describir el comportamiento de los datos? Explique.
- j. Con base en los datos suministrados, construya el diagrama de caja para la variable resultado de la aplicación del instrumento estandarizado para la medición de los aprendizajes en preescolares de 5 años.
- k. Al analizar la gráfica de caja, ¿qué se puede afirmar sobre la dispersión de los datos? Explique. Al analizar la gráfica de caja, ¿qué se puede concluir sobre la forma de la distribución de los datos (simetría o asimetría)? Explique.

26. Una trabajadora social está preocupada por el consumo de drogas ilegales entre los estudiantes de secundaria de una comunidad marginal del país llamada ciudadela Los Robles. Una muestra de 18 estudiantes varones del liceo local reportó las siguientes cifras como la edad en que se inició el consumo de drogas ilegales:

14	17	16	13	15	16	13	14	15	12	16	18	15	14	11	13	15	8
----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	---

Con base en los datos anteriores, calcule e interprete cada una de las siguientes medidas descriptivas:

- Moda
- Mediana
- Media aritmética
- Rango intercuartil
- Percentil 85
- Desviación estándar
- ¿Qué puede decirse de la simetría o asimetría de la distribución de los datos anteriores sobre la variable edad de inicio en el consumo de drogas?
- Establezca un intervalo que incluya aproximadamente al 75% de las observaciones, tomando como base los datos de la pregunta uno sobre las edades de inicio en el consumo de drogas.
- Un estudio similar en otra comunidad, llamada San Juan, determinó una edad promedio de inicio en el consumo de drogas entre varones estudiantes de secundaria de 15,85 años con una desviación estándar de 3,1 años.
 - ¿Cuál de las dos comunidades presenta mayor homogeneidad en cuanto a las edades en el inicio del consumo de drogas de los varones estudiantes de secundaria? Explique basándose en los cálculos adecuados.
 - Suponga que Luis vive en Los Robles y empezó a consumir drogas a los 16 años. Roberto vive en San Juan y empezó a consumir drogas a los 18 años. En términos de relativos, ¿cuál de los dos empezó a consumir drogas a menor edad?
- Con respecto a la variable edad de inicio del consumo de drogas en la comunidad de Los Robles:
 - Construya el diagrama de caja. Incluya todos los cálculos y la gráfica con la escala adecuada.
 - ¿Qué puede concluirse sobre la dispersión de los datos a partir de la gráfica anterior? Básese en las medidas de variabilidad asociadas con la mediana.
 - ¿Qué puede concluirse sobre la simetría o asimetría de los datos? Utilice la gráfica para responder.

27. Según un estudio de una agencia de publicidad sobre el número de horas de televisión por semana que ven los jóvenes entre 12 y 18 de una zona del país poseen las siguientes características (todas en horas por semana):

Moda: 13	Mediana: 14,34	Media: 16,67
Cuartil 1: 11,04	Cuartil 3: 18,92	Decil 2: 10,87
Decil 8: 19,56	Decil 9: 22,35	Percentil 95: 25,88
Desviación estándar: 4,5	Mínimo: 6	Máximo: 28

Con base en los datos anteriores, determine lo siguiente:

- a. Interprete cada una de las medidas anteriores.
 - b. Calcule e interprete el coeficiente de variación.
 - c. ¿Cuáles de las siguientes afirmaciones son falsas y cuáles verdaderas, basándose en los datos?
 - i. "Yo veo muy poca TV, pues por semana estoy frente al televisor solo 20 horas"
 - ii. "Yo creo que la mayoría de los muchachos ven menos TV que mi hijo. Es increíble que gaste 11 horas semanales frente al televisor"
 - iii. "El 95% de los jóvenes miran entre 7,67 y 23,67 horas de TV por mes"
 - iv. "La distribución del número de horas de televisión por semana que ven los jóvenes entre 12 y 18 de una zona del país es asimétrica negativa"

Examen del capítulo

EJERCICIOS DE REPASO
En cada caso seleccione la opción que mejor contesta cada pregunta. (las respuestas a los ejercicios se encuentran en la página de internet del texto: www.auladeeconomia.com/raeep.html).

6. La moda del siguiente conjunto de datos 7, 8, 20, 20, 20, 13, 14, 12, 6, 9, 20, 6 es:
(a) 20 (b) 10
(c) 13 (d) 11,36

7. El primer cuartil del siguiente conjunto de datos 7, 8, 20, 20, 13, 14, 12, 6, 9, 10, 6 es:
(a) 7 (b) 20
(c) 13 (d) 6

8. El tercer cuartil del siguiente conjunto de datos 7, 8, 20, 20, 13, 14, 12, 6, 9, 10, 6 es:
(a) 14 (b) 20
(c) 13 (d) 17

9. El percentil 30 del siguiente conjunto de datos 7, 8, 20, 20, 13, 14, 12, 6, 9, 10, 6 es:
(a) 7,6 (b) 10
(c) 7 (d) 6

10. El percentil 70 del siguiente conjunto de datos 7, 8, 20, 20, 13, 14, 12, 6, 9, 10, 6 es:
(a) 7 (b) 20
(c) 13,4 (d) 6

11. La desviación estándar del siguiente conjunto de datos 7, 8, 20, 20, 13, 14, 12, 6, 9, 10, 6 es:
(a) 25,45 (b) 5,05
(c) 1,52 (d) 44,4

12. La varianza del siguiente conjunto de datos 7, 8, 20, 20, 13, 14, 12, 6, 9, 10, 6 es:
(a) 25,45 (b) 5,05
(c) 1,52 (d) 44,4

13. Si en una muestra, la media es igual a la moda y a la mediana, entonces se concluye que:
A. Los datos son iguales
B. La desviación estándar es cero

Con respecto a las dos afirmaciones anteriores, es correcto que con toda certeza:

- (a) Ambas son verdaderas
 - (b) Solo A es verdadera
 - (c) Solo B es verdadera
 - (d) Ambas son falsas

14. Si dos valores en un grupo de datos ocurren más a menudo que otros cualesquiera, la distribución de los datos será _____. La opción que mejor completa la frase anterior es:

 - (a) Simétrica
 - (b) Bimodal
 - (c) Asimétrica positiva
 - (d) Asimétrica negativa

15. Considere las dos siguientes afirmaciones:

- A. Los valores extremos en un conjunto de datos influyen profundamente en la mediana.
- B. Para un arreglo de datos con 50 observaciones, la mediana será el valor de la vigésima quinta observación en el arreglo.

Con respecto a las dos afirmaciones anteriores, es correcto que:

- (a) Ambas son verdaderas
- (b) Solo A es verdadera
- (c) Solo B es verdadera
- (d) Ambas son falsas

16. Considere las dos siguientes afirmaciones:

- A. Cuando la población tiene sesgo negativo o positivo, a menudo es preferible utilizar la mediana como la mejor medida de localización, pues siempre se encuentra entre la media y la moda.
- B. Cuando una distribución es simétrica y tiene una moda, el punto más alto en la curva es la mediana y la media.

Con respecto a las dos afirmaciones anteriores, es correcto que:

- (a) Ambas son verdaderas
- (b) Solo A es verdadera
- (c) Solo B es verdadera
- (d) Ambas son falsas

17. ¿Cuál de los siguientes enunciados no es una ventaja del uso de la mediana?

- (a) Los valores extremos afectan a la mediana menos intensamente que a la media
- (b) La mediana es fácil de entender
- (c) Una mediana puede calcularse para descripciones cualitativas
- (d) Ninguna de las anteriores

18. Considere las dos siguientes afirmaciones:

- A. Las medidas de tendencia central en un conjunto de datos se refieren al grado de dispersión de las observaciones.
- B. La diferencia entre las observaciones más grandes y las más pequeñas en un conjunto de datos se llama media geométrica.

Con respecto a las dos afirmaciones anteriores, es correcto que:

- (a) Ambas son verdaderas
- (b) Solo A es verdadera
- (c) Solo B es verdadera
- (d) Ambas son falsas

19. Considere las dos siguientes afirmaciones:

- A. La desviación estándar se mide en las mismas unidades que las observaciones en el conjunto de datos.
- B. Una desventaja del uso del rango para medir la dispersión consiste en que ignora la naturaleza de las variaciones entre la mayor parte de las observaciones.

Con respecto a las dos afirmaciones anteriores, es correcto que:

- (a) Ambas son verdaderas
- (b) Solo A es verdadera
- (c) Solo B es verdadera
- (d) Ambas son falsas

20. Si un grupo de datos tiene tan sólo una moda y el valor de la moda es menor que el de la media, podremos llegar a la conclusión de que la gráfica de la distribución es:
- Simétrica
 - Sesgada a la izquierda
 - Sesgada a la derecha
 - Platicúrtica
21. ¿Cuál de los siguientes enunciados NO es correcto?
- Algunos conjuntos de datos no tienen media.
 - En los cálculos de la media influyen los valores extremos de datos.
 - Una media ponderada ha de emplearse cuando es necesario tener en cuenta la importancia de cada valor.
 - Todos estos enunciados son correctos.
22. ¿Cuál de los siguientes enunciados es el primer paso en el cálculo de la mediana de un conjunto de datos?
- Obtener el promedio de los dos valores de la mitad en un conjunto de datos.
 - Ordenar los datos en un arreglo.
 - Determinar los pesos relativos de los valores de los datos por orden de importancia.
 - Ninguno de los anteriores.
23. ¿Cuál de los siguientes casos es un ejemplo de una medida relativa de dispersión?
- Desviación estándar
 - Varianza
 - Coeficiente de variación
 - Las opciones a y b pero no c
24. Si p es el mayor de tres enteros consecutivos, entonces el promedio de los tres números es:
- p
 - $p - 1$
 - $p - 3$
 - $3p - 1$
25. La edad promedio de un grupo de 5 amigos es 17,4 años. Si se incorpora al grupo un amigo de 18 años, la edad promedio de nuevo grupo es:
- 17,5 años
 - 17,7 años
 - 21 años
 - 20,4 años
26. El gerente de un centro de llamadas desea evaluar el desempeño de los agentes de servicio y para ello decide basarse en los tiempos de espera de los clientes para ser atendidos (medido en segundos) y el grado de satisfacción que los clientes manifiesten al recibir el servicio (valorado como bueno, regular o malo). La semana anterior tomó una muestra de 12 llamadas por día de lunes a miércoles y 20 llamadas el jueves y el viernes. Los siguientes son los tiempos de las muestras tomadas de lunes a miércoles. El lunes y el martes se tomaron tiempos de llamadas atendidas y el miércoles solo de llamadas no atendidas:

Día	Número de muestras (tiempo en segundos)											
	1	2	3	4	5	6	7	8	9	10	11	12
Lunes	13	15	15	16	16	16	16	17	17	19	11	21
Martes	12	15	15	16	16	15	16	18	15	20	30	40
Miércoles	30	40	45	45	50	50	45	35	60	50	80	100

Las preguntas de la 26 a la 43 se basan en la información anterior.

La distribución de los tiempos de las llamadas del día lunes es:

27. La distribución de los tiempos de las llamadas del día martes es:

28. El tiempo medio de espera de los clientes de la muestra del día martes es, en segundos:

29. La mediana del tiempo de espera de los clientes de la muestra del día martes es, en segundos:

30. La moda del tiempo de espera de los clientes de la muestra del día lunes es, en segundos:

- (a) 15
(b) 15,5
(c) 16
(d) Ninguna de las anteriores

31. Con respecto a los datos del día miércoles es verdadero que:

- (a) La distribución es asimétrica negativa
(c) La distribución es bimodal
(b) No se presentan valores extremos
(d) Todas las anteriores son verdaderas

32. El cuartil 1 de los tiempos de espera del día lunes es, en segundos:

33. El cuartil 3 de los tiempos de espera del día martes es, en segundos:

34. El percentil 80 de los tiempos de espera del día miércoles es, en segundos:

35. La varianza de los tiempos de espera del día miércoles es, aproximadamente, en segundos²:

36. La desviación estándar de los tiempos de espera del día martes es, en segundos:
- (a) 5,5
 - (b) 8
 - (c) 64
 - (d) Ninguna de las anteriores
37. El coeficiente de variación de los tiempos de espera del lunes es:
- (a) 15,99%
 - (b) 2,55
 - (c) 6,25%
 - (d) Ninguna de las anteriores
38. Con relación a la variabilidad relativa de los tiempos de espera es verdadero que el día cuyos tiempos tienen una dispersión relativa más baja es:
- (a) Lunes
 - (b) Martes
 - (c) Miércoles
 - (d) Falta información para determinarlo
39. El decil 4 de los tiempos de espera del día lunes es, en segundos:
- (a) 5,2
 - (b) 15
 - (c) 16
 - (d) Ninguna de las anteriores
40. Si el primer cuartil para los tiempos de espera del día viernes es de 14,6 segundos, entonces es falso que:
- (a) Un 25% de los clientes de ese día esperaron 14,6 segundos o menos
 - (b) Un 75% de los clientes de ese día esperaron 14,6 segundos o más
 - (c) Un cliente que esperó 12 segundos esperó poco con respecto a los demás
 - (d) Ninguna de las anteriores
41. Si la mediana para los tiempos de espera del día viernes es de 19,8 segundos, entonces es falso, con toda certeza, que:
- (a) Un 50% de los clientes de ese día esperaron 19,8 segundos o menos
 - (b) El tiempo de espera promedio de ese día fue de 19,8 segundos
 - (c) La mayoría de los clientes esperaron más de 19,8 segundos
 - (d) Ninguna de las anteriores
42. Si la media para los tiempos de espera del día viernes es de 21,3 segundos, entonces es verdadero que:
- (a) La mitad de los clientes esperaron 21,3 segundos o menos
 - (b) Un 50% de los clientes de ese día esperaron 21,3 segundos o más
 - (c) El tiempo más frecuente fue 21,3 segundos
 - (d) Todas las anteriores son falsas
43. Si la media para los tiempos de espera es de 21,3 segundos y la desviación estándar para los tiempos de espera del día viernes es de 7,6 segundos, entonces es verdadero que:
- (a) Los tiempos de espera tuvieron una variabilidad de 7,6 segundos con relación a su media
 - (b) Aproximadamente un 68,3% de los clientes esperaron entre 13,7 y 28,9 segundos
 - (c) Aproximadamente un 95,4% de los clientes esperaron entre 6,1 y 36,5 segundos
 - (d) Todas las anteriores son verdaderas

CAPÍTULO 4

DISTRIBUCIÓN DE FRECUENCIAS

OBJETIVOS

Al concluir el capítulo, será capaz de:

- Construir la tabla de una distribución de frecuencias.
 - Representar gráficamente los datos provenientes de una distribución de frecuencias.
 - Calcular e interpretar las principales medidas de posición en datos agrupados.
 - Calcular e interpretar las principales medidas de variabilidad en datos agrupados.

4.1 Necesidad de resumir la información

Cuando se tiene un conjunto grande de datos, por ejemplo de 2000 datos, no es fácil poder determinar por la simple observación de todos esos datos características de la forma en que se distribuyen ni conocer medidas que describen al conjunto, como la media, la mediana, la desviación estándar, etc. Entonces es necesario contar con alguna herramienta que permita resumir la información. Un modo de hacer esto es agrupando las observaciones en grupos o clases, o sea, en una distribución de frecuencias.

En una distribución de frecuencias se crean clases y a cada clase se le asocia con una frecuencia. Es decir, se determina cuántas veces aparecen datos en cada una de las clases. La **distribución de frecuencias** es una tabla de datos estadísticos agrupados en clases que muestra la frecuencia asociada con cada clase.

DISTRIBUCIÓN DE FRECUENCIAS: Es una tabla que agrupa los datos en clases o categorías mutuamente excluyentes y en la que se indica la cantidad de observaciones correspondientes a cada categoría.

Las distribuciones de frecuencias pueden ser útiles para clasificar tanto variables cualitativas como cuantitativas, sean éstas discretas o continuas. A continuación se dan algunos ejemplos:

– **Distribución para un atributo** (variable cualitativa):

Ventas por producto
para el mes de julio 2004

Producto	Ventas (millones€)
A	25,5
B	84,2
C	12,8
Total	122,6

– **Distribución para una variable discreta:**

Número de reclamos de clientes
durante Agosto de 2004

Número de reclamos	Número de clientes
1	30
2	26
3	18
4	9
5	4
6 o más	7
Total	94

– Distribución para una variable continua:

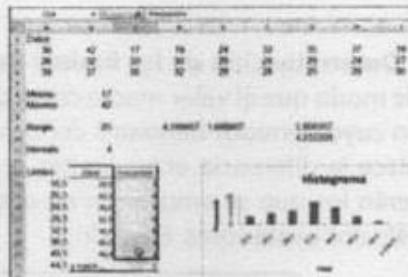
Nivel de ingreso familiar mensual
San José, 2004

Ingresos familiar mensual	Número de familias
Menos de ₡150.000	20
De ₡150.000 a ₡300.000	100
De ₡300.000 a ₡450.000	85
Más de ₡450.000	80
Total	285

Material audiovisual

En la página de internet de este texto podrá encontrar una presentación y un video que exponen este tema.

Solicite su usuario y contraseña escribiendo al correo electrónico info@auladeeconomia.com. Para ello deberá indicar el número de su factura de compra de texto.



Construcción de distribuciones de frecuencias

Para ilustrar el proceso de la construcción de una distribución de frecuencias se empleará el siguiente ejemplo: Se tienen los siguientes datos que corresponden a la edad de 30 personas:

19	25	32	40	21	28	56	27	31	29
41	36	32	18	50	48	25	33	35	26
28	24	22	27	35	26	43	34	43	39

– Determinación del rango o amplitud total: Esto consiste en encontrar la diferencia entre el dato más alto y el más bajo. En este caso:

Dato mayor: 56

Dato menor: 18

$$\text{Rango} = \text{dato mayor} - \text{dato menor} = 56 - 18 = 38$$

– Selección del intervalo de clase (c): No es práctico que la distribución de frecuencias tenga muchas clases, pero tampoco demasiado pocas. Generalmente se recomienda entre un mínimo de 5 clases y un máximo de 15. Entonces se divide el rango entre 6 y entre 15:

$$38 \div 6 = 7,6$$

$$38 \div 15 = 2,53$$

Esto quiere decir que cualquier número entero entre 2,53 y 7,6 puede escogerse como intervalo de clase. Generalmente se prefieren los números impares, por ejemplo $c = 5$.

También, para determinar cuántas clases crear, se puede utilizar la siguiente fórmula (fórmula de Sturges):

$$\text{Número de clases} = 1 + 3,322 \log n$$

donde n es el número total de datos. Si al aplicar la fórmula se obtiene un número decimal, se aproxima al siguiente entero.

– **Determinación de los límites de clase:** Los límites de clase deben elegirse de modo que el valor medio coincida, hasta donde sea posible, con los valores en cuyo derredor tiendan a concentrarse los datos. Cabe señalar que se establece la diferencia entre límites reales y límites indicados. Los límites reales serán los que se emplearán en el cálculo de los puntos medios y los demás cálculos posteriores. Ejemplo:

Límites reales	Límites indicados
17,5 - 22,5	18 - 22
22,5 - 27,5	23 - 27
27,5 - 32,5	28 - 32
32,5 - 37,5	33 - 37
37,5 - 42,5	38 - 42
42,5 - 47,5	43 - 47
47,5 - 52,5	48 - 52
52,5 - 57,5	53 - 57

Ejercicio de revisión

Suponga que se tienen los siguientes datos relacionados con el tiempo de espera de los usuarios de un servicio (en minutos):

12, 16, 8, 22, 14, 12, 13, 19, 17, 10,
21, 25, 23, 18, 14, 9, 14, 16, 10, 12,
15, 16, 16, 17, 12, 11, 11, 19, 20, 15

Determine cuáles serían los límites reales si se desea construir la tabla de la distribución de frecuencias empleando 6 clases.

Ver solución de este ejercicio en la página de internet de este texto:

www.auladeeconomia.com/raeep.html

– **Tabulación de los datos:** Esto consiste en distribuir los datos entre las clases seleccionadas. Generalmente se redondea a la unidad inmediata superior si el primer decimal del dato es 5 o más.

Límites reales	Recuento	Frecuencia
17,5 - 22,5		4
22,5 - 27,5		5
27,5 - 32,5		8
32,5 - 37,5		5
37,5 - 42,5		3
42,5 - 47,5		2
47,5 - 52,5		2
52,5 - 57,5		1
Total		30

Ejercicio de revisión

Suponga que se tienen los siguientes datos relacionados con el tiempo de espera de los usuarios de un servicio (en minutos):

12, 16, 8, 22, 14, 12, 13, 19, 17, 10,
21, 25, 23, 18, 14, 9, 14, 16, 10, 12,
15, 16, 16, 17, 12, 11, 11, 19, 20, 15

Si los siguientes son los límites reales, determine las frecuencias absolutas de cada clase:

Límites reales	Frecuencia absoluta
7,5 - 10,5	
10,5 - 13,5	
13,5 - 16,5	
16,5 - 19,5	
19,5 - 22,5	
22,5 - 25,5	
Total	30

Ver solución de este ejercicio en la página de internet de este texto:
www.auladeeconomia.com/raeep.html

– **Cálculo de los puntos medios:** El punto medio es el valor central de la clase. Se obtiene como el promedio entre el límite superior y el límite inferior de cada clase. Se denotarán como χ_i .

Límites indicados	Límites reales	Puntos medio x_i	Frecuencia f_i
18 - 22	17,5 - 22,5	20	4
23 - 27	22,5 - 27,5	25	5
28 - 32	27,5 - 32,5	30	8
33 - 37	32,5 - 37,5	35	5
38 - 42	37,5 - 42,5	40	3
43 - 47	42,5 - 47,5	45	2
48 - 52	47,5 - 52,5	50	2
53 - 57	52,5 - 57,5	55	1
Total:			30

Ejercicio de revisión

Suponga que se tienen los siguientes datos relacionados con el tiempo de espera de los usuarios de un servicio (en minutos), determine los puntos medios de cada clase:

Límites reales	Puntos medios
7,5 - 10,5	
10,5 - 13,5	
13,5 - 16,5	
16,5 - 19,5	
19,5 - 22,5	
22,5 - 25,5	

Ver solución de este ejercicio en la página de internet de este texto:
www.auladeeconomia.com/raeep.html

– **Cálculo de frecuencias relativas simples:** Determinan la proporción o porcentaje del total de observaciones que se clasifican en cada clase. Se calculan dividiendo la frecuencia absoluta de cada clase entre el total de observaciones. El total de las frecuencias relativas siempre debe ser uno.

Límites reales	Puntos medios x_i	Frecuencia absoluta f_i	Frecuencia relativa $f_r = f_i / n$
17,5 - 22,5	20	4	0,1333
22,5 - 27,5	25	5	0,1667
27,5 - 32,5	30	8	0,2667
32,5 - 37,5	35	5	0,1667
37,5 - 42,5	40	3	0,1000
42,5 - 47,5	45	2	0,0667
47,5 - 52,5	50	2	0,0667
52,5 - 57,5	55	1	0,0333
30			1

Ejercicio de revisión

Suponga que se tienen los siguientes datos relacionados con el tiempo de espera de los usuarios de un servicio (en minutos), determine las frecuencias relativas de cada clase:

Límites reales	Frecuencia absoluta	Frecuencia relativa
7,5 - 10,5	4	
10,5 - 13,5	7	
13,5 - 16,5	9	
16,5 - 19,5	5	
19,5 - 22,5	3	
22,5 - 25,5	2	
Total	30	

Ver solución de este ejercicio en la página de internet de este texto:
www.auladeeconomia.com/raeep.html

– **Frecuencias absolutas acumuladas:** Se puede dar la acumulación en dos sentidos: 1. "Menos de" que da el número de observaciones menores o iguales al límite real superior de la clase correspondiente, y 2. "Más de" que da el número de observaciones mayores o iguales al límite inferior de la clase correspondiente.

Límites reales	Puntos medios	Frecuencia absoluta	Frecuencia relativa	Frecuencia absoluta acumulada	Frecuencia absoluta acumulada
	x_i	f_i	$f_r = f_i / n$	"menos de"	"más de"
17,5 - 22,5	20	4	0,1333	4	30
22,5 - 27,5	25	5	0,1667	9	26
27,5 - 32,5	30	8	0,2667	17	21
32,5 - 37,5	35	5	0,1667	22	13
37,5 - 42,5	40	3	0,1000	25	8
42,5 - 47,5	45	2	0,0667	27	5
47,5 - 52,5	50	2	0,0667	29	3
52,5 - 57,5	55	1	0,0333	30	1
		30		1	

Ejercicio de revisión

Suponga que se tienen los siguientes datos relacionados con el tiempo de espera de los usuarios de un servicio (en minutos), determine las frecuencias absolutas acumuladas a menos de y a más de cada clase:

Límites reales	Frecuencia absoluta	Frecuencia acumulada	
		a menos de	a más de
7,5 - 10,5	4		
10,5 - 13,5	7		
13,5 - 16,5	9		
16,5 - 19,5	5		
19,5 - 22,5	3		
22,5 - 25,5	2		
Total	30		

Ver solución de este ejercicio en la página de internet de este texto:
www.auladeeconomia.com/raeep.html

– **Frecuencias relativas acumuladas:** Se puede dar la acumulación en dos sentidos: 1. "Menos de" que da la proporción de observaciones menores o iguales al límite real superior de la clase correspondiente, y 2. "Más de" que da la proporción de observaciones mayores o iguales al límite inferior de la clase correspondiente.

Límites reales	Puntos medios	Frecuencia absoluta	Frecuencia relativa	Frecuencia absoluta acumulada	"menos de"	Frecuencia absoluta acumulada	Frecuencia relativa acumulada	Frecuencia relativa acumulada
17,5 - 22,5	20	4	0,1333	4	30	0,1333	1,0000	
22,5 - 27,5	25	5	0,1667	9	26	0,3000	0,8667	
27,5 - 32,5	30	8	0,2667	17	21	0,5667	0,7000	
32,5 - 37,5	35	5	0,1667	22	13	0,7333	0,4333	
37,5 - 42,5	40	3	0,1000	25	8	0,8333	0,2667	
42,5 - 47,5	45	2	0,0667	27	5	0,9000	0,1667	
47,5 - 52,5	50	2	0,0667	29	3	0,9667	0,1000	
52,5 - 57,5	55	1	0,0333	30	1	1,0000	0,0333	
		30	1					

Ejercicio de revisión

Suponga que se tienen los siguientes datos relacionados con el tiempo de espera de los usuarios de un servicio (en minutos), determine las frecuencias relativa acumuladas a menos de y a más de de cada clase:

Límites reales	Frecuencia relativa	Frecuencia acumulada	
		a menos de	a más de
7,5 - 10,5	4		
10,5 - 13,5	7		
13,5 - 16,5	9		
16,5 - 19,5	5		
19,5 - 22,5	3		
22,5 - 25,5	2		
Total	30		

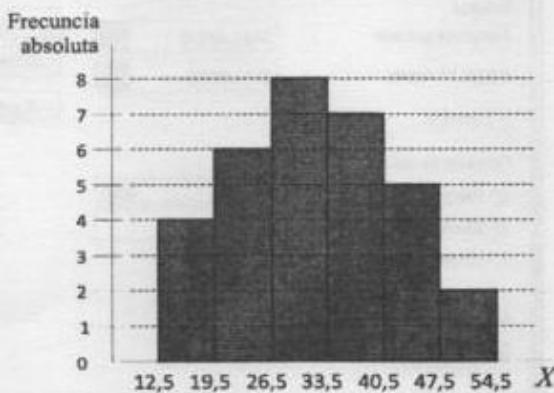
Ver solución de este ejercicio en la página de internet de este texto:
www.auladeeconomia.com/raeep.html

4.2 Gráficos de las distribuciones de frecuencia

Para mostrar gráficamente una distribución de frecuencias se emplean diversos tipos de gráficos, tales como los histogramas, polígonos de frecuencias y ojivas.

Un **histograma** es un gráfico que se forma construyendo rectángulos sobre cada uno de los límites reales de cada intervalo, cuya altura equivale a la frecuencia absoluta o relativa de cada clase. En el histograma el valor de la variable aparece en el eje horizontal y la frecuencia absoluta o relativa en el eje vertical. El histograma se emplea para presentar los datos cuando estos corresponden a los valores de una variable cuantitativa continua, no una variable discreta ni cualitativa. Es por eso que, para señalar la continuidad de la variable, no se dejan espacios entre las barras.

El siguiente es un ejemplo de un histograma:



Uso de Excel y Minitab para construir histogramas

Ejemplo

Se tiene la edad de 30 personas en la tabla siguiente:

19	25	32	40	21	28	56	27	31	29
41	36	32	18	50	48	25	33	35	26
28	24	22	27	35	26	43	34	43	39

Utilice Excel y Minitab para construir un histograma que represente dichos datos.

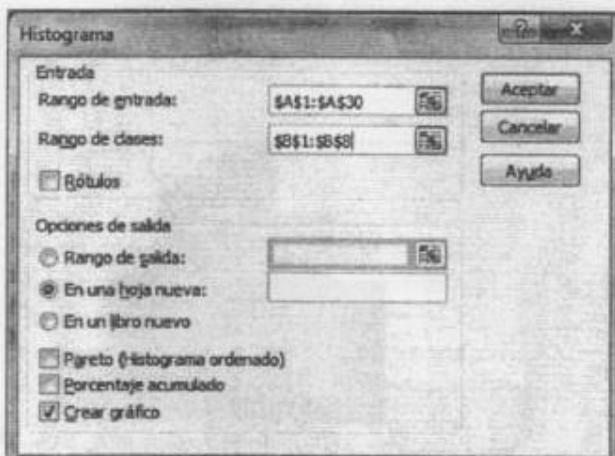
Solución

En Excel, primero se introducen los datos en una columna (o una fila). En este caso se introducen los datos en el rango de celdas A1 hasta A30.

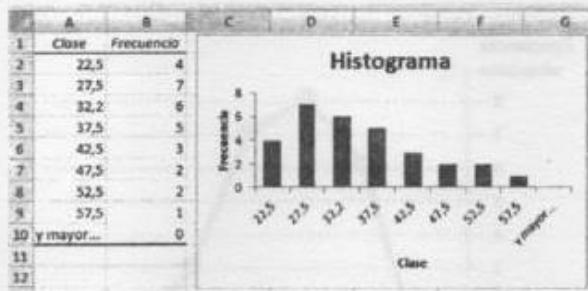
En otro rango de celdas se introducen los límites de las clases. Solo es necesario indicar los límites superiores de las clases, que en este caso serían 22.5, 27.5, 32.5, 37.5, 42.5, 47.5, 52.5 y 57.5. Estos límites se introducirán en este caso en las celdas de B1 hasta B8.

Luego se da clic en la pestaña Datos, y se selecciona Análisis de datos. Si no aparece el botón de Análisis de datos, se puede instalar dando clic al botón de Office (en la esquina superior izquierda del programa), y en el menú se da clic en Opciones de Excel. Ahí se elige en el menú de la izquierda se da clic en Complementos, y en los complementos de aplicaciones inactivas se elige Herramientas para análisis. Después se da clic en el botón Ir que se haya en la parte inferior del cuadro de diálogo, y en la lista de complementos disponibles se marca Herramientas para análisis y después se presiona Aceptar).

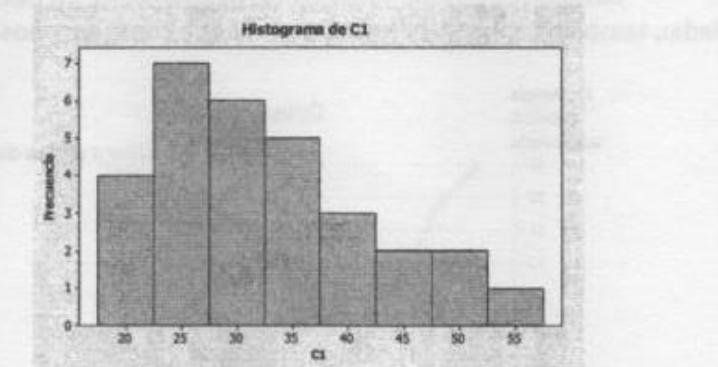
Ahora al dar clic en el botón de Análisis de datos, en la lista se elige Histograma y se completa el cuadro de diálogo siguiente:



En rango de entrada se indican los datos de la serie a graficar, que en este caso están en las celdas de A1 hasta A30. En rango de clases se indican los límites, los cuales están en las celdas de B1 hasta B8. Luego hay que marcar la opción Crear gráfico, y se da clic en Aceptar. Excel genera una tabla y un gráfico como el siguiente:



En Minitab, primero se introducen los datos en la hoja de trabajo. Luego se da clic en el menú Gráfica y se elige Histograma. En el cuadro de diálogo se selecciona la opción Simple. Después, en el cuadro se selecciona como variables de gráficas la columna C1 y se da clic en Aceptar:



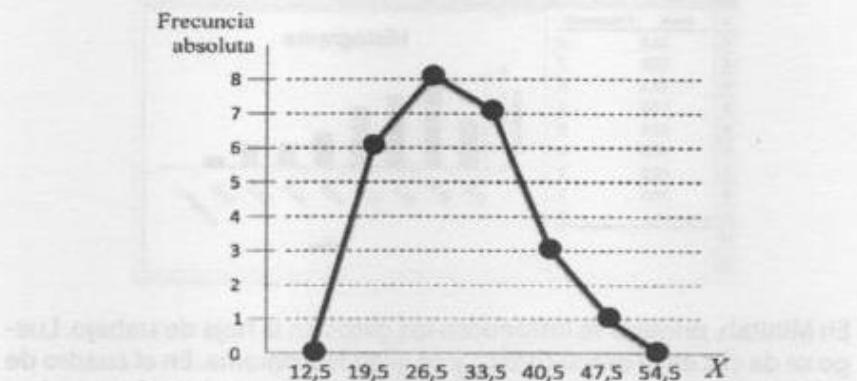
Ejercicio de revisión

Suponga que se tienen los siguientes datos relacionados con el tiempo de espera de los usuarios de un servicio (en minutos), construya un histograma para esta variable:

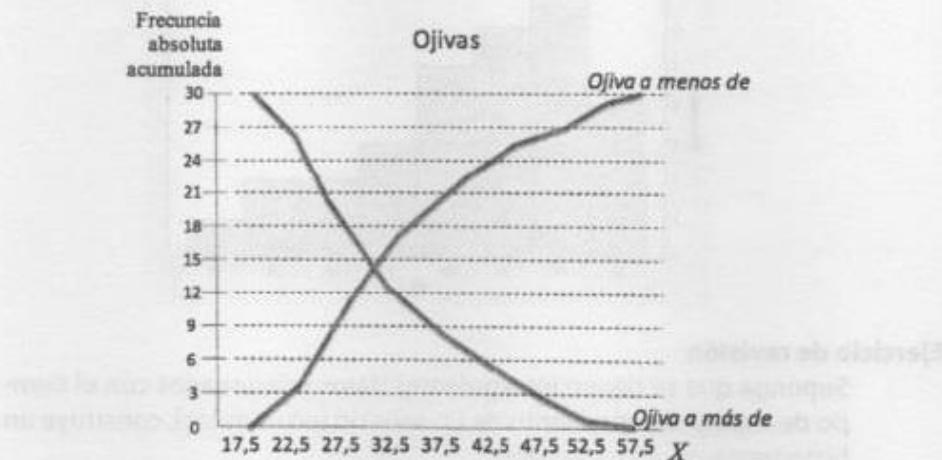
Límites reales	Frecuencia absoluta
7,5 - 10,5	4
10,5 - 13,5	7
13,5 - 16,5	9
16,5 - 19,5	5
19,5 - 22,5	3
22,5 - 25,5	2
Total	30

Ver solución de este ejercicio en la página de internet de este texto:
www.auladeeconomia.com/raeep.html

Un **polígono de frecuencias** es una línea que conecta los puntos medios de todas y cada una de las barras de un histograma. En el polígono de frecuencias el valor de la variable aparece en el eje horizontal y la frecuencia absoluta o relativa en el eje vertical. El siguiente es un ejemplo de un polígono de frecuencias:



También es posible construir los diagramas de las frecuencias acumuladas, también llamadas ojivas, tanto a más de como a menos de:



4.3 Cálculo de medidas de posición y variabilidad en distribuciones de frecuencias

La moda

En una distribución de frecuencias con datos agrupados no es posible conocer con exactitud el valor de la moda, pero si es posible identificar la clase modal, y a partir de ella obtener un valor aproximado de la moda. Esto quiere decir que no es posible saber con toda precisión el valor de la moda.

Para hallar la moda se emplea la siguiente fórmula, la cual se aplica a la clase modal, es decir, la clase con la frecuencia absoluta más alta:

$$M_o = L_i + c \frac{d_1}{d_1 + d_2}$$

donde:

M_o : valor de la moda

L_i : límite inferior real de la clase modal

c : intervalo de la clase modal

d_1 : diferencia entre la frecuencia de la clase modal y la pre modal

d_2 : diferencia entre la frecuencia de la clase modal y la pos modal

Ejemplo

Considere la siguiente distribución de frecuencias:

Límites reales	Puntos medios x_i	Frecuencia absoluta f_i	Frecuencia relativa $f_r = f_i / n$	Frecuencia relativa acumulada "menos de"
17,5 - 22,5	20	4	13,33%	13,33%
22,5 - 27,5	25	5	16,67%	30,33%
27,5 - 32,5	30	8	26,67%	56,67%
32,5 - 37,5	35	5	16,67%	73,33%
37,5 - 42,5	40	3	10,00%	83,33%
42,5 - 47,5	45	2	6,67%	90,00%
47,5 - 52,5	50	2	6,67%	96,67%
52,5 - 57,5	55	1	3,33%	100,00%
Total		30	100,00%	

Calcule la moda.

Solución

La clase modal es la tercera clase, cuyos límites reales son 27,5 – 32,5, su límite inferior real es 27,5, su intervalo de clase es 5 (límite superior menos límite inferior = 32,5 – 27,5 = 5) y su frecuencia absoluta es 8. La clase pre modal (22,5 – 27,5) tiene frecuencia 5 (por tanto $d_1 = 8 - 5 = 3$) y la pos modal (32,5 – 37,5) tiene frecuencia también de 5 (por tanto $d_2 = 8 - 5 = 3$). El cálculo es:

$$M_o = L_i + c \frac{d_1}{d_1 + d_2} = 27,5 + 5 \frac{(8-5)}{(8-5)+(8-5)} = 30$$

La moda es 30.

Ejercicio de revisión

Suponga que se tienen los siguientes datos relacionados con el tiempo de espera de los usuarios de un servicio (en minutos), calcule la moda:

Límites reales	Frecuencia absoluta
7,5 - 10,5	4
10,5 - 13,5	7
13,5 - 16,5	9
16,5 - 19,5	5
19,5 - 22,5	3
22,5 - 25,5	2
Total	30

Ver solución de este ejercicio en la página de internet de este texto:
www.auladeeconomia.com/raeep.html

Material audiovisual

En la página de internet de este texto podrá encontrar una presentación y un video que exponen el cálculo de las .

Solicite su usuario y contraseña escribiendo al correo electrónico info@auladeeconomia.com. Para ello deberá indicar el número de su factura de compra de texto.

Ejercicio:			
Los siguientes datos corresponden al consumo mensual (en kg.) para 30 familias. Completa el cuadro.			
Consumo mensual	Frecuencia absoluta	Frecuencia relativa	Freq. relativa acum. a metros de
1,25 - 1,75	4		
1,75 - 2,25			10
2,25 - 2,75		20	
2,75 - 3,25			
3,25 - 3,75	5		70
3,75 - 4,25			
Total	30		

La mediana

Tampoco es posible conocer el valor de la mediana con toda precisión en el caso de distribuciones de frecuencias con datos agrupados, pero si es posible identificar la clase mediana (clase donde está la mediana), como aquella cuya frecuencia acumulada "menos de" es apenas mayor a $n/2$. El cálculo de la mediana implica el uso de la siguiente fórmula:

$$Med = L_i + c \left[\frac{n/2 - F_{i-1} \downarrow}{f_i} \right]$$

donde:

Med: valor de la mediana

L_i: límite inferior real de la clase mediana

c: intervalo de la clase mediana

n: número total de datos en la distribución de frecuencias

F_{i-1}↓: frecuencia acumulada "menos de" de la clase pre mediana

f_i: frecuencia absoluta de la clase mediana

Ejemplo

Considere la siguiente distribución de frecuencias:

Límites reales	Puntos medios	Frecuencia absoluta	Frecuencia relativa	Frecuencia relativa acumulada
	x_i	f_i	$f_e = f_i / n$	"menos de"
17,5 - 22,5	20	4	13,33%	13,33%
22,5 - 27,5	25	5	16,67%	30,33%
27,5 - 32,5	30	8	26,67%	56,67%
32,5 - 37,5	35	5	16,67%	73,33%
37,5 - 42,5	40	3	10,00%	83,33%
42,5 - 47,5	45	2	6,67%	90,00%
47,5 - 52,5	50	2	6,67%	96,67%
52,5 - 57,5	55	1	3,33%	100,00%
Total		30	100,00%	

Calcule la mediana.

Solución

Se tiene que $n = 30$, por tanto $n/2 = 30/2 = 15$, lo que quiere decir que la clase mediana será la tercera clase, ya que su frecuencia acumulada menos de es 17, que es la que apenas supera a 15. El límite inferior de la clase es 27,5, el intervalo de la clase es 5, la frecuencia acumulada de la clase pre mediana es 9 y la frecuencia de la clase mediana es 8. Aplicando la fórmula:

$$Med = L_i + c \left[\frac{n/2 - F_{i-1} \downarrow}{f_i} \right] = 27,5 + 5 \left[\frac{30/2 - 9}{8} \right] = 27,5 + 5 \cdot \frac{6}{8} = 31,25$$

La mediana es 31,25.

Ejercicio de revisión

Suponga que se tienen los siguientes datos relacionados con el tiempo de espera de los usuarios de un servicio (en minutos), calcule la mediana

Límites reales	Frecuencia absoluta
7,5 - 10,5	4
10,5 - 13,5	7
13,5 - 16,5	9
16,5 - 19,5	5
19,5 - 22,5	3
22,5 - 25,5	2
Total	30

Ver solución de este ejercicio en la página de internet de este texto:
www.auladeeconomia.com/raeep.html

La media aritmética

Para calcular la media aritmética se emplean los puntos medios de cada clase y sus frecuencias absolutas, como sigue:

$$\bar{x} = \frac{\sum_{i=1}^k x_i f_i}{n}$$

Ejemplo

Considere la siguiente distribución de frecuencias:

Límites reales	Puntos medios	Frecuencia absoluta	Frecuencia relativa	Frecuencia
				relativa acumulada
17,5 - 22,5	20	4	13,33%	13,33%
22,5 - 27,5	25	5	16,67%	30,33%
27,5 - 32,5	30	8	26,67%	56,67%
32,5 - 37,5	35	5	16,67%	73,33%
37,5 - 42,5	40	3	10,00%	83,33%
42,5 - 47,5	45	2	6,67%	90,00%
47,5 - 52,5	50	2	6,67%	96,67%
52,5 - 57,5	55	1	3,33%	100,00%
Total		30	100,00%	

Calcule la media aritmética.

Solución

Para el cálculo es útil el empleo de una tabla auxiliar:

Puntos medio x_i	Frecuencia absoluta f_i	$x_i f_i$
20	4	80
25	5	125
30	8	240
35	5	175
40	3	120
45	2	90
50	2	100
55	1	55
Total:	30	

Aplicando la fórmula:

$$\bar{x} = \frac{\sum_{i=1}^k x_i f_i}{n} = \frac{985}{30} = 32,83$$

La media es 32,83.

Ejercicio de revisión

Suponga que se tienen los siguientes datos relacionados con el tiempo de espera de los usuarios de un servicio (en minutos), calcule la media:

Límites reales	Frecuencia absoluta
7,5 - 10,5	4
10,5 - 13,5	7
13,5 - 16,5	9
16,5 - 19,5	5
19,5 - 22,5	3
22,5 - 25,5	2
Total	30

Ver solución de este ejercicio en la página de internet de este texto:

www.auladeeconomia.com/raep.html

Aplicación

Índice Dow Jones

Possiblemente en alguna oportunidad usted habrá escuchado noticias sobre el mundo de los negocios que se relacionen con la Bolsa de Valores de Nueva York y entonces habrá escuchado hablar sobre el Dow Jones, el cual es el indicador bursátil más conocido en el planeta.

El Índice Promedio Industrial Dow Jones (DJIA, por sus siglas en inglés) es un promedio ponderado que se calcula a partir de los precios de las acciones de las 30 mayores empresas transadas en la Bolsa de Valores de Nueva York. En el índice se incluyen acciones de empresas como General Electric, Exxon Mobil, DuPont, 3M, Coca Cola, AT&T, Intel, etc.

Originalmente el índice era un promedio simple, es decir, se calculaba dividiendo la suma total (en dólares) del valor de las acciones de las empresas incluidas en el índice dividido entre el número de acciones de las mismas. Sin embargo, fue necesario realizar algunas modificaciones para hacerlo más representativo, y entonces se convirtió en un promedio ponderado, en el cual las empresas más grandes tienen un peso relativamente superior.

En la práctica el valor del DJIA no es tan importante como su variación relativa. Por ejemplo, el lunes 19 de octubre de 1987 este índice sufrió una caída de 508 puntos, los cuales representaron una reducción de 22,6%. Esta fue una de las caídas más fuertes que ha sufrido el índice en toda su historia, de ahí que a ese día se le ha llamado el "lunes negro". A inicios de octubre de 2012 el índice tenía un valor de alrededor de 13.610 puntos, por lo que una contracción de 508 puntos habría representado apenas un 3,7%.

Percentiles y cuartilos

El cálculo de los percentiles y los cuartilos es similar al de la mediana. Igual que en el caso de datos no agrupados, los cuartiles, deciles y quintiles se convierten al percentil equivalente. Para el cálculo se emplea la fórmula siguiente:

$$P_m = L_i + c \left[\frac{\frac{m}{100} n - F_{i-1} \downarrow}{f_i} \right]$$

Ejemplo

Consideré la siguiente distribución de frecuencias:

Límites reales	Puntos medios	Frecuencia absoluta	Frecuencia relativa	Frecuencia relativa acumulada "menos de"
x_i	f_i	$f_r = f_i / n$		
17,5 - 22,5	20	4	13,33%	13,33%
22,5 - 27,5	25	5	16,67%	30,33%
27,5 - 32,5	30	8	26,67%	56,67%
32,5 - 37,5	35	5	16,67%	73,33%
37,5 - 42,5	40	3	10,00%	83,33%
42,5 - 47,5	45	2	6,67%	90,00%
47,5 - 52,5	50	2	6,67%	96,67%
52,5 - 57,5	55	1	3,33%	100,00%
Total	30		100,00%	

Calcule el tercer cuartil.

Solución

El cálculo es muy similar al de la mediana. El tercer cuartil equivale al percentil 75, por lo tanto se puede buscar en la columna de la frecuencia relativa acumulada a menos de aquel valor que es el primero en exceder 75%. Esto se da en quinta clase, por lo que el límite inferior de la clase es 37,5, el intervalo de la clase es 5, la frecuencia acumulada de la clase previa es 22 y la frecuencia de la clase es 3. Aplicando la fórmula:

$$Q_3 = P_{75} = L_i + c \left[\frac{mn/100 - F_{i-1}}{f_i} \right] = 37,5 + 5 \left[\frac{75 \times 30/100 - 22}{3} \right]$$

$$= 37,5 + 5 \cdot \frac{0,5}{3} = 38,33$$

La tercer cuartil es 38,33.

Ejercicio de revisión

Suponga que se tienen los siguientes datos relacionados con el tiempo de espera de los usuarios de un servicio (en minutos), calcule el primer cuartil y el percentil 70:

Límites reales	Frecuencia absoluta
7,5 - 10,5	4
10,5 - 13,5	7
13,5 - 16,5	9
16,5 - 19,5	5
19,5 - 22,5	3
22,5 - 25,5	2
Total	30

Ver solución de este ejercicio en la página de internet de este texto:

www.auladeeconomia.com/raEEP.html

Varianza y desviación estándar

Para el cálculo de la varianza basta modificar las fórmulas vistas anteriormente, donde la primera corresponde a datos de una población y las dos siguientes en el caso de muestras:

$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2 f_i}{N}$$

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 f_i}{n-1}$$

$$s^2 = \frac{\sum_{i=1}^n x_i^2 f_i - \left(\frac{\sum_{i=1}^n x_i f_i}{n} \right)^2}{n-1}$$

La desviación estándar puede seguir calculándose como la raíz cuadrada de la varianza.

Ejemplo

Considere la siguiente distribución de frecuencias:

Límites reales	Puntos medios	Frecuencia absoluta	Frecuencia relativa	Frecuencia relativa acumulada
	x_i	f_i	$f_r = f_i / n$	"menos de"
17,5 - 22,5	20	4	13,33%	13,33%
22,5 - 27,5	25	5	16,67%	30,33%
27,5 - 32,5	30	8	26,67%	56,67%
32,5 - 37,5	35	5	16,67%	73,33%
37,5 - 42,5	40	3	10,00%	83,33%
42,5 - 47,5	45	2	6,67%	90,00%
47,5 - 52,5	50	2	6,67%	96,67%
52,5 - 57,5	55	1	3,33%	100,00%
Total		30	100,00%	

Calcule la varianza y la desviación estándar.

Solución

Se supondrá que los datos corresponden a una muestra y se usará la segunda fórmula de las señaladas anteriormente. También es útil construir una tabla auxiliar. La media se calculó anteriormente y es de 32.83.

Puntos medios	Frecuencia absoluta	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 f_i$
X_i	f_i			
20	4	-12,83	164,69	658,78
25	5	-7,83	61,36	306,81
30	8	-2,83	8,03	64,22
35	5	2,17	4,69	23,47
40	3	7,17	51,36	154,08
45	2	12,17	148,03	296,06
50	2	17,17	294,69	589,39
55	1	22,17	491,36	491,36
	30		Total	2584,17

Aplicando la fórmula:

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 f_i}{n-1} = \frac{2584,17}{30-1} = 89,11$$

La varianza es de 89,11. Para calcular la desviación estándar se saca la raíz cuadrada al resultado anterior:

$$s = \sqrt{s^2} = \sqrt{89,11} = 9,44$$

Ejercicio de revisión

Suponga que se tienen los siguientes datos relacionados con el tiempo de espera de los usuarios de un servicio (en minutos), calcule la varianza y la desviación estándar:

Límites reales	Frecuencia absoluta
7,5 - 10,5	4
10,5 - 13,5	7
13,5 - 16,5	9
16,5 - 19,5	5
19,5 - 22,5	3
22,5 - 25,5	2
Total	30

Ver solución de este ejercicio en la página de internet de este texto:
www.auladeeconomia.com/raEEP.html

Apoyo audiovisual y uso de la tecnología

En la página de internet www.auladeeconomia.com podrá encontrar una presentación de diapositivas que expone este tema y es una parte importante de este texto. Esta presentación presenta el tema en forma visual, pues emplea fotografías, esquemas u otros recursos visuales, e incluso recursos resueltos paso a paso.

Adicionalmente puede encontrar algunos videos explicativos.

Solicite su usuario y contraseña escribiendo al correo electrónico info@auladeeconomia.com. Para ello deberá indicar el número de su factura de compra de texto.

4.4 Ejercicios

Ejercicios de desarrollo

Resuelva los ejercicios que a continuación se le presentan (las respuestas a los ejercicios se encuentran en la página de internet del texto: www.auladeeconomia.com/raeep.html).

1. Un médico ha reunido los siguientes datos sobre la estancia hospitalaria en horas de una muestra de 40 neonatos en la unidad de cuidados intensivos de un hospital:

150	70	95	120	120	72	240	84	300	48
100	120	75	86	90	120	220	180	165	36
48	72	240	270	128	80	110	270	250	112
160	170	120	80	50	60	78	90	48	75

Con base en estos datos, construya una tabla de distribución de frecuencias (con 8 clases) y elabore el histograma para las frecuencias relativas.

2. En los datos recopilados por la enfermera encontró los datos de los siguientes tiempos de espera de los pacientes, en minutos. Los siguientes son algunos de los datos.

15	8	12	20	28	45	43	19	9	5	32	33	36	52	47
34	28	27	10	12	8	15	16	18	19	24	27	28	26	23
38	40	42	44	46	48	40	38	37	37	32	22	26	31	34
32	37	33	35	36	34	31	39	29	51	48	21	38		

Con base en esos datos, construya una tabla de una distribución de frecuencias de 7 clases. Indique el rango, el intervalo de clase, los límites de clase reales, frecuencia absoluta, frecuencia relativa, frecuencia absoluta acumulada menos de y frecuencia relativa acumulada menos de.

3. En una muestra de 12 personas con miopía elevada superior a 12,00 D se midieron algunos parámetros biométricos de sus ojos, como la longitud axial. Los resultados fueron los siguientes:

25	28	26	30	28	27	25	29	31	30	31	33	29	31	29	32	29
----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----

Construya el histograma para esta variable (utilice 6 clases) e indique si la distribución de esta variable es simétrica o asimétrica.

4. Los siguientes son los tiempos de funcionamiento, en horas, de un tipo de componente electrónico: 100, 110, 80, 95, 97, 82, 113, 128, 116, 98, 99, 105, 117, 114, 121, 104, 118, 109, 110, 80, 95, 97, 82, 113, 128, 116, 98, 99, 105, 127, 114, 121, 104, 118, 122, 121, 86, 92, 96, 101. Construya el histograma para las frecuencias absolutas (utilice 5 clases).
5. Un ingeniero ha recopilado datos sobre la vida útil de una muestra de filtros de gasolina del mismo tipo y marca. Los datos son los siguientes (en miles de kilómetros): 12, 14, 16, 15, 10, 20, 13, 15, 16, 14, 12, 13, 13, 15, 16, 13, 14, 14, 12, 8, 10, 16, 14, 12, 13, 11, 13, 9, 19, 22. Construya la tabla de la distribución de frecuencias.
6. La presencia de manganeso en la gasolina puede significar un problema, pues genera ciertos daños a los vehículos, a la salud de las personas y al ambiente. Ante la duda de altas concentraciones de esta sustancia en la gasolina vendida en el país se toman muestras en los distintos planteles de la compañía refinadora de petróleo para que vaya a ser estudiadas en un laboratorio. Los siguientes son datos recopilados para un plantel de la refinadora durante el mes de octubre anterior (concentración de manganeso en mg/L de gasolina):

18	25	20	17	24	12	16	12	18	8
12	13	16	21	19	14	13	22	17	10
24	25	20	19	22	19	16	17	17	10
21	23	26	29	17	14	30	20	18	23

Con base en los datos anteriores:

- Construya la tabla de la distribución de frecuencias (utilice 8 clases).
- Calcule e interprete la media, la mediana y la moda.
- Construya el histograma para las frecuencias absolutas.
- Tomando como base el cálculo de la media, mediana y moda, indique la forma de la distribución de las concentraciones de manganeso (simétrica o asimétrica positiva o negativa). ¿Es consistente ese resultado con lo que se observa en la gráfica?
- Si se considera como aceptable que la gasolina contenga a lo sumo 18 mg/L de manganeso, ¿qué puede usted concluir en este caso?

7. Despues de examinar los registros de facturación mensual de una compañía de ventas por internet, un auditor toma una muestra de 30 cuentas por cobrar, obteniendo las siguientes cantidades que distintos clientes deben a la empresa (en miles de dólares):

9	10	10	10	11	11	11	12	14	15
15	17	17	18	18	19	19	19	19	21
25	26	26	32	33	34	34	37	38	40

- a. ¿Qué monto máximo adeuda el 40% de las cuentas por cobrar observadas?
b. ¿Qué monto mínimo correspondió al 15% de las cuentas por cobrar con mayores montos?
8. Un contador está realizando una auditoría de los cheques extendidos por una compañía. En una muestra de 20 cheques se obtuvieron los siguientes montos, en miles dólares:

15, 17, 22, 8, 4, 2, 16, 6, 8, 14, 15, 25, 20, 5, 6, 8, 12, 19, 11, 12

- a. Construya la tabla de la distribución de frecuencias de los montos de los cheques extendidos por la compañía.
b. Construya el histograma para las frecuencias relativas de los montos de los cheques extendidos por la compañía.
c. Use la Calcula la media, mediana y moda de los montos de los cheques.
9. Los siguientes datos corresponden a la demanda, en unidades al mes, que los consumidores de un producto estarían dispuestos a comprar a un determinado precio: 5, 8, 12, 14, 21, 18, 6, 14, 15, 24, 32, 17, 15, 4, 5, 9, 7, 28, 26, 22, 20, 17, 15, 14, 13, 11, 8, 9, 16, 10, 11, 12, 16, 15, 17, 12, 13, 14, 18, 15, 12, 11, 16.
- a. Construya la tabla de la distribución de frecuencias con 5 clases.
b. Construya la tabla de la distribución de frecuencias con 10 clases.
c. Construya el histograma para cada una de las tablas de los incisos anteriores. Comente las diferencias.
10. Dado el siguiente conjunto de datos sobre el número de turistas que se hospedan en un hotel por semana en una muestra de 22 semanas seleccionadas al azar:

25	81	92	44	58	62	34	40	90	57	42
40	57	57	57	57	62	40	65	58	50	18

Construya la tabla de la distribución de frecuencias (con 6 clases) y a partir de ella calcule e intérprete cada una de las siguientes medidas descriptivas: coeficiente de variación y desviación cuartil.

11. Los siguientes datos corresponden a una muestra de personas que aplicaron un test para medir su cociente intelectual: 102, 98, 95, 100, 99, 112, 125, 85, 83, 100, 98, 129, 78, 100, 116, 101, 102, 101, 96, 100, 103, 97, 100, 97, 98, 99, 103, 102, 95, 101, 98, 100, 98, 99, 101, 102, 99, 103, 101. Construya la tabla de la distribución de frecuencias y con base en ella calcule los percentiles 20, 40, 60, 80 y 90.

12. Una trabajadora social está preocupada por el consumo de drogas ilegales entre los estudiantes de secundaria de una comunidad marginal del país llamada ciudadela Los Robles. Una muestra de 18 estudiantes varones del liceo local reportó las siguientes cifras como la edad en que se inició el consumo de drogas ilegales:

14	17	16	13	15	16	13	14	15
12	16	18	15	14	11	13	15	8
12	15	16	15	19	9	10	11	10

Con base en los datos anteriores, construya la tabla de la distribución de frecuencias (usando 5 clases) y a partir de ella calcule e intérprete cada una de las siguientes medidas descriptivas:

- a. Moda
b. Mediana
c. Media aritmética
d. Cuartil 3
e. Percentil 85
f. Desviación estándar

13. Con relación a los datos suministrados a cerca de la siguiente distribución de frecuencias, complete la tabla.

14. Con relación a la tabla realice lo siguiente:

Límites reales	Frecuencia absoluta
28,5 - 30,5	16
30,5 - 32,5	68
32,5 - 34,5	33
34,5 - 36,5	12
36,5 - 38,5	8
38,5 - 40,5	5
40,5 - 42,5	1

- Calcule: moda, percentil 90, tercer cuartil, coeficiente de variación.
- Construya las ojivas a menos de y a más de.
- Suponga que los datos mostrados en la tabla corresponden a una muestra de las cantidades en mililitros de un producto que una empresa envasa.
 - ¿Qué puede concluirse sobre la simetría de la distribución al observar la relación entre moda, mediana y media y qué importancia tiene eso?
 - Si se define la distancia intercuartílica como la diferencia entre el tercer cuartil y el primer cuartil, calcule la diferencia o distancia intercuartílica, e indique si sería mejor que fuera menor o que fuera mayor. Explique.
 - Si la etiqueta del producto dice que el envase debe contener 40 ml y la empresa tolera como aceptable una diferencia de 2 ml hacia arriba o hacia abajo, ¿qué puede decirse del grado de cumplimiento de esta especificación?

15. Con relación a la tabla realice lo siguiente:

Límites reales	Frecuencia absoluta
28,5 - 30,5	6
30,5 - 32,5	48
32,5 - 34,5	33
34,5 - 36,5	16
36,5 - 38,5	8
38,5 - 40,5	5
40,5 - 42,5	1

- Calcule: moda, mediana, percentil 70, tercer cuartil, coeficiente de variación.
- Construya el histograma.
- Suponga que los datos mostrados en la tabla corresponden a una muestra de los pesos en gramos de un producto que una empresa envasa.
 - ¿Qué puede concluirse al observar la relación entre moda, mediana y media?
 - Interprete el tercer cuartil.
 - Con respecto al coeficiente de variación, ¿sería mejor que fuera menor o que fuera mayor? Explique.
 - Si la etiqueta del producto dice que el envase debe contener 40 gramos y la empresa tolera como aceptable una diferencia de 2 gramos hacia arriba o hacia abajo, ¿qué puede decirse del grado de cumplimiento de esta especificación?

16. Con relación a los datos suministrados a cerca de la siguiente distribución de frecuencias, complete la tabla.

Limites reales	Puntos medios	Frecuencia absoluta	Frecuencia relativa	Frecuencia absoluta acumulada "menos de"	Frecuencia absoluta acumulada "más de"	Frecuencia relativa acumulada "menos de"	Frecuencia relativa acumulada "más de"
?	?	?	?	?	?	?	?
?	?	?	?	?	?	?	0.9
?	?	?	0.45	?	?	0.7	?
?	?	?	?	?	?	?	?
?	85	?	?	?	4	?	0.05
Total		?					

Además se sabe que todas las clases tienen el mismo intervalo de clase, que es de 9.

17. Con relación a los datos suministrados a cerca de la siguiente distribución de frecuencias, complete la tabla.

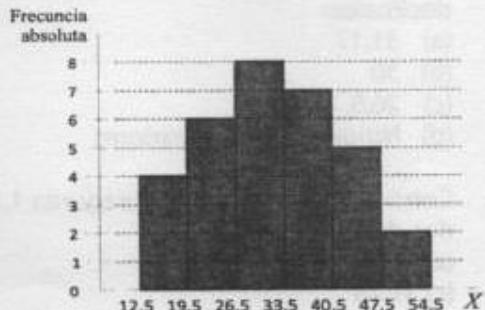
Limites reales	Puntos medios	Frecuencia absoluta	Frecuencia relativa	Frecuencia absoluta acumulada "menos de"	Frecuencia absoluta acumulada "más de"	Frecuencia relativa acumulada "menos de"	Frecuencia relativa acumulada "más de"
?	?	?	?	?	?	?	?
?	?	?	?	?	?	?	0.92
?	?	?	0.45	?	?	0.7	?
?	?	?	?	?	?	?	?
?	85	?	?	?	4	?	0.04
Total		?					

Además se sabe que todas las clases tienen el mismo intervalo de clase, que es de 7.

Examen del capítulo:

En cada caso seleccione la opción que mejor contesta cada pregunta. (las respuestas a los ejercicios se encuentran en la página de internet del texto: www.auladeeconomia.com/raEEP.html).

1. Considere el siguiente gráfico:



- Con respecto a esta gráfica es falso con toda certeza que:
- (a) La variable X es cuantitativa continua
 - (b) La gráfica es un polígono de frecuencias
 - (c) Los datos corresponden a una población
 - (d) La gráfica corresponde a un histograma
2. Con relación a la gráfica de la pregunta 1, es verdadero que:
- (a) El intervalo de clase es 12,5
 - (b) El valor de n es 32
 - (c) El punto medio de la segunda clase es 24
 - (d) El límite superior de la cuarta clase es 38
3. Con relación a la gráfica de la pregunta 1, es verdadero que:
- (a) La frecuencia relativa acumulada de la cuarta clase es 0,78125
 - (b) La frecuencia relativa de la sexta clase es $2/30$ pues hasta ahí se acumulan 30 datos
 - (c) El punto medio de la segunda clase es 24
 - (d) La frecuencia absoluta acumulada de la tercera clase es 8
4. Suponga que la variable X de la gráfica de la pregunta 1 corresponde al tiempo, en segundos, entre la llegada de dos autos consecutivos a un peaje en una autopista durante periodo aleatoriamente seleccionado. Con respecto a esta afirmación es falso con toda certeza que:
- (a) Los datos no son confiables pues la muestra es muy pequeña
 - (b) El 56,25% de los tiempos entre la llegada de dos autos es de 33,5 segundos
 - (c) La mayoría de los tiempos registrados se da entre 26,5 y 33,5 segundos
 - (d) Los tiempos de llegada entre dos autos sucesivos nunca son mayores a 54,5 segundos
5. Considere el gráfico de la pregunta 1, la media aritmética es:
- (a) 33,5
 - (b) 31,97
 - (c) 1023
 - (d) Ninguna de las anteriores
6. Considere el gráfico de la pregunta 1, la mediana es:
- (a) 26,5
 - (b) 28,83
 - (c) 31,75
 - (d) Ninguna de las anteriores
7. Considere el gráfico de la pregunta 1, la moda es, redondeando a dos decimales:
- (a) 31,17
 - (b) 30
 - (c) 26,5
 - (d) Ninguna de las anteriores
8. Considere el gráfico de la pregunta 1, el primer cuartil es, redondeando a dos decimales:
- (a) 24,17
 - (b) 22,30
 - (c) 31,75
 - (d) Ninguna de las anteriores

Con respecto a las dos afirmaciones anteriores, es correcto que:

11. El gerente de un centro de llamadas desea evaluar el desempeño de los agentes de servicio y para ello decide basarse en los tiempos de espera de los clientes para ser atendidos (medido en segundos) y el grado de satisfacción que los clientes manifiesten al recibir el servicio (valorado como bueno, regular o malo). La semana anterior tomó una muestra de 12 llamadas por día de lunes a miércoles y 20 llamadas el jueves y el viernes.

La siguiente tabla corresponde a las frecuencias de los tiempos de espera de los clientes para la muestra de llamadas atendidas durante el día jueves.

Las preguntas de la 11 a la 22 se basan en esta tabla.

Límites reales	Frecuencia absoluta
10,5 - 14,5	2
14,5 - 18,5	4
18,5 - 22,5	8
22,5 - 26,5	5
26,5 - 30,5	1

El punto medio de la tercera clase es:

12. La frecuencia porcentual de la cuarta clase es:

(a) 25% (b) 30%
(c) 95% (d) 5%

14. La frecuencia relativa acumulada a más de la tercera clase es:
- (a) 14
 - (b) 40%
 - (c) 0,70
 - (d) Ninguna de las anteriores
15. La frecuencia absoluta acumulada a menos de correspondiente a la tercera clase significa que:
- (a) 14 clientes esperaron 18,5 segundos o más
 - (b) 14 clientes esperaron 18,5 segundos o menos
 - (c) 14 clientes esperaron 14,5 segundos o más
 - (d) Ninguna de las anteriores
16. La gráfica apropiada para representar las frecuencias absolutas relacionadas con sus puntos medios se llama:
- (a) Gráfico de barras horizontales
 - (b) Polígono de frecuencias
 - (c) Ojiva a menos de
 - (d) Diagrama de frecuencias acumuladas
17. El tiempo medio de espera de los clientes de la muestra del día jueves es, en segundos:
- (a) 20,5
 - (b) 20,3
 - (c) 18,5
 - (d) Ninguna de las anteriores
18. La mediana del tiempo de espera de la muestra del día jueves es, en segundos:
- (a) 20,5
 - (b) 20,3
 - (c) 18,5
 - (d) Ninguna de las anteriores
19. La desviación estándar del tiempo de espera de la muestra del día jueves es, en segundos:
- (a) 4,2
 - (b) 17,64
 - (c) 1,21
 - (d) Ninguna de las anteriores
20. El primer cuartil del tiempo de espera de la muestra del día jueves es, en segundos:
- (a) 17,5
 - (b) 16,5
 - (c) 21,5
 - (d) Ninguna de las anteriores
21. El percentil 78 del tiempo de espera de la muestra del día jueves es, en segundos:
- (a) 23,78
 - (b) 19,78
 - (c) 22,88
 - (d) Ninguna de las anteriores
22. Con respecto al tercer cuartil para los tiempos de espera del día jueves es verdadero que:
- (a) Un 25% de los clientes de ese día esperaron 23,3 segundos o menos
 - (b) Un 75% de los clientes de ese día esperaron 23,3 segundos o menos
 - (c) Un 25% de los clientes de ese día esperaron 22,5 segundos o menos
 - (d) Un 75% de los clientes de ese día esperaron 22,5 segundos o menos

edad se soparece. Aun así, sobre el tema no se habló mucho. Si el tema de probabilidad es algo que no te interesa, es porque no te interesa la probabilidad. Sin embargo, las probabilidades no tienen que ver con lo que te interesa. La probabilidad es una medida de la posibilidad de que algo suceda.

CAPÍTULO 5

INTRODUCCIÓN A LAS PROBABILIDADES

OBJETIVOS

Al concluir el capítulo, será capaz de:

1. Reconocer la importancia y uso del concepto de probabilidad
2. Aplicar conceptos básicos de conteo
3. Calcular probabilidades empleando la definición clásica de probabilidad
4. Aplicar los principales teoremas y axiomas de probabilidad

Este capítulo introduce el concepto de probabilidad y sus aplicaciones más comunes.

Se presentan conceptos básicos de probabilidad, como la definición clásica de probabilidad, la probabilidad condicional y la probabilidad conjunta. Se muestra cómo calcular la probabilidad de un evento dado un otro evento. Se presentan ejemplos de aplicaciones de la probabilidad en la vida cotidiana, como la probabilidad de que llueva mañana o la probabilidad de que un paciente reciba una enfermedad determinada.

Este capítulo también introduce el concepto de probabilidad condicional y la probabilidad conjunta. Se muestra cómo calcular la probabilidad de un evento dado un otro evento. Se presentan ejemplos de aplicaciones de la probabilidad en la vida cotidiana.

Este capítulo introduce el concepto de probabilidad y sus aplicaciones más comunes.

5.1 Probabilidad

La probabilidad es un tema de todos los días. Cada vez que se habla del clima, por ejemplo si va a llover o no en un día determinado, o bien, la posibilidad de sufrir un accidente, está implícito el concepto de probabilidad. En general, se habla de probabilidad en cualquier situación en la que no haya certeza del resultado.

PROBABILIDAD Una probabilidad es un valor numérico de la incertidumbre de que un suceso específico pueda ocurrir.

Las probabilidades se pueden obtener como resultado de experimentos realizados aleatoriamente, como en los juegos de azar, en donde todos los elementos tienen la misma posibilidad de ser seleccionados. Por ejemplo, al lanzar un dado, cada uno de los números tiene la misma posibilidad de salir. Ahora bien, las probabilidades también se pueden obtener en forma empírica, o sea, de la observación de la frecuencia relativa con la que ocurre un fenómeno en particular. Por ejemplo, al tratar de determinar la probabilidad de que ocurra un accidente de aviación se puede observar la frecuencia relativa con la que este evento se presenta.

Cuando se habla de probabilidad se menciona el concepto de evento o suceso. Los eventos son los resultados de experimentos aleatorios. Por ejemplo, cuando se lanza un dado, un evento es que salga el 4, o bien, en el caso de la fabricación de un artículo el hecho de que éste tenga un defecto menor puede ser un evento, los otros eventos posibles son que tenga un defecto mayor, que esté en perfecta condición, etc.

EVENTO Es un posible resultado de un experimento aleatorio.

Los eventos pueden ser simples o compuestos. Un evento simple es aquel que no se puede descomponer en otros eventos, mientras que el evento compuesto sí se puede descomponer en otros eventos. En el caso del lanzamiento de un dado, un evento simple es que al tirar el dado salga un cuatro. Un evento compuesto es que al tirar el dado, salga un número par. Este es un evento compuesto porque se compone de los eventos dos, cuatro y seis.

Por otro lado, el espacio muestral de un experimento aleatorio es el conjunto de todos los posibles eventos que pueden ocurrir. Por ejemplo en el lanzamiento de un dado el espacio muestral está dado por los eventos 1, 2, 3, 4, 5, 6.

ESPACIO MUESTRAL Conjunto de todos los eventos simples posibles de un experimento aleatorio.

Material audiovisual

En la página de internet de este texto podrá encontrar una presentación y un video que exponen los conceptos y teoremas de probabilidad.

Solicite su usuario y contraseña escribiendo al correo electrónico info@auladeeconomia.com. Para ello deberá indicar el número de su factura de compra de texto.



5.2 Conteo

Muchas veces para calcular probabilidades es necesario enumerar sucesos y sus respectivas posibilidades, para lo cual es necesario aplicar principios de análisis combinatorio. Por lo tanto se repasan tres conceptos básicos:

Principio de multiplicación de conteo

Si una actividad requiere una primera elección que se puede hacer de n_1 formas distintas, una segunda elección que se puede hacer de n_2 formas diferentes, hasta una k -ésima elección que se puede hacer de n_k formas distintas, entonces la actividad puede ser realizada de $n_1 \times n_2 \times \dots \times n_k$ formas diferentes.

Ejemplo

Si usted invita a 8 personas a comer y hay una mesa con 8 sillas, ¿de cuántas formas distintas pueden sentarse a la mesa?

Solución

La primer persona que se sienta dispone de 8 posibilidades, la segunda de sólo 7 (ya que la primera ya se sentó), la tercera tiene 6 posibilidades, la cuarta 5 y así sucesivamente. Por tanto se pueden sentar de:

$$8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 40320 \text{ formas distintas}$$

Ejercicio de revisión

1. Un restaurante ofrece las siguientes opciones para almorzar:
Tres tipos de plato fuerte: pollo, res, chuleta
Dos tipos de refrescos: frutas, cola
Dos tipos de postre: flan, helado
¿Cuántas órdenes distintas pueden efectuarse?
2. Si una contraseña para retirar dinero de un cajero automático se compone de 4 dígitos. ¿Cuántas contraseñas distintas son posibles?

Ver solución de este ejercicio en la página de internet de este texto:
www.auladeeconomia.com/raeep.html

Factorial: El factorial de un número n , que se denota $n!$, se define como:

$$n! = n \times (n-1) \times (n-2) \times \dots \times 2 \times 1$$

Es importante señalar que se define $1! = 1$ y $0! = 1$.

Ejemplo

Calcule el factorial de 5.

Solución

El factorial de 5 es:

$$5! = 5 \times 4 \times 3 \times 2 \times 1 = 120$$

Ejercicio de revisión

Calcule el factorial de los siguientes números:

1. $5! =$
2. $6! =$
3. $10! =$
4. $0! =$
5. $1! =$
6. $70! =$
7. $20! =$

Ver solución de este ejercicio en la página de internet de este texto:
www.auladeeconomia.com/raeep.html

Uso de Excel y Minitab para el cálculo del factorial

Ejemplo

Use Excel y Minitab para calcular el factorial de 5.

Solución

En Excel se emplea la función FACT, la cual tiene la siguiente sintaxis:

$$=FACT(número)$$

Donde "número" indica la celda donde se halla el número del cual se desea calcular el factorial, o bien, simplemente se escribe dicho número. Entonces, en este caso se digita en la celda en que se desea obtener el resultado la función:

$$=FACT(5)$$

Y así se obtiene el resultado 120.

En Minitab se requiere dar clic al menú Calc y elegir Calculadora. En el cuadro de diálogo se debe completar la columna de la hoja de trabajo en la cual se desea almacenar el resultado, por ejemplo, la columna C1. Luego en expresión se emplea la función FACTORIAL, la cual emplea la sintaxis:

FACTORIAL(número de elementos)

Donde "número de elementos" es el número del cual se desea obtener el factorial, o bien, la columna en la que se hallan esos números. En este caso, si se indica solo el número, entonces la función quedaría:

FACTORIAL(5)

Después se da clic en Aceptar y el resultado 120 se obtiene en la hoja de trabajo en la celda que se haya indicado.

Permutaciones

PERMUTACIONES Una permutación es un arreglo ordenado de n elementos distintos tomados r a la vez sin repetición.

El número de permutaciones, que se denota $P(n, r)$, se calcula como:

$$P(n, r) = \frac{n!}{(n - r)!}$$

Ejemplo

Calcule el número de permutaciones de 5 elementos tomados de 3 en 3.

Solución

Se tiene que $n = 5$ y $r = 3$:

$$P(5, 3) = \frac{5!}{(5 - 3)!} = \frac{5!}{2!} = \frac{120}{2} = 60$$

Ejercicio de revisión

Calcule las siguientes permutaciones:

1. $P(8, 5) =$
2. $P(6, 0) =$
3. $P(10, 1) =$
4. $P(5, 5) =$
5. $P(300, 1) =$
6. $P(200, 2) =$
7. $P(n, n) =$
8. $P(n, 1) =$

9. $P(n, 0) =$
10. $P(n, n - 1) =$

Ver solución de este ejercicio en la página de internet de este texto:
www.auladeeconomia.com/raeep.html

Uso de Excel y Minitab para el cálculo de permutaciones

Ejemplo

Use Excel y Minitab para calcular el número de permutaciones de 5 elementos tomados de 3 en 3.

Solución

En Excel se emplea la función PERMUTACIONES, la cual tiene la siguiente sintaxis:

=PERMUTACIONES(número; tamaño)

Donde "número" indica la celda donde se halla el valor de n , o bien, simplemente se escribe dicho valor de n . Luego "tamaño" es la celda en la cual se haya el valor de r o simplemente el valor de r . Entonces, en este caso se digita en la celda en que se desea obtener el resultado la función:

=PERMUTACIONES(5; 3)

Y así se obtiene el resultado 60.

En Minitab se requiere dar clic al menú Calc y elegir Calculadora. En el cuadro de diálogo se debe completar la columna de la hoja de trabajo en la cual se desea almacenar el resultado, por ejemplo, la columna C1. Luego en expresión se emplea la función PERMUTATIONS, la cual emplea la sintaxis:

PERMUTACIONES(número de elementos.número para elegir)

Donde "número de elementos" es el valor de n , o bien, la columna en la que se halla el valor de n . Luego, "número para elegir" es el valor de r , o la columna en la que se encuentra el valor de r . En este caso, si se indican solo los números, entonces la función quedaría:

PERMUTACIONES(5.3)

Después se da clic en Aceptar y el resultado 60 se obtiene en la hoja de trabajo en la celda que se haya indicado.

Combinaciones

COMBINACIONES Una combinación es un arreglo no ordenado de n elementos distintos tomados r a la vez sin repetición.

El número de combinaciones, que se denota $C(n, r)$, se calcula como:

$$C(n, r) = \frac{n!}{r!(n-r)!}$$

Ejemplo

Calcule el número de combinaciones de 5 elementos tomados 3 en 3

Solución

Se tiene que $n = 5$ y $r = 3$:

$$C(5, 3) = \frac{5!}{3!(5-3)!} = \frac{5!}{3! \cdot 2!} = \frac{120}{6 \cdot 2} = 10$$

La expresión $\binom{n}{r}$ es conocida como coeficiente binomial y es igual a $C(n, r)$.

Ejercicio de revisión

Calcule las siguientes probabilidades:

1. $C(8, 5) =$
2. $C(6, 0) =$
3. $C(10, 1) =$
4. $C(5, 5) =$
5. $C(300, 1) =$
6. $C(200, 2) =$
7. $C(n, n) =$
8. $C(n, 1) =$
9. $C(n, 0) =$
10. $C(n, n - 1) =$

Resuelva los siguientes ejercicios:

¿Cuántas directivas de tres miembros (presidente, secretario y tesorero) se pueden formar de un grupo de 8 personas elegibles?

¿Cuántos comités de tres miembros se pueden formar de un grupo de 8 personas elegibles?

¿Cuántos comités de tres estudiantes y dos profesores se pueden formar si hay un grupo de 10 estudiantes y 5 profesores elegibles?

Ver solución de este ejercicio en la página de internet de este texto:
www.auladeeconomia.com/raeep.html

Uso de Excel y Minitab para calcular combinaciones

Ejemplo

Use Excel y Minitab para calcular el número de combinaciones de 5 elementos tomados de 3 en 3.

Solución

En Excel se emplea la función COMBINAT, la cual tiene la siguiente sintaxis:

=COMBINAT(número; tamaño)

Donde "número" indica la celda donde se halla el valor de n , o bien, simplemente se escribe dicho valor de n . Luego "tamaño" es la celda en la cual se haya el valor de r o simplemente el valor de r . Entonces, en este caso se digita en la celda en que se desea obtener el resultado la función:

=COMBINAT(5; 3)

Y así se obtiene el resultado 10.

En Minitab se requiere dar clic al menú Calc y elegir Calculadora. En el cuadro de diálogo se debe completar la columna de la hoja de trabajo en la cual se desea almacenar el resultado, por ejemplo, la columna C1. Luego en expresión se emplea la función COMBINATIONS, la cual emplea la sintaxis:

COMBINATIONS(número de elementos.número para elegir)

Donde "número de elementos" es el valor de n , o bien, la columna en la que se halla el valor de n . Luego, "número para elegir" es el valor de r , o la columna en la que se encuentra el valor de r . En este caso, si se indican solo los números, entonces la función quedaría:

COMBINATIONS(5.3)

Después se da clic en Aceptar y el resultado 10 se obtiene en la hoja de trabajo en la celda que se haya indicado.

Permutaciones con elementos repetidos:

La definición anterior de permutaciones señalaba que los elementos debían ser distintos, pero cuando en el total de elementos hay algunos grupos de elementos iguales y se desea calcular el número de permutaciones, entonces cambia el cálculo del número de permutaciones. Si hay n elementos en total y n_1 elementos iguales entre sí, n_2 elementos iguales entre sí, hasta n_k

elementos iguales entre sí, entonces el número de permutaciones se calcula como:

$$\frac{n!}{n_1! \cdot n_2! \cdot \dots \cdot n_k!}$$

Ejemplo

Suponga que en un grupo de 10 bolas hay 5 de color rojo, 3 azules y dos blancas, ¿cuántas permutaciones son posibles?

Solución

Aplicando la fórmula de permutaciones con elementos repetidos:

$$\frac{n!}{n_R! \cdot n_A! \cdot n_B!} = \frac{10!}{5! \cdot 3! \cdot 2!} = \frac{3628800}{120 \cdot 6 \cdot 2} = \frac{3628800}{1440} = 2520$$

5.3 Definición de probabilidad

Definición clásica de probabilidad

El concepto *clásico* o matemático de probabilidad, se originó en los juegos de azar, y parte de si un experimento aleatorio puede tener N resultados igualmente posibles y mutuamente excluyentes, como en el caso del lanzamiento de un dado, pues hay $N = 6$ resultados posibles que son igualmente posibles y mutuamente excluyentes (o sea, que no pueden salir dos números distintos al mismo tiempo); y si a de estos resultados corresponden a un evento A , entonces la probabilidad del evento A , denotada $P(A)$, es:

$$P(A) = \frac{\text{Número de casos favorables al evento } A}{\text{Número de casos posibles}} = \frac{a}{N}$$

En esta definición la probabilidad de un evento determinado representa la proporción de veces que se espera que ocurra ese resultado específico si el experimento se realiza un gran número de veces bajo idénticas circunstancias.

Ejemplo

¿Cuál es la probabilidad de que al tirar un dado perfecto se obtenga un número par mayor que 2?

Solución

Se define el evento A como obtener un número par mayor de 2. Al tirar el dado los seis lados tienen igual posibilidad de quedar hacia arriba. Los números pares mayores que 2 son 4 y 6, por lo tanto:

$$P(A) = \frac{a}{N} = \frac{2}{6} = 0,3333$$

Ejercicio de revisión

Se lanzan dos dados y se suman los puntos. Si X es la suma de los puntos, calcule las siguientes probabilidades:

1. $P(X = 3) =$
2. $P(X = 6) =$
3. $P(X = 7) =$
4. $P(X = 11) =$
5. $P(X = 12) =$
6. $P(X = 15) =$

Ver solución de este ejercicio en la página de internet de este texto:
www.auladeeconomia.com/raep.html

Este concepto de probabilidad tiene algunas limitaciones. Por ejemplo, cuando no es posible conocer el número de resultados posibles (espacio muestral), o bien, este espacio muestral es infinito, o cuando los eventos no son mutuamente excluyentes ni igualmente posibles.

5.4 La frecuencia relativa como probabilidad

En muchas situaciones, los resultados de un experimento no tienen la misma posibilidad de ocurrir, por lo que las probabilidades deben asignarse con base en una evidencia experimental o con base en datos históricos. En este caso se dice que la probabilidad del evento se puede determinar como la frecuencia relativa de ocurrencia de dicho suceso, es decir, su frecuencia absoluta con respecto al total de casos observados:

$$P(A) = Fr(A) = \frac{F(A)}{N}$$

Así, por ejemplo, al decir que la probabilidad de obtener una pieza defectuosa en un proceso es del 6%, se está usando el valor de la frecuencia relativa que se obtuvo al llevar un registro de las piezas defectuosas durante un período razonablemente largo de tiempo.

Ejemplo

En un lote de 3.000 piezas producidas en una máquina se encontraron 96 defectuosas. Calcule la probabilidad de piezas defectuosas de esa máquina.

Solución

Si d es el evento obtener una pieza defectuosa, entonces su frecuencia es 96, lo que da la probabilidad:

$$P(d) = \frac{96}{3.000} = 0,0320$$

Ejercicio de revisión

En una ciudad en la que habitan 5.000 personas, se sabe que 2.700 son mujeres. Si se selecciona una persona al azar, ¿cuál es la probabilidad de que sea mujer?

Ver solución de este ejercicio en la página de internet de este texto:
www.auladeeconomia.com/raeep.html

5.5 Concepto intuitivo o subjetivo de probabilidad

En muchas situaciones existe poca o ninguna información numérica para determinar la probabilidad. Entonces se emplean los juicios u opiniones basados en un conocimiento profundo de la situación, e incluso basados en la intuición, acerca de la probabilidad de que ocurra un evento. Por ejemplo, el caso en que un especialista emite su criterio con relación al estado futuro de la economía y estima una probabilidad de que haya recesión y otra probabilidad de que haya expansión.

Ejercicio de revisión

En cada caso, indique cuál enfoque se emplearía para determinar la probabilidad de que el evento dado ocurra:

- Ganar en un juego de ruleta.
- Enfermar de cáncer de piel.
- Que un nuevo producto desarrollado por una compañía sea un éxito.
- Que la realización de un proyecto dure más de lo esperado.
- Que una computadora nueva falle en un plazo de tres años o menos.

Ver solución de este ejercicio en la página de internet de este texto:
www.auladeeconomia.com/raeep.html

5.6 Axiomas y teoremas de probabilidad

Estos se basan en el concepto matemático o clásico de probabilidad. Se presentan algunos ejemplos para ilustrar los conceptos.

Axioma I: Para cualquier evento A , la probabilidad no puede ser menor que 0, ni mayor que 1:

$$0 \leq P(A) \leq 1$$

Por ejemplo, si un club cuenta con 1.200 miembros de ambos sexos ($N = 1200$). El evento de interés (A) es ser mujer, por tanto $P(A) = a/N$. El mínimo

número de mujeres en el club es cero ($a = 0$), ninguna mujer en el club, en cuyo caso $P(A) = 0/1200 = 0$. El máximo número posible de mujeres es 1200, no hombres en el club, por lo que $P(A) = 1200/1200 = 1$. No puede haber menos de 0 ni más de 1200 mujeres en el grupo de 1200 personas, por lo tanto:

$$0 \leq P(A) \leq 1$$

De lo anterior se puede concluir que:

- La probabilidad de un evento imposible es cero.
- La probabilidad de un evento que ocurre con toda certeza es uno.

Para el siguiente axioma hay que recordar el concepto de espacio muestral, el cual es el conjunto de todos los resultados posibles de un experimento aleatorio.

Axioma 2. La suma de la probabilidad de todos los eventos A , contenidos en el espacio muestral S es 1, o sea, $P(S) = 1$.

Donde S es el espacio muestral, o sea, el conjunto total de eventos simples posibles, y equivale a $\Sigma P(A)$. Esto quiere decir que la suma de las probabilidades de todos los eventos posibles es igual a uno.

Para el axioma siguiente es necesario introducir el concepto de eventos excluyentes.

EVENTOS EXCLUYENTES Dos eventos son excluyentes si no pueden ocurrir simultáneamente.

Por ejemplo, al jugar un partido de futbol hay tres resultados posibles, ganar, empatar o perder, los cuales son excluyentes, pues dos o más de esos eventos no pueden ocurrir simultáneamente. De modo similar se pueden definir los eventos no excluyentes.

EVENTOS NO EXCLUYENTES Dos eventos son excluyentes si pueden ocurrir simultáneamente.

Si se tiene el evento ser mujer y el evento estudiar ingeniería industrial, entonces estos son eventos no excluyentes, porque pueden ocurrir simultáneamente, es decir, una mujer puede haber estudiado ingeniería industrial.

Axioma 3. Si A y B son eventos mutuamente excluyentes en el espacio muestral S , entonces,

$$P(A \cup B) = P(A) + P(B)$$

Más sobre axiomas de probabilidad y su relación con la estadística

comúnmente se dice que la probabilidad de que ocurra el evento A es igual a

Ejemplo

En una bodega hay 400 televisores (T), 100 videogramadoras (V), 200 cámaras fotográficas (F) y 300 computadoras (C). Si se selecciona un aparato al azar, ¿cuál es la probabilidad de que sea un televisor o una computadora?

Solución

Dado que los eventos televisor (T) y computadora (C) son excluyentes se calcula cada probabilidad por separado y se suman ambas probabilidades. Además, en la bodega hay un total de 1000 aparatos:

$$P(T \text{ o } C) = \frac{400}{1000} + \frac{300}{1000} = \frac{700}{1000} = 0,7$$

Ejercicio de revisión

Suponga que se tiene un grupo de 40 ingenieros que se agrupan por carrera y por sexo según la tabla. Suponga que ninguno ha estudiado dos carreras. Si se selecciona al azar un profesional, ¿cuál es la probabilidad de que sea ingeniero civil o ingeniero industrial?

	Industrial	Civil	Electrónica	Otras	Total
Masculino	8	6	6	6	26
Femenino	7	2	4	1	14
Total	15	8	10	7	40

Ver solución de este ejercicio en la página de internet de este texto:
www.auladeeconomia.com/raeep.html

Por inducción matemática el axioma 3 conduce a los siguientes teoremas:

Teorema 1. Si E_1, E_2, \dots, E_m son eventos mutuamente excluyentes en el espacio muestral S , entonces:

$$P(E_1 \text{ o } E_2 \text{ o } \dots \text{ o } E_m) = P(E_1) + P(E_2) + \dots + P(E_m)$$

Ejemplo

En una bodega hay 400 televisores (T), 100 videogramadoras (V), 200 cámaras fotográficas (F) y 300 computadoras (C). Si se selecciona un aparato al azar, ¿cuál es la probabilidad de que sea un televisor o una computadora o una cámara fotográfica?

Solución

Dado que los eventos televisor (T), computadora (C) y cámara fotográfica (F) son excluyentes se calcula cada probabilidad por separado y se suman ambas probabilidades. Además, en la bodega hay un total de 1000 aparatos:

$$P(T \text{ o } C \text{ o } F) = \frac{400}{1000} + \frac{300}{1000} + \frac{200}{1000} = \frac{900}{1000} = 0,9$$

Teorema 2. Regla general de la suma o adición de probabilidades. Si A y B son eventos cualesquiera (excluyentes o no excluyentes) en un espacio muestral S , la probabilidad de la unión de los eventos (o sea, que ocurra un evento o el otro) está dada por:

$$P(A \text{ o } B) = P(A) + P(B) - P(AB)$$

Donde $P(AB)$ es la probabilidad de la intersección de los dos eventos, o sea, la probabilidad de que los eventos ocurran simultáneamente.

Ejemplo

En una bodega hay 400 televisores (T), 100 videogramadoras (V), 200 cámaras fotográficas (F) y 300 computadoras (C). Además, se tienen algunos datos sobre su nivel de calidad, como perfectos (P) o con defectos (D), según la tabla:

	Tipo de aparato				Total
	T	V	F	C	
P	350	80	150	270	850
D	50	20	50	30	150
Total	400	100	200	300	1000

Si se selecciona un aparato al azar, ¿cuál es la probabilidad de que sea un televisor o que tenga defectos?

Solución

En la bodega hay un total de 1000 aparatos. Dado que los eventos televisor (T) y que el aparato tenga defectos (D) no son excluyentes se calcula cada probabilidad por separado y se suman ambas probabilidades, pero también se resta la probabilidad de que ocurran a la vez:

$$P(T \text{ o } D) = \frac{400}{1000} + \frac{150}{1000} - \frac{50}{1000} = \frac{500}{1000} = 0,5$$

Ejercicio de revisión

Suponga que se tiene un grupo de 40 ingenieros que se agrupan por carrera y por sexo según la tabla. Suponga que ninguno ha estudiado dos carreras. Si se selecciona al azar un profesional, ¿cuál es la probabilidad de que sea ingeniero civil o mujer?

	Industrial	Civil	Electrónica	Otras	Total
Masculino	8	6	6	6	26
Femenino	7	2	4	1	14
Total	15	8	10	7	40

Ver solución de este ejercicio en la página de internet de este texto:
www.auladeeconomia.com/raeep.html

El teorema 2 puede generalizarse a tres eventos:

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC)$$

Para el teorema siguiente se requiere el concepto de eventos complementarios.

EVENTOS COMPLEMENTARIOS Dos eventos A y B son complementarios si su unión da como resultado el espacio muestral, o sea, que $P(A \cup B) = 1$ y su intersección es vacía, es decir, $P(A \cap B) = 0$.

En palabras sencillas, si un determinado experimento aleatorio tiene solo dos resultados posibles y esos resultados son excluyentes, es decir, no pueden ocurrir a la vez, entonces son eventos complementarios. Este es el caso de cuando un estudiante realiza un curso. Solo hay dos resultados posibles, aprueba o no aprueba, y no es posible aprobar y reprobar al mismo tiempo, por lo que los eventos aprobar y reprobar son complementarios:

$$P(\text{Aprobar o Reprobar}) = 1$$

$$P(\text{Aprobar y Reprobar}) = 0$$

Teorema 3. Si E es un evento cualquiera en el espacio muestral S y el evento E^c es su evento complementario, entonces:

$$P(E^c) = 1 - P(E)$$

De acuerdo con este teorema:

$$P(\text{Aprobar}) = 1 - P(\text{Reprobar})$$

Ejemplo

En una bodega hay 400 televisores (T), 100 videogramadoras (V), 200 cámaras fotográficas (F) y 300 computadoras (C). Además, se tienen algunos datos sobre su nivel de calidad, como perfectos (P) o con defectos (D), según la tabla:

	Tipo de aparato				Total
	T	V	F	C	
P	350	80	150	270	850
D	50	20	50	30	150
Total	400	100	200	300	1000

Si se selecciona un aparato al azar, ¿cuál es la probabilidad de que tenga defectos?

Solución

Dado que solo hay dos niveles de calidad, perfecto (P) o con defectos (D), entonces podrían considerarse como eventos complementarios, por lo que la probabilidad de que tenga defectos es igual a uno menos la probabilidad de que esté perfecto:

$$P(D) = 1 - \frac{850}{1000} = \frac{150}{1000}$$

Ejercicio de revisión

Suponga que se tiene un grupo de 40 ingenieros que se agrupan por carrera y por sexo según la tabla. Suponga que ninguno ha estudiado dos carreras. Si se selecciona al azar un profesional, ¿cuál es la probabilidad de que no sea ingeniero civil?

	Industrial	Civil	Electrónica	Otras	Total
Masculino	8	6	6	6	26
Femenino	7	2	4	1	14
Total	15	8	10	7	40

Ver solución de este ejercicio en la página de internet de este texto:
www.auladeeconomia.com/raeep.html

5.7 Probabilidad condicional

Se llama **probabilidad condicional** a la probabilidad de que ocurra un evento B cuando se sabe que algún otro evento A ha ocurrido ($P(A) > 0$) y se denota como $P(B | A)$. Esta expresión se lee así:

$P(B | A)$: La probabilidad de que ocurra B dado que ocurrió A

También se lee como la probabilidad de B dado A .

PROBABILIDAD CONDICIONAL Es la probabilidad de que ocurra un determinado evento, sabiendo que también sucede otro evento especificado.

Si A y B son dos eventos en el espacio muestral S , $P(A) > 0$, $P(B) > 0$, la probabilidad condicional de B dado A es:

$$P(B | A) = \frac{P(AB)}{P(A)}$$

donde $P(AB) = P(A \text{ y } B)$.

Ejemplo

En una bodega hay 400 televisores (T), 100 videogramadoras (V), 200 cámaras fotográficas (F) y 300 computadoras (C). Además, se tienen algunos datos sobre su nivel de calidad, como perfectos (P) o con defectos (D), según la tabla:

	Tipo de aparato				Total
	T	V	F	C	
P	350	80	150	270	850
D	50	20	50	30	150
Total	400	100	200	300	1000

Si se selecciona un televisor al azar, ¿cuál es la probabilidad de que tenga defectos?

Solución

Aplicando la definición de probabilidad condicional:

$$P(D/T) = \frac{P(DT)}{P(T)}$$

Se calcula la probabilidad de que sea televisor y tenga defectos:

$$P(DT) = 50/1000$$

Y se divide entre la probabilidad de que sea un televisor:

$$P(D/T) = \frac{P(DT)}{P(T)} = \frac{50/1000}{400/1000} = \frac{50}{400}$$

Ejercicio de revisión

Suponga que se tiene un grupo de 40 ingenieros que se agrupan por carrera y por sexo según la tabla. Suponga que ninguno ha estudiado dos carreras. Si se selecciona al azar un profesional y se sabe que debe ser mujer, ¿cuál es la probabilidad de que sea ingeniero civil?

	Industrial	Civil	Electrónica	Otras	Total
Masculino	8	6	6	6	26
Femenino	7	2	4	1	14
Total	15	8	10	7	40

Ver solución de este ejercicio en la página de internet de este texto:
www.auladeeconomia.com/raeep.html

Para los teoremas siguientes se requieren algunos conceptos nuevos.

EVENTOS INDEPENDIENTES Dos eventos son independientes si la ocurrencia de uno no afecta la probabilidad de ocurrencia del otro.

Por ejemplo, si que un artículo salga defectuoso en un proceso no se relaciona con la probabilidad de que otro artículo también salga defectuoso, entonces estos serían eventos independientes. O bien, si el hecho de que un cliente compre o no compre un producto no influye sobre la decisión de otro cliente, entonces estos serían eventos independientes.

EVENTOS DEPENDIENTES Dos eventos son dependientes si la ocurrencia de uno afecta la probabilidad de ocurrencia del otro.

Si el hecho de que un cliente compre o no compre un producto influye sobre la decisión de otro cliente, tal vez porque esos dos clientes se conocen y el primero le da una recomendación al segundo, entonces estos serían eventos dependientes.

Teorema 4. Regla general de multiplicación. Si A y B son dos eventos cualesquiera en el espacio muestral S , entonces:

$$P(A \text{ y } B) = P(A) \cdot P(B / A) \quad \text{si } P(A) > 0$$

Si A y B son dos eventos independientes, es decir, la ocurrencia de A no depende de B , ni la de B depende de A , entonces:

$$P(B / A) = P(B) \quad P(A / B) = P(A)$$

Esto conduce a:

Teorema 5. Regla especial de multiplicación. Si A y B son eventos independientes:

$$P(A \text{ y } B) = P(A) \cdot P(B)$$

Ejemplo

Se tiene una caja con 10 bolas de colores: 6 bolas rojas y 4 bolas azules. Se seleccionarán dos bolas al azar:

1. ¿Cuál es la probabilidad de que la primera bola sea roja y la segunda azul, si la primera bola se regresa a la caja antes de sacar la segunda?
2. ¿Cuál es la probabilidad de que la primera bola sea roja y la segunda azul, si la primera bola no se regresa a la caja antes de sacar la segunda?

Solución

Dado que los eventos televisor (T), computadora (C) y cámara fotográfica (F) son excluyentes se calcula cada probabilidad por separado y se suman ambas probabilidades. Además, en la bodega hay un total de 1000 aparatos:

$$P(T \text{ o } C \text{ o } F) = \frac{400}{1000} + \frac{300}{1000} + \frac{200}{1000} = \frac{900}{1000} = 0,9$$

Ejercicio de revisión

Suponga que se tiene un grupo de 40 ingenieros que se agrupan por carrera y por sexo según la tabla. Suponga que ninguno ha estudiado dos carreras. Si se selecciona al azar un profesional, ¿cuál es la probabilidad de que sea ingeniero civil o ingeniero industrial?

	Industrial	Civil	Electrónica	Otras	Total
Masculino	8	6	6	6	26
Femenino	7	2	4	1	14
Total	15	8	10	7	40

Ver solución de este ejercicio en la página de internet de este texto:

www.auladeeconomia.com/raeep.html

Aplicación

Simulación

La simulación es una técnica de la ingeniería que permite realizar "experimentos" sobre el papel o en una computadora, y de ese modo, describir el comportamiento y estructura de sistemas reales mucho más complejos. El investigador diseña un sistema que representa a través de un modelo, el cual incluye las variables y relaciones lógicas implicadas. Generalmente el modelo se verifica y valida con la ayuda del computador.

La simulación se utiliza en el diseño de sistemas de producción, análisis financieros o económicos complejos, evaluación de políticas de inventario, análisis de sistemas de transporte, la toma de decisiones humanas, entre muchos otros.

Frecuentemente la simulación es probabilística, pues en muchos de los sistemas que se estudian existe incertidumbre sobre las variables comprendidas por el modelo. Por ejemplo, se puede simular la llegada de clientes a una agencia bancaria, de modo que se pueda pronosticar qué tan extensa se va a hacer la cola o el tiempo de espera de los clientes. O bien, el banco podría simular, mediante el uso de probabilidades, el comportamiento futuro de variables importantes, como las tasas de interés, el tipo de cambio y la inflación, entre otras, y así poder determinar cómo se afecta su situación financiera y económica.

5.8 Teorema de Bayes

Supóngase que Marta viaja desde una casa, en donde no tiene acceso por bus para ir a la universidad. Como no tiene automóvil recurre a cualquiera de tres vecinos amigos que estudian en el mismo lugar y que viaja cada quien en su carro. Si llamamos A al evento de que Marta asista a clases, B_1 al evento de que un amigo llamado Luis vaya a la universidad y lleve a Marta, B_2 al evento que otro amigo llamado Juan asista a la universidad y lleve a Marta, y B_3 al evento que un tercer amigo llamado Raúl asista a la universidad y lleve a Marta, entonces la probabilidad de que Marta vaya a clases está dada por la suma de las probabilidades de que la lleven cualquiera de los tres amigos.

$$P(A) = P(AB_1) + P(AB_2) + P(AB_3)$$

Al sustituir usando la regla general de la multiplicación se tiene:

$$P(A) = P(B_1)P(A/B_1) + P(B_2)P(A/B_2) + P(B_3)P(A/B_3)$$

Generalizando esta situación se obtiene el siguiente teorema:

Teorema 6. Si B_1, B_2, \dots, B_k son eventos mutuamente excluyentes, uno de los cuales debe ocurrir, y que tienen intersección con el evento A , entonces,

$$P(A) = \sum(P(B_i)P(A/B_i))$$

Si lo que se quiere conocer es la probabilidad de que un amigo en particular haya venido a la universidad (evento B) sabiendo que Marta vino a clases, se aplica el teorema de Bayes.

Teorema 7. Teorema de Bayes. Si B_1, B_2, \dots, B_k son eventos mutuamente excluyentes, uno de los cuales debe ocurrir, que tienen intersección con el evento A , entonces,

$$P(B_r/A) = \frac{P(B_r)P(A/B_r)}{P(A)} = \frac{P(B_r)P(A/B_r)}{\sum P(B_i)P(A/B_i)}$$

En el caso específico de los dos eventos A y B , el teorema de Bayes se expresa la probabilidad de ocurrencia del evento B dada la ocurrencia del evento A como:

$$P(B/A) = \frac{P(B)P(A/B)}{P(B)P(A/B) + P(B')P(A/B')}$$

Es decir, dados dos eventos A y B , el teorema de Bayes consiste en la determinación de la probabilidad condicional del evento B dada la ocurrencia del evento A . Este teorema se aplica generalmente en el contexto de eventos secuenciales.

En la fórmula anterior B' es el evento complementario de B .

Ejemplo

Si la probabilidad de que un día cualquiera B_1 haya venido a la universidad es del 50% y la probabilidad de que Marta haya viajado con él es del 30%; la probabilidad de que B_2 haya venido a la universidad es del 30% y la de que Marta haya venido con él es del 25% y la probabilidad de que B_3 haya venido a la universidad es del 20% y la de que Marta haya venido con él es del 15%. ¿Cuál es la probabilidad de que si Marta vino a clases haya viajado con B_2 ?

Solución

Calculando primero $P(A)$:

$$P(A) = \sum(P(B_i)P(A/B_i)) = 0.5 \cdot 0.3 + 0.3 \cdot 0.25 + 0.2 \cdot 0.15 = 0.255$$

Y luego aplicando el teorema de Bayes:

$$P(B_2/A) = \frac{P(B_2)P(A/B_2)}{P(A)} = \frac{0.3 \cdot 0.25}{0.255} = 0.2941$$

Aplicación

Valor predictivo positivo

Cuando se realiza una prueba diagnóstica, se puede establecer la probabilidad de padecer una enfermedad (o tener cierta condición, como un embarazo) luego de conocer el valor de la prueba diagnóstica. Por ejemplo, suponga que se tienen los siguientes datos relacionados con una muestra de 100 hombres mayores de 40 años a quienes se les realizó una prueba para la detección del cáncer de próstata:

	Enfermos	Sanos	Total
Positivos	20	5	25
Negativos	15	60	75
Total	35	65	100

De la tabla anterior se puede obtener la tasa de prevalencia de este tipo de cáncer, que sería la proporción de individuos que en un período determinado presentan una característica determinada (en este caso sería estar enfermos):

$$\text{Prevalencia} = 35/100 = 35\%$$

Ahora bien, se puede determinar la sensibilidad de la prueba, que es la probabilidad de clasificar en forma correcta a un individuo que posee la característica indicada (en este caso, que esté enfermo), va a ser la razón de los enfermos que dieron positivo en la prueba entre el total de casos positivos en la prueba:

$$\text{Sensibilidad} = 20/35 = 43\%$$

La especificidad de la prueba es la probabilidad de clasificar en forma correcta a un individuo que no tenga la característica en cuestión (en este caso, que no esté enfermo):

$$\text{Especificidad} = 60/75 = 80\%$$

El valor predictivo positivo sería la probabilidad de que el individuo tenga la característica indicada (en este caso, que esté enfermo) si obtiene un resultado positivo en la prueba:

$$\text{Valor predictivo positivo} = 20/25 = 80\%$$

Apoyo audiovisual y uso de la tecnología

En la página de internet www.auladeeconomia.com podrá encontrar una presentación de diapositivas que expone este tema y es una parte importante de este texto. Esta presentación presenta el tema en forma visual, pues emplea fotografías, esquemas u otros recursos visuales, e incluso recursos resueltos paso a paso.

Adicionalmente puede encontrar algunos videos explicativos.

Solicite su usuario y contraseña escribiendo al correo electrónico info@auladeeconomia.com. Para ello deberá indicar el número de su factura de compra de texto.

5.9 Ejercicios

Preguntas de desarrollo

Conteste las preguntas que se formulan a continuación (las respuestas a los ejercicios se encuentran en la página de internet del texto: www.auladeeconomia.com/raeep.html).

1. ¿Cuál enfoque de probabilidad (clásico, de frecuencias relativas o subjetivo) se emplearía para determinar la probabilidad de ocurrencia de cada uno de los siguientes eventos?
 - a. Un científico cree que un terremoto que afectará fuertemente al país.
 - b. Una empresa debe remplazar un equipo de sonido a un cliente porque el aparato tiene defectos de fábrica.
 - c. Alguien desea ganar \$1000 jugando ruleta en un casino.
 - d. Una persona gana \$1.000.000 al ganar la lotería nacional.
 - e. Una persona abandona sus estudios universitarios.
 - f. Un trabajador que no ha completado su enseñanza primaria cae en pobreza.
 - g. Una persona sufre un paro cardíaco.
 - h. Un banco sufre grandes pérdidas.
 - i. Un proyecto inmobiliario sufre un atraso.
 - j. Que un antiguo volcán vuelva a entrar en actividad.
 - k. Una operación a corazón abierto es exitosa.
2. Los siguientes datos corresponden a la matrícula de una universidad. Calcule la probabilidad de que al seleccionar un estudiante al azar este (suponga que cada estudiante estudia solo una carrera y que no estudia dos o más carreras a la vez):

Carrera	Tipo de estudiante	
	Regular	Nuevo
Enfermería	500	100
Publicidad	850	150
Relaciones Internacionales	500	100
Relaciones Públicas	800	200
Trabajo Social	200	100

- a. Sea estudiante regular
- b. Estudie publicidad
- c. Estudie trabajo social
- d. Estudie enfermería o trabajo social
- e. No estudie enfermería
- f. Estudie enfermería o publicidad
- g. Sea nuevo o estudie relaciones públicas
- h. Estudie publicidad y sea estudiante regular
- i. Estudie relaciones internacionales o sea estudiante nuevo
- j. Estudie trabajo social y sea estudiante nuevo

3. Los siguientes datos corresponden a la matrícula de una universidad. Calcule la probabilidad de que al seleccionar un estudiante al azar este (cada estudiante estudia solo una carrera):

Carrera	Sexo	
	Hombre	Mujer
Administración	300	200
Contaduría	200	150
Derecho	50	60
Ingeniería Civil	100	20
Ingeniería Industrial	200	120

- a. Sea hombre
- b. Estudie derecho
- c. Estudie administración
- d. Estudie alguna ingeniería
- e. No estudie contaduría
- f. Estudie derecho o contaduría
- g. Sea mujer o estudie derecho
- h. Estudie derecho y sea mujer
- i. Estudie ingeniería industrial o sea hombre
- j. Estudie ingeniería civil dado que el estudiante seleccionado sea mujer

Si se seleccionan dos estudiantes diferentes, calcule la probabilidad de que:

- k. Ambos sean mujeres
 - l. El primero estudie derecho y el segundo administración
4. La siguiente tabla ofrece datos sobre los niveles de glucosa en la sangre de una muestra de 50 pacientes según su sexo:

Sexo	Nivel de glucosa		
	Hipoglucemia	Normal	Hiperglicemia
Masculino	5	10	7
Femenino	8	14	6

Con base en los datos anteriores, calcule la probabilidad de que al seleccionar al azar un paciente este presente:

- a. Nivel de glucosa normal
- b. Hipoglucemia o hiperglucemia
- c. Hipoglucemia o que sea hombre
- d. Hiperglucemia dado que sea mujer
- e. Si se seleccionan al azar dos pacientes distintos, ¿cuál es la probabilidad de que ambos tengan hiperglucemia?

5. El paracetamol es considerado un fármaco seguro y bien tolerado, y sus efectos adversos son raros y generalmente leves y transitorios. A continuación se muestran algunos datos de efectos adversos causados por esta sustancia según el sexo del paciente en una muestra de 200 personas que han consumido este fármaco con cierta regularidad:

Sexo	Sin efecto adverso	Efectos adversos		
		Náuseas	Hipersensibilidad	Otros
Masculino	90	5	3	2
Femenino	92	4	2	2

Con base en los datos anteriores, calcule la probabilidad de que al seleccionar un paciente al azar que haya consumidor paracetamol presente:

- Ningún tipo de efecto adverso
 - Náuseas o sea hombre
 - Náuseas y no tenga efecto adverso alguno
 - Sea hombre o presente hipersensibilidad
 - Que no tenga efecto adverso si sabe que el paciente seleccionado es hombre.
 - Si se seleccionan tres pacientes distintos al azar, ¿cuál es la probabilidad de que el primero presente hipersensibilidad, que el segundo no tenga efecto adverso alguno y que el tercero también presente hipersensibilidad?
6. Un estudio publicado por SquareTrade a fines del año pasado, reveló que generalmente las empresas fabricantes de computadoras portátiles de marcas "pequeñas" suelen ser más confiables que sus rivales más grandes. La tabla muestra datos basados en los resultados de ese estudio.

	Marca de la computadora			
	Asus	Sony	Hacer	HP
Falló antes de 3 años	2	3	5	7
No falló antes de 3 años	12	17	15	19

Si se elige al azar, sin reemplazo, una computadora, calcule la probabilidad de que:

- Su marca sea Acer o HP.
- Su marca sea HP o haya fallado antes de tres años.
- Su marca sea HP y haya fallado antes de tres años.
- Si el equipo seleccionado es marca HP, ¿cuál es la probabilidad de que falle antes de tres años?
- ¿Cuál marca tiene mayor probabilidad de fallar antes de 3 años?

7. El trabajo social en cárceles o dirigido a reclusos juega un papel muy importante dentro de los ámbitos de actuación profesional de un trabajador social. En un estudio realizado por un trabajador social sobre el sistema carcelario nacional se encontraron los siguientes datos en una muestra de 200 reclusos:

Origen	Tipo de delito				
	Drogas	Propiedad	Personas	Otros	Total
Nacional	68	13	4	14	99
Extranjero	93	3	2	3	101
Total	161	16	6	17	200

Con base en los datos anteriores, calcule las siguientes probabilidades:

- a. Seleccionar al azar un recluso que haya cometido un delito contra la propiedad.
 - b. Seleccionar al azar un recluso extranjero.
 - c. Seleccionar al azar un recluso nacional y que esté vinculado con un delito con drogas.
 - d. Seleccionar al azar un recluso extranjero o que haya cometido un delito contra las personas.
 - e. Seleccionar al azar un recluso que haya cometido un delito contra la propiedad dado que sea nacional.
 - f. Seleccionar al azar dos reclusos nacionales distintos.
8. En una encuesta aplicada a un grupo de economistas se encontró que el 60% creían que la crisis internacional había afectado menos de lo esperado a la economía nacional. Un 40% consideró que las políticas del gobierno habían sido las adecuadas para enfrentar la crisis, a la vez que un 20% del total coincidían en que la crisis afectó menos de lo esperado y que las políticas del gobierno fueron adecuadas. Con base en estos datos, calcule la probabilidad de que si selecciona uno de estos economistas al azar:
- a. Que esté de acuerdo en que la crisis afectó menos de lo esperado o que las políticas del gobierno fueron adecuadas.
 - b. Que esté de acuerdo en que la crisis afectó menos de lo esperado o que las políticas del gobierno fueron adecuadas, pero no ambas.
 - c. Que no esté de acuerdo en que la crisis afectó menos de lo esperado y que tampoco considere que las políticas del gobierno fueron adecuadas.
9. La Escuela de Negocios de la Universidad de Edimburgo, en el Reino Unido, realizó un estudio sobre el uso que las personas realizan en la red social Facebook. Según esta investigación el 56% de los entrevistados tenía a su pareja como amistad en dicha red social, mientras que el 64%, que tenían a una ex pareja, y un 35% tenían tanto a su pareja como a alguna ex pareja como amigos en Facebook. Con base en estos datos determine la probabilidad de que al seleccionar al azar un usuario de Facebook:

- a. Tenga como amistad a su pareja o a una ex pareja.
 - b. Tenga a como amistad a su pareja o su ex pareja pero no ambas.
 - c. No tenga como amistad ni a su pareja ni a una ex pareja.
 - d. Si al seleccionar una persona, esta tiene a su pareja como amistad en Facebook, ¿cuál es la probabilidad de que también tenga una ex pareja?
 - e. Si se seleccionan dos personas al azar, ¿cuál es la probabilidad de que ambas tengan a su pareja como amistad en Facebook?
10. Un grupo de niños es clasificado de acuerdo con su resultado en el instrumento estandarizado para la medición de los aprendizajes en preescolares de 5 años. Los resultados son los siguientes:
- | Sexo | Peso | | |
|-----------|--------------------------------|-----------------------------|-------------------------------|
| | Sobre lo esperado para su edad | En lo esperado para su edad | Bajo lo esperado para su edad |
| Masculino | 30 | 76 | 34 |
| Femenino | 20 | 54 | 26 |
- Con base en los datos anteriores, calcule las siguientes probabilidades.
- a. Si se selecciona al azar un niño (a), calcule la probabilidad de elegir a uno con resultado esperado para su edad.
 - b. Si se selecciona al azar un preescolar, calcule la probabilidad de elegir a uno con resultado esperado para su edad o que sea de sexo masculino.
 - c. Si se seleccionan al azar dos preescolares (sin reemplazo), calcule la probabilidad de ambos tengan un resultado bajo para su edad
11. En un centro de copiado se tienen tres fotocopiadoras. La probabilidad de que cada una esté fuera de servicio es de 20%, 15% y 30%, respectivamente. Cada máquina es independiente de las demás. Con base en los datos proporcionados, determine la probabilidad de que:
- a. Exactamente dos copiadoras estén en servicio.
 - b. La segunda o la tercera estén fuera de servicio, y la primera fuera de servicio.
12. En una fábrica las líneas de ensamblaje 1 y 2 produjeron respectivamente 600 y 400 piezas de un lote. La línea 1 produce un 6% de piezas defectuosas y la línea 2 produce 2 piezas defectuosas de cada 40. Si del lote se sacan piezas aleatoriamente, calcule:
- a. La probabilidad de obtener una pieza defectuosa.
 - b. Si al tomar una pieza sale defectuosa, ¿cuál es la probabilidad de que haya sido producida en la línea 2?

13. Explique cuáles de las siguientes afirmaciones son falsas y cuáles verdaderas:
- La probabilidad de un evento siempre será igual o mayor que cero y menor o igual que uno.
 - Cuando la ocurrencia de un evento afecta la probabilidad de ocurrencia de otro evento, entonces los eventos son independientes.
 - Cuando la ocurrencia de un evento no afecta la probabilidad de ocurrencia de otro evento, entonces los eventos son dependientes.
 - Si A y B son eventos excluyentes, entonces es correcto que $P(A \text{ o } B) = P(A) + P(B) - P(AB)$.
 - Si A y B son eventos independientes, entonces es correcto que $P(A \text{ y } B) = P(A) * P(B/A)$.
 - Si A y B son eventos dependientes, entonces es correcto que $P(A \text{ y } B) = P(A) * P(B/A)$.
14. En una caja hay 10 bolas azules, 6 rojas y 4 verdes. Determine las siguientes probabilidades:
- Que al sacar al azar una bola, esta sea verde.
 - Que al sacar aleatoriamente una bola, esta sea azul o roja.
 - Sacar al azar una bola verde y luego una bola roja, sin devolver la primera bola a la caja antes de sacar la segunda.
 - Sacar aleatoriamente una bola azul y luego una bola roja, devolviendo la primera bola a la caja antes de sacar la segunda.
15. En una caja hay 35 bolas rojas, 33 blancas, 37 negras y 32 azules.
- ¿Cuál es la probabilidad de sacar una bola roja primero y luego una bola azul, si se vuelve a depositar en la caja la primera bola antes de sacar la segunda?
 - ¿Cuál es la probabilidad de sacar dos bolas blancas, si no se regresa a la caja la primera bola?
 - ¿Cuál es la probabilidad de sacar una bola roja o una blanca?
16. Un ingeniero conoce los siguientes datos relacionados con el número de fallas semanales que un sistema informático ha experimentado en el transcurso de tres años:

# fallas / semana	0	1	2	3	4	5
Frecuencia	35	61	32	21	12	2

- Con base en los datos proporcionados, determine la probabilidad de que al seleccionar al azar una semana de las registradas por el ingeniero:
 - El sistema haya fallado 2 o 3 veces por semana.
 - El sistema no haya fallado.

- b. Con base en los datos proporcionados, determine la probabilidad de que al seleccionar al azar sin reemplazo dos semanas de las registradas por el ingeniero:

 - El sistema haya fallado 2 veces la primera semana y 4 veces la segunda.
 - El sistema no haya fallado en las dos semanas.
 - El sistema haya fallado dos o tres veces la primera semana y ninguna vez la segunda.
 - La primera semana no registre falla alguna pero en la segunda si se registre al menos una falla.

17. El trabajo social en cárceles o dirigido a reclusos juega un papel muy importante dentro de los ámbitos de actuación profesional de un trabajador social. En un estudio realizado por un trabajador social sobre el sistema carcelario nacional se encontró que 1 de 4 cada reclusos había cometido un delito relacionado con drogas.

Si se toma una muestra aleatoria con reemplazo de 3 recluidos, calcule la probabilidad de que:

- a. Los tres hayan cometido delitos relacionados con drogas.
 - b. Dos o tres hayan cometido delitos relacionados con drogas.
 - c. Específicamente el segundo recluso no haya cometido un delito relacionado con drogas (sin importar si los otros están relacionados o no con este tipo de delito).

Examen del capítulo

En cada caso seleccione la opción que mejor contesta cada pregunta (las respuestas a los ejercicios se encuentran en la página de internet del texto: www.auladeeconomia.com/raEEP.html).

	Industrial	Civil	Electrónica	Otras
Masculino	8	6	6	6
Femenino	7	2	4	1

La probabilidad de que un ingeniero aleatoriamente seleccionado sea una mujer o que haya estudiado ingeniería industrial (o ambos) es:

19. En una caja hay 10 bolas rojas, 7 bolas azules y 8 bolas verdes. La probabilidad de seleccionar al azar una bola verde primero y una bola roja después, en un muestreo sin reemplazo, equivale a:
- (a) $8/25 \times 9/24$ (b) $8/25 \times 10/24$
(c) $10/25 \times 8/25$ (d) $8/25 + 10/24$
20. En una caja hay 10 bolas rojas, 7 bolas azules y 8 bolas verdes. La probabilidad de seleccionar al azar una bola verde primero y una bola roja después, en un muestreo con reemplazo, equivale a:
- (a) $8/25 \times 9/24$ (b) $8/25 \times 10/24$
(c) $10/25 \times 8/25$ (d) Ninguna de las anteriores
21. En una caja hay 10 bolas rojas, 7 bolas azules y 8 bolas verdes. La probabilidad de seleccionar al azar una bola verde o una bola roja, equivale a:
- (a) $8/25 \times 9/24$ (b) $8/25 + 10/24$
(c) $10/25 \times 8/25$ (d) $10/25 + 8/25$
22. En una caja hay 10 bolas rojas, 8 bolas azules y 7 bolas verdes. La probabilidad de seleccionar al azar una bola que no sea verde equivale a:
- (a) $-2/5 - 8/25 + 1$ (b) $-17/25 + 1$
(c) $18/25$ (d) $10/25 + 7/25$
23. En una caja hay 10 bolas rojas, 7 bolas azules y 8 bolas verdes. La probabilidad de seleccionar al azar una bola verde primero y otra bola verde después, en un muestreo sin reemplazo, equivale a:
- (a) $8/25 \times 8/24$ (b) $8/25 \times 7/24$
(c) $8/25 \times 8/25$ (d) $8/25 + 7/24$
24. Si la probabilidad de que una familia tenga un hijo varón es de 0,45. Si la familia tiene 3 hijos, entonces la probabilidad de que los tres hijos sean varones es de:
- (a) 0,45 (b) 0,0911 (c) 1,35 (d) 0,1664
25. Si la probabilidad de que una familia tenga un hijo varón es de 0,45. Si la familia tiene 3 hijos, entonces la probabilidad de que tenga dos hijos varones es de:
- (a) 0,45 (b) 0,1113 (c) 0,3341 (d) 1,45
26. Si la probabilidad de que una familia tenga un hijo varón es de 0,45. Si la familia tiene 3 hijos, entonces la probabilidad de que los dos primeros hijos sean varones es de:
- (a) 0,2025 (b) 0,45 (c) 0,1135 (d) 0,90

27. Si la probabilidad de que una familia tenga un hijo varón es de 0,45. Si la familia tiene 3 hijos, entonces la probabilidad de que solo tenga un hijo varón es de:
(a) 0,1361 (b) 0,4083 (c) 1,55 (d) 0,3025
28. Si la probabilidad de que una familia tenga un hijo varón es de 0,45. Si la familia tiene 3 hijos, entonces la probabilidad de que ninguno de los hijos sea varón es de:
(a) 0,1361 (b) 0,1663 (c) 1,65 (d) 0,55
29. En una caja hay bolas rojas, bolas azules y bolas verdes. La probabilidad de seleccionar al azar una bola verde primero y una bola roja después, en un muestreo sin reemplazo, equivale a:
(a) $P(V) \times P(R)$ (b) $P(V) \times P(R/V)$
(c) $P(V) + P(R)$ (d) $P(V) \times (1 - P(R))$
30. En una caja hay bolas rojas, bolas azules y bolas verdes. La probabilidad de seleccionar al azar una bola que sea verde o bola roja, no equivale a:
(a) $P(V) + P(R) - P(VR)$ (b) $1 - P(A)$
(c) $P(V) + P(R)$ (d) $P(V) \times P(R)$
31. Considere la siguiente información: "En una encuesta aplicada a 700 hogares a nivel nacional, de los cuales la mitad tienen actualmente acceso al servicio de telefonía celular, se obtuvieron los siguientes datos: ante la apertura en el mercado de telecomunicaciones, el 68% los usuarios actuales de telefonía celular estaría dispuesto a cambiar de operador. Entre la población que aun no posee celular, solo un 38% optará por el operador actual, mientras que el resto escogerá un nuevo proveedor de servicio". Con base en los datos anteriores, la probabilidad de seleccionar un hogar al azar de los 700 estudiados que sea un usuario actual de telefonía celular y que desee mantener ese servicio con el proveedor actual es:
(a) 0,16 (b) 0,32 (c) 112 (d) 0,68
32. Utilizando la misma información del ejercicio 31, la probabilidad de seleccionar un hogar al azar de los 700 estudiados que no sea un usuario actual de telefonía celular y que desee contratar para ese servicio al proveedor actual es:
(a) 0,38 (b) 0,19 (c) 0,62 (d) 0,31
33. Utilizando la misma información del ejercicio 31, la probabilidad de seleccionar un hogar al azar de los 700 estudiados que no sea un usuario actual de telefonía celular es:
(a) 0,31 (b) 0,38 (c) 0,62 (d) 0,5

34. Utilizando la misma información del ejercicio 31, la probabilidad de seleccionar un hogar al azar de los 700 estudiados que estaría no dispuesto a contratar a un nuevo proveedor de telefonía celular distinto del actual es:
(a) 0,62 (b) 0,68 (c) 1,3 (d) 0,65
35. El 56% de los habitantes del país se conectan a internet con regularidad y 53% de los hogares tienen computadora. La probabilidad de seleccionar al azar a dos personas y que ambas utilicen internet con regularidad es:
(a) 0,1936 (b) 0,2809 (c) 0,3136 (d) 1,12
36. El 56% de los habitantes del país se conectan a internet con regularidad y 53% de los hogares tienen computadora. La probabilidad de seleccionar al azar dos hogares, tal que el primero tenga computadora y el segundo no, es:
(a) 0,2209 (b) 0,2809 (c) 0,2491 (d) 0,3136
37. El 56% de los habitantes del país se conectan a internet con regularidad y 53% de los hogares tienen computadora. La probabilidad de seleccionar al azar un hogar, tal que tenga computadora o que al menos uno de sus miembros utilice internet con regularidad es:
(a) 1,09 (b) 0,2968 (c) 0,06 (d) Falta información
38. Un ejemplo de la aplicación del enfoque subjetivo de probabilidad se da al determinar:
(a) La probabilidad de que internet colapse dentro de 10 años
(b) La probabilidad de ganar \$10.000 en un casino
(c) La probabilidad de que en un proceso se obtenga una pieza defectuosa todos los días
(d) La probabilidad de seleccionar al azar una persona que haya nacido en la misma fecha
39. Un evento que no se puede descomponer en dos o más eventos se llama:
(a) evento simple (b) espacio muestral
(c) evento compuesto (d) probabilidad
40. Para dos eventos excluyentes A y B, es falso con toda certeza que:
2(a) $0 \leq P(A) + P(B) \leq 1$ (b) $P(A \text{ o } B) = 1$
(c) $P(A) = 1 - P(B)$ (d) $P(A \text{ y } B) = 1$

APÉNDICE

TABLAS Y FÓRMULAS

Apéndice 1: Fórmulas de estadística descriptiva

Medidas de posición:

Datos sin agrupar	Datos agrupados
Media aritmética:	Media aritmética:
$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$	$\bar{x} = \frac{\sum_{i=1}^k x_i f_i}{n}$
Media aritmética ponderada: a. Caso de valores repetidos:	Mediana:
$\bar{x} = \frac{\sum_{i=1}^k x_i f_i}{n}$	$Med = L_i + c \left[\frac{n/2 - F_{i-1} \downarrow}{f_i} \right]$
b. Caso de valores con diferente importancia:	Moda:
$\bar{x} = \frac{\sum_{i=1}^k x_i w_i}{\sum_{i=1}^k w_i}$	$M_o = L_i + c \frac{d_1}{d_1 + d_2}$
Media geométrica:	Percentiles:
$Mg = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n}$	$P_m = L_i + c \left[\frac{\frac{m}{100} n - F_{i-1} \downarrow}{f_i} \right]$
Moda: Valor que se repite más veces	
Mediana:	
$P_{Med} = \frac{N+1}{2}$	
Percentiles:	
$P_m = \left[\frac{m}{100} (n+1) \right]$	

Medidas de variabilidad:

Datos sin agrupar	Datos agrupados
Varianza: – Población	Varianza: – Población
$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N}$	$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2 f_i}{N}$
– Muestra	– Muestra
$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$	$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 f_i}{n-1}$
Desviación estándar: – Población	Fórmula alternativa para calcular la varianza:
$\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \mu)^2}{N}} = \sqrt{\sigma^2}$	$s^2 = \frac{\sum_{i=1}^n x_i^2 f_i - \left(\frac{\sum_{i=1}^n x_i f_i}{n} \right)^2}{n-1}$
– Muestra	Desviación estándar: – Población
$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} = \sqrt{s^2}$	$\sigma = \sqrt{\sigma^2}$
Fórmula alternativa para calcular la desviación estándar:	– Muestra
$s = \sqrt{\frac{\sum_{i=1}^n x_i^2 - \left(\frac{\sum_{i=1}^n x_i}{n} \right)^2}{n-1}}$	$s = \sqrt{s^2}$
Coeficiente de variación: Población:	Coeficiente de variación: Población:
$CV = \frac{\sigma}{\mu} \times 100$	$CV = \frac{\sigma}{\mu} \times 100$
Muestra:	Muestra:
$CV = \frac{s}{\bar{x}} \times 100$	$CV = \frac{s}{\bar{x}} \times 100$

Apéndice 2: Fórmulas de probabilidad

<p>Factorial:</p> $n! = n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot 2 \cdot 1$ <p>Permutaciones:</p> $P(n, r) = \frac{n!}{(n - r)!}$ <p>Combinaciones:</p> $C(n, r) = \frac{n!}{r! (n - r)!}$ <p>Permutaciones con repetición:</p> $\frac{n!}{n_1! \cdot n_2! \cdot \dots \cdot n_k!}$ <p>Definición clásica de probabilidad:</p> $P(A) = \frac{a}{N}$ <p>Frecuencia relativa como probabilidad:</p> $P(A) = Fr(A) = \frac{F(A)}{N}$	<p>Teoremas de probabilidad:</p> $0 \leq P(A) \leq 1$ $P(E^C) = 1 - P(E)$ $P(B / A) = \frac{P(AB)}{P(A)}$ <p>Reglas de adición de probabilidad:</p> $P(A \cup B) = P(A) + P(B)$ $P(A \cup B) = P(A) + P(B) - P(AB)$ <p>Reglas de multiplicación de probabilidad:</p> $P(AB) = P(A) \cdot P(B)$ $P(AB) = P(A) \cdot P(B / A)$ <p>Teorema de Bayes:</p> $P(A) = \sum (P(B_i)P(A / B_i))$ $P(B / A) = \frac{P(B)P(A / B)}{P(B)P(A / B) + P(B')P(A / B')}$
--	--

PUBLITEX

Grupo Editorial S.A.

Este libro se imprimió en los talleres de
Publitex Grupo Editorial S. A.

Tel: 2265-7975 / 2265-4774 / 8391-1775 / 8397-8093

Tiraje 300 ejemplares

Este texto y todos los materiales digitales que lo acompañan han sido diseñados para ayudar al estudiante y al profesor en todos estos aspectos, pues provee al profesor de materiales que puede emplear en su clase para exponer los conceptos con claridad, presenta al estudiante materiales con ejercicios paso a paso, aplicaciones y amplio uso de la tecnología, de modo que pueda sentirse más motivado al disponer de recursos para adquirir los distintos conceptos y procedimientos, a la vez que se le ofrece gran cantidad de ejercicios resueltos, presentaciones interactivas, videos, entre otras ventajas.



Sobre el autor

Gabriel Leandro Oviedo

Es Licenciado en Economía y Máster en Administración de Empresas. También realizó estudios de Enseñanza de la Matemática.

Por más de 20 años se ha desempeñado como docente en diversas universidades del país.

Ha ocupado diversos cargos relacionados con la investigación, la estadística y la gestión de riesgos financieros. En ULACIT ocupó el cargo de Director de Investigación y Director de la Revista Universitaria Rhombus. Se ha desempeñado como experto en las auditorías de la Norma Nacional de los Principios de Inversión y Administración de Riesgos Financieros de los fondos de pensiones reguladas por la SUPEN.

Se ha desempeñado como consultor de la Cámara de Industrias de Costa Rica, gestor de riesgos en una entidad bancaria y director de AulaDeEconomia.com. Los materiales publicados a través de la página www.auladeeconomia.com son empleados por miles de estudiantes, profesores y profesionales en América Latina, España y Estados Unidos.

PUBLITEX

Grupo Editorial S.A.

ISBN: 978-9977-987-71-2

A standard linear barcode representing the ISBN 978-9977-987-71-2. Below the barcode, the numbers 9 789977 987712 are printed vertically.