Bloque 4

Unidad 5

Estadística

BECU_M1_B4_P144_173.indd 144 4/22/14 2:02 PM

Antes de empezar

1. Reduce las siguientes operaciones.

a. $\frac{4}{5} + \frac{12}{4} + \frac{7}{3}$

b. 25% 1 75% 1 31%

c. $\frac{6 \cdot 2 + 25 \cdot 8 + 13 \cdot 3 + 31 \cdot 13 + 44 \cdot 7}{30}$

2. Ordena los siguientes números en forma creciente.

14, 12, 32, 5, 24, 72, 7, 36, 23, 17, 10, 9, 30

3. Halla los siguientes valores.

a. El 5% de 20 c. El 25% de 89

b. El 80% de 47 d. El 50% de 550

4. Simplifica hasta su mínima expresión cada fracción y, luego, expresar cada fracción simplificada como número decimal.

a. $\frac{88}{36}$ c. $\frac{33}{11}$ e. $\frac{12}{6}$

b. $\frac{25}{100}$ d. $\frac{104}{120}$

5. Halla las siguientes potencias.

a. 2^5 c. 2^4 e. 1^{10}

b. 5^2 d. 4^2

Status, estado, estadística

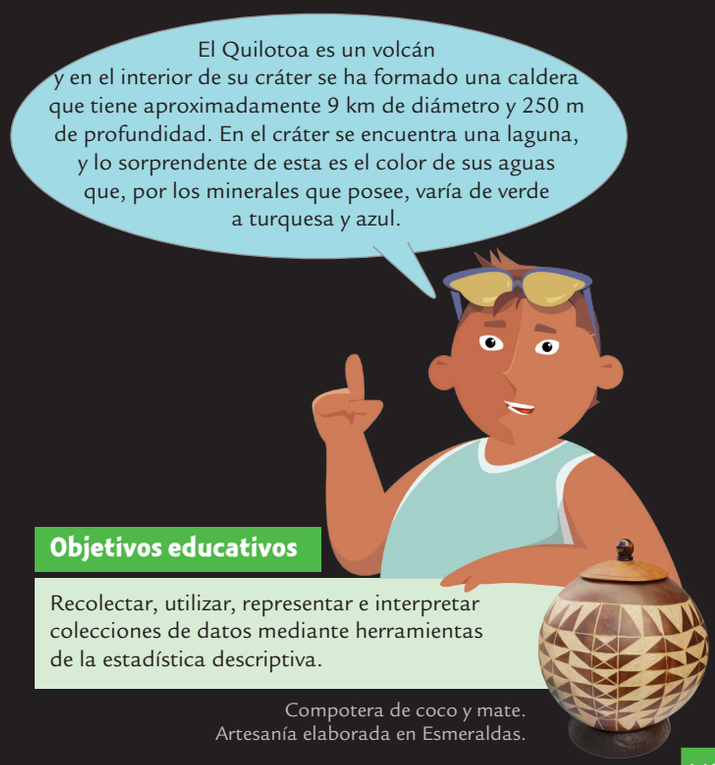
El origen de la estadística se remonta a los comienzos de la Historia. En los antiguos monumentos egipcios se han encontrado documentos según los cuales, a partir del año 3050 a. de C., se llevaban cuentas de los movimientos de población y continuamente hacían censos, bajo la dirección del faraón.

En la Biblia, por ejemplo, en el libro de los Números, se encuentra registrado el censo que realizó Moisés después de la salida de Egipto.

En China, Confucio en uno de sus clásicos Shu-King escrito hacia el año 550 a. de C., narra cómo el rey Yao, ordenó hacer una estadística agrícola, industrial y comercial.

Con Carlo Magno, en Francia, regresaron las estadísticas a Europa, teniendo un carácter netamente financiero y administrativo. En Inglaterra Guillermo el Conquistador mandó a realizar una especie de catastro, que constituyó un documento estadístico administrativo.

A mediados del siglo XVII, gracias a Vito Seckendorff y, sobre todo, a German Conring la estadística se empezó a considerar como la descripción de los hechos notables de un estado. Conring perfeccionó y mejoró notablemente la nueva tendencia, sistematizando los conocimientos y los datos. El mejor de sus seguidores fue Godofredo Achenwall, quien consolidó definitivamente los postulados de esta nueva ciencia y le dio el nombre de *estadística*, palabra que etimológicamente se deriva de la palabra *status*, que significa 'estado' o 'situación'.



Objetivos educativos

Recolectar, utilizar, representar e interpretar colecciones de datos mediante herramientas de la estadística descriptiva.

Compotera de coco y mate.
Artesanía elaborada en Esmeraldas.

BECU_M1_B4_P144_173.indd 145 4/22/14 11:58 AM

Estadística descriptiva

Destrezas con criterio de desempeño:

- Identificar **conceptos básicos** que se utilizan en la estadística descriptiva. **(C)**
- Comprender situaciones de la vida cotidiana a través de la interpretación de **datos estadísticos**. **(M)**

La estadística consiste en un conjunto de técnicas y procedimientos que permiten recoger datos, presentarlos, ordenarlos y analizarlos, de manera que, a partir de ellos, se puedan inferir conclusiones.

Conocimientos previos

Indica a qué tipos de variables estadísticas corresponden los siguientes ejemplos.

- Estatura
- Peso
- Número de hijos
- Marcas de automóviles

POBLACIÓN Y MUESTRA

La **población** es un conjunto de objetos o de individuos que se desea estudiar y que, a su vez, presentan una característica que interesa medir.

Generalmente, el tamaño de la población se denota con la letra **N**.

Se llama **muestra** a un subconjunto representativo de la población que se desea estudiar. Generalmente, el tamaño de la muestra se denota con la letra **n**.

VARIABLES ESTADÍSTICAS

Una **variable** estadística corresponde a una o varias características que se miden en la muestra. Las variables pueden ser **cuantitativas** o **cualitativas**.

Variables cualitativas

Son aquellas que no se pueden medir numéricamente y están relacionadas con características. Los valores que toma este tipo de variables son etiquetas que representan categorías o cualidades.

Una variable cualitativa puede ser **nominal** u **ordinal**.

Las **variables nominales** corresponden a aquellas en las cuales no existe ninguna ordenación; por ejemplo, el estado civil, el sexo de un individuo, etc.

Las **variables ordinales** son aquellas en las cuales existe un orden intuitivo; por ejemplo, nivel educacional (básico, medio, superior), situación económica (baja, media, alta), etc.

Variables cuantitativas

Son aquellas que se pueden medir numéricamente, es decir, los valores que toma este tipo de variables son números.

Una variable cuantitativa puede ser **discreta** o **continua**.

Las **variables discretas** son aquellas en las cuales los posibles valores surgen frecuentemente de un conteo. En cada tramo o intervalo, la variable solo puede tomar un número determinado de valores (enteros).

Glosario

estadística. Derivada del latín *status*, que significa 'estado' o 'situación'.

Toma en cuenta

Una muestra se toma cuando la población es muy grande y no es posible realizar un estudio con todos los integrantes de esta. Es importante que la muestra se escoja correctamente, pues de lo contrario, las conclusiones obtenidas no serán representativas de la población.

Ejemplos

1. Número de hijos.
2. Número de páginas de un libro, etc.

BECU_M1_B4_P144_173.indd 146 4/22/14 11:58 AM

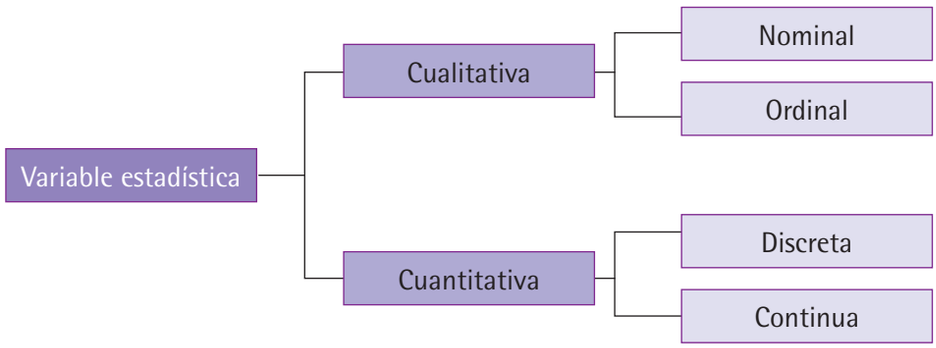
Trabajo individual

Escribe dos ejemplos de cada uno de los tipos de variables estadísticas.

Las **variables continuas** son aquellas en las cuales los posibles valores surgen frecuentemente de una medición. Estas variables pueden tomar tantos valores (reales) como sea posible en un tramo.

Ejemplos

1. La estatura de una persona.
2. El peso de alguien, etc.



Recuerda

Estadística e historia

En el año 2000 a. C. en China ya se realizaban estudios estadísticos relacionados con el censo de la población.

Por otro lado, los romanos, cada cinco años, realizaban un recuento de la población, que consideraba cantidad de nacimientos, defunciones, ganado, etc.

ESTUDIO ESTADÍSTICO

Para realizar un estudio estadístico, generalmente se siguen los siguientes pasos.

1º Recolección, orden y recuento de datos.

2º Cálculo de las medidas de centralización y localización.

3º Representación gráfica de los resultados.

4º Planteamiento de las conclusiones.

Actividades

Determina elementos de un estudio estadístico.

1. En la unidad educativa, durante el primer quimestre del año se propondrá realizar algunos estudios, como:

- ¿Qué características comunes tienen los estudiantes dedicados a la música?
- ¿En qué usan el tiempo libre los estudiantes de básica?
- ¿Qué oportunidades de esparcimiento y diversión tienen en la ciudad?
- ¿Cuáles son los hábitos deportivos de los profesores?
- ¿A qué dedican el tiempo libre que comparten las familias de la unidad educativa?

Determina:

a. Entre cinco y diez variables que se puedan considerar en el estudio.

b. Valores de cada variable considerada.

Clasifica las variables estadísticas.

2. En un estudio sobre el uso de la televisión por parte de los estudiantes de primer año de bachillerato del colegio Avancemos, se averiguará:

a. Edad

b. Género

c. Número de horas semanales dedicadas a ver televisión.

d. Tipo de programas preferidos.

e. Cantidad de canales de los que disponen en casa.

Determina las variables e indica de qué clase son.

BECU_M1_B4_P144_173.indd 147 4/22/14 11:58 AM

Tablas de frecuencias

Destrezas con criterio de desempeño:

- Reconocer y elaborar cuadros de **frecuencias absolutas y frecuencias acumuladas**, con datos simples y con datos agrupados. **(C, P)**
- Comprender situaciones de la vida cotidiana a través de la interpretación de **datos estadísticos**. **(M)**

Conocimientos previos

Determina los siguientes porcentajes.

a. 30% de 300

b. 75% de 250

c. 40% de 1 200

Toma en cuenta

De la tabla de frecuencias presentada en el ejemplo, se pueden responder preguntas tales como las siguientes:

- ¿Cuántas veces se obtuvo 4?
- ¿Cuántos resultados menores o iguales que 4 se obtuvo?
- ¿Qué porcentaje obtuvo el resultado 5?

Observa

notación. Para denotar la frecuencia absoluta se utiliza la expresión $\mathbf{f_i}$. Para denotar la frecuencia relativa se utiliza la expresión $\mathbf{h_i}$.

TABLAS DE FRECUENCIA PARA DATOS NO AGRUPADOS

Al ordenar los datos correspondientes a un cierto estudio, es usual agruparlos en clases o categorías, para lo cual, generalmente, se utilizan tablas de frecuencias.

Frecuencia absoluta

Es el número de veces que aparece o se repite un cierto valor en la variable de medición.

Frecuencia absoluta acumulada

Representa el número de datos cuyo valor es menor o igual al valor considerado. Se obtiene sumando sucesivamente las frecuencias absolutas.

Frecuencia relativa

Representa la razón de ocurrencia respecto al total. Se calcula como el cociente entre la frecuencia absoluta y el tamaño total de la muestra.

La suma de todas las frecuencias relativas da como resultado 1.

Frecuencia relativa porcentual

Corresponde a la frecuencia relativa expresada en porcentaje. Se calcula como el producto de la frecuencia relativa por 100.

La suma de todas las frecuencias relativas porcentuales da como resultado 100%.

Ejemplo

Al lanzar un dado 10 veces, se obtienen los siguientes resultados: 1, 1, 3, 4, 4, 4, 5, 5, 6, 6. Se puede observar que la variable de estudio es el resultado del lanzamiento del dado. Luego, la tabla de frecuencias correspondiente es:

Resultado	F. absoluta f_i	F. acumulada F_i	F. relativa h_i	F. relativa %
1	2	2	2/10	20%
2	0	2	0/10	0%
3	1	3	1/10	10%
4	3	6	3/10	30%
5	2	8	2/10	20%
6	2	10	2/10	20%

TABLAS DE FRECUENCIA PARA DATOS AGRUPADOS

Si el conjunto de datos que se recolecta es muy numeroso, o bien, si el **rango** (diferencia entre el mayor y menor valor de una variable) es muy amplio, es usual presentarlos agrupados y ordenados en **intervalos** (rango de valores).

BECU_M1_B4_P144_173.indd 148 4/22/14 11:58 AM

Toma en cuenta

Que los intervalos tengan igual tamaño facilita la interpretación, análisis y conclusión de los estudios.

Sin embargo, en ocasiones los intervalos no coinciden en tamaño.

Amplitud de una clase. Se calcula mediante la diferencia entre el límite superior y el límite inferior de un intervalo.

Recuerda

De la tabla de frecuencias dada en el ejemplo se pueden responder preguntas tales como las siguientes:

- ¿Qué porcentaje de pacientes presentan un nivel de colesterol menor a 200 mg/d?
- ¿Qué porcentaje de pacientes supera los 200 mg/d de colesterol total?
- ¿Cuántos pacientes tienen una medición entre 190 mg/d y 199 mg/d?

Tamaño de un intervalo

El tamaño de cada intervalo se puede calcular dividiendo el valor del rango para la cantidad de intervalos que se desean obtener.

Marca de clase

Es un valor representativo de cada intervalo (o clase), que corresponde al punto medio del intervalo. Se calcula como la suma del límite inferior (menor valor) y el límite superior (mayor valor) del intervalo, dividido entre 2.

Ejemplo

Un grupo de 20 pacientes entre 50 y 60 años se realizaron un examen para medir su nivel de colesterol (en mg/d). Los resultados obtenidos fueron los siguientes.

183	206	193	172 ← valor menor
182	205	199	177
180	195	199	175
185	197	190	201
200	210	219	219 ← valor mayor

Como se puede observar, los valores de la variable de estudio (nivel de colesterol) presentan un rango amplio. Los datos se agruparon en 5 intervalos de tamaño 9, ya que $\frac{219 - 172}{5} = \frac{47}{5} \approx 9$, es decir, cada intervalo es de amplitud 9. Luego, la tabla de frecuencias correspondiente es:

Nivel de colesterol	F. absoluta	F. acumulada	F. relativa	F. relativa %
170 – 179	3	3	3/10	15%
180 – 189	4	7	4/10	20%
190 – 199	6	13	6/10	30%
200 – 209	4	17	4/10	20%
210 – 219	3	20	3/10	15%

Actividades

Analiza problemas de estadística.

1. Los siguientes datos se obtuvieron durante una encuesta a los estudiantes de primer año de bachillerato sobre el número de años que están en el colegio.

4	5	2	9	4	5	7
2	8	2	5	1	9	6
10	2	5	5	5	4	5
3	2	10	3	10	8	3
3	2	6	2	10	3	
7	5	5	5	9	3	

a. Elabora la tabla de frecuencias e indica a cuántos se ha encuestado.

b. Indica cuál es la variable.

c. Encuentra el porcentaje de los que tienen más tiempo en el colegio.

d. Halla el porcentaje de los que tienen menos tiempo.

e. Responde ¿Cuántos estudiantes tienen menos de 5 años en el colegio?

BECU_M1_B4_P144_173.indd 149 4/22/14 11:58 AM

Gráfico de frecuencias

Destrezas con criterio de desempeño:

- Reconocer en diferentes **diagramas estadísticos** (tallo y hojas, polígonos de frecuencia, gráfico de barras, caja y bigotes, histogramas, etc.) la información que estos proporcionan. **(C)**
- Interpretar un **diagrama estadístico** a través de los parámetros representados en él. **(C)**
- Representar los resultados de cuadros de frecuencias absolutas y frecuencias acumuladas mediante los **diferentes diagramas** (tallo y hojas, polígonos de frecuencia, gráfico de barras, histogramas, etc.). **(P)**

Conocimientos previos

Elabora una tabla y encuentra las frecuencias absolutas y relativas.

Las edades en un grupo de jóvenes son:

14	15	14
16	15	14
15	15	14
16	13	14
15	13	13
14	15	16
16	14	

HISTOGRAMA

Es una representación gráfica de una distribución de frecuencias, generalmente de variables cuantitativas agrupadas en intervalos. Está formado por barras cuyas bases representan el intervalo al que corresponden los valores de la variable, y las alturas están dadas por las frecuencias de cada categoría.

GRÁFICO CIRCULAR

Este gráfico, también conocido como **diagrama de sectores**, se utiliza para representar cualquier tipo de frecuencias aunque, generalmente, se emplea para frecuencias relativas porcentuales.

Los datos son representados mediante sectores de un círculo. Cada sector indica diferentes categorías de la variable y cada ángulo de los sectores circulares es proporcional al valor de la variable.

Ángulos de los sectores de un gráfico circular

La medida de los ángulos de los sectores circulares, se obtiene multiplicando las frecuencias absolutas de la categoría por 360° y dividiendo para el número total de datos, es decir:

$$\text{Ángulo} = \frac{f_i \cdot 360^o}{N}$$

Donde f_i: frecuencia absoluta y N: total de datos.

POLÍGONO DE FRECUENCIAS

Un polígono de frecuencias se obtiene al unir los puntos medios de los intervalos representados por cada barra en un histograma, es decir, al unir la marca de clase de cada intervalo mediante una línea poligonal.

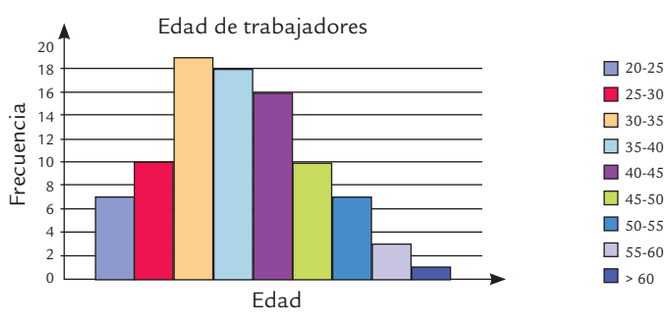
Ejemplo

Se encuestó al personal de una empresa con la finalidad de conocer y registrar las edades de sus trabajadores.

La tabla de frecuencias con los resultados obtenidos es la siguiente:

Edades	[20 – 25[	[25 – 30[	[30 – 35[	[35 – 40[	[40 – 45[	[45 – 50[	[50 – 55[	[55 – 60[	> 60
f_i	7	10	19	18	16	10	7	3	1

- El histograma correspondiente es:



Glosario

ejes. En un histograma los intervalos se representan en el eje de las abscisas (x), mientras que en el eje de las ordenadas (y) se representan las frecuencias.

BECU_M1_B4_P144_173.indd 150 4/22/14 11:58 AM

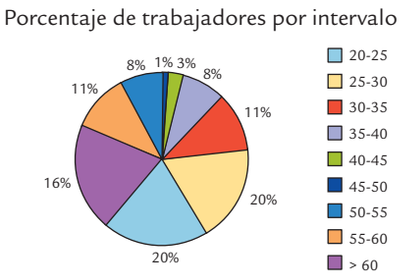
Toma en cuenta

Gráfico según variable

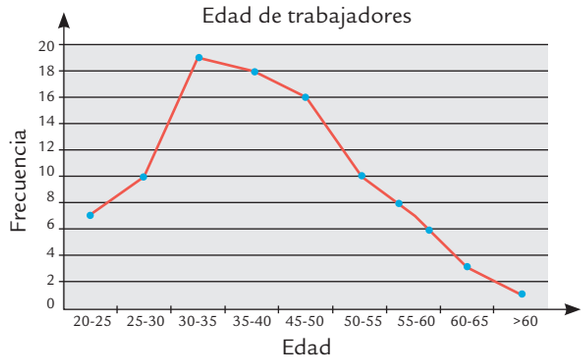
La elección del gráfico más adecuado para representar cierto tipo de variable dependerá de si esta es cualitativa o cuantitativa. Generalmente se utilizan los siguientes tipos de gráficos.

Variable	Nombre
Cualitativa	· Circular · Gráfico de barras
Cuantitativa discreta	· Circular · Polígono de frecuencias
Cuantitativa continua	· Histograma · Polígono de frecuencias

- El gráfico circular que muestra las frecuencias relativas correspondientes a cada categoría es:



- El polígono de frecuencias correspondiente al ejemplo es:



Un polígono de frecuencias puede ser construido a partir de las frecuencias absolutas, o bien, de las frecuencias relativas. Aunque estos valores sean distintos, la forma del gráfico es similar, ya que las proporciones entre ellas se mantienen.

Actividades

Elabora una tabla de funciones.

1. Los datos corresponden a la duración en horas del uso continuo de 40 dispositivos electrónicos iguales, sometidos a un control de calidad.

480	570	775	826	890
496	802	712	560	590
724	795	683	794	750
780	886	830	676	489
801	714	560	760	725
830	452	810	720	680
660	490	895	660	666
746	668	880	570	680

Construye una tabla de distribución de frecuencias agrupadas que considere las columnas *intervalo*, *frecuencia absoluta* y *frecuencia relativa*.

Representa la información en una diagrama de barras.

2. Grafica un diagrama de barras con la siguiente información.
Las estaturas en centímetros de 27 jóvenes son las siguientes: 155, 178, 170, 165, 173, 168, 160, 166, 176, 169, 158, 170, 179, 161, 164, 156, 170, 171, 167, 151, 163, 158, 164, 174, 176, 164 y 154.

Traza un histograma.

3. Realiza el histograma y el polígono de frecuencias con la siguiente información.
Se mide la altura de 200 estudiantes de Bachillerato y se extraen los resultados de la siguiente tabla.

Altura (cm)	[140-150[	[150-160[	[160-170[	[170-180[	[180-190[
fi	12	58	91	46	23

BECU_M1_B4_P144_173.indd 151 4/22/14 11:58 AM

Observa

En el ejemplo de gráfico de frecuencias acumuladas, en el eje *x* se ha representado el límite superior del intervalo y en el eje *y*, la frecuencia acumulada.

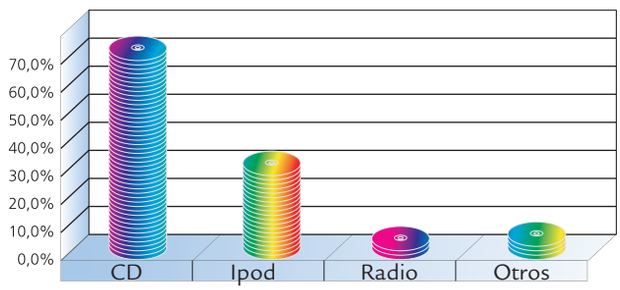
PICTOGRAMA

Este tipo de gráfico se utiliza para representar variables cualitativas. Para cada valor de la variable, se utiliza una figura cuyo tamaño es proporcional a la frecuencia.

Ejemplo

El siguiente pictograma corresponde a una encuesta de consumo cultural y tiempo libre.

Medios o formatos empleados para escuchar música (año 2013)



Toma en cuenta

Es posible construir un gráfico a partir de la frecuencia relativa acumulada.

GRÁFICO DE FRECUENCIAS ACUMULADAS (OJIVA)

Recuerda

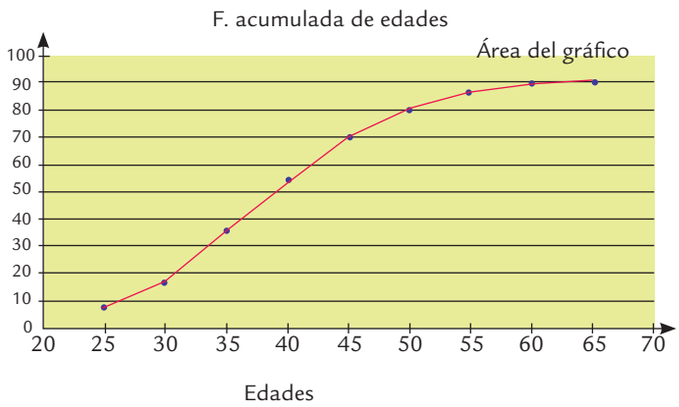
En el gráfico de frecuencias acumuladas se marcan los puntos de la forma (x, y), donde:
x: límite superior
y: frecuencia acumulada.

En un gráfico de distribución de frecuencias acumuladas, se puede notar que esta frecuencia de un intervalo corresponde a todas las observaciones menores que el límite superior de ese intervalo.

Ejemplo

Si se considera el caso de las edades del personal de una empresa, se obtiene la siguiente tabla de frecuencias acumuladas, en la cual se ha incluido el límite superior de cada intervalo y el respectivo gráfico.

Edades	Límite superior	F. absoluta	F. acumulada
[20 – 25[	25	7	7
[25 – 30[	30	10	17
[30 – 35[	35	19	36
[35 – 40[	40	18	54
[40 – 45[	45	16	70
[45 – 50[	50	10	80
[50 – 55[	55	7	87
[55 – 60[	60	3	90
[60 – 65[	65	1	91



BECU_M1_B4_P144_173.indd 152 4/22/14 11:58 AM

Toma en cuenta

John Wilder Tukey (1915-2000)

Inventó el diagrama de tallo y hoja, como un modo rápido de representar un conjunto de datos de manera gráfica.

Recuerda

En general, se utilizan diagramas de tallo y hoja para estudiar la dispersión de los valores de una muestra.

Además, la elección del tallo y hojas es arbitraria pues estos se seleccionan de acuerdo a la conveniencia y dependerán de la magnitud de los datos.

El diagrama correspondiente a los datos: 20, 21, 35, 40, 43 y 45 se puede representar como:

2	0 1
3	5
4	0 3 5

DIAGRAMA DE TALLO Y HOJA

Este diagrama tiene por objetivo resumir u ordenar un conjunto de datos, con el fin de conocer intuitivamente la forma de su distribución. También permite comparar la distribución de dos o más grupos diferentes.

Este tipo de gráfico se construye separando los valores de cada observación en dos partes, la primera corresponde al tallo y se ubica a la izquierda de una línea vertical; la segunda incluye a las hojas y se ubica a la derecha.

Si se tienen muchas hojas en cada tallo, es posible separarlas en dos tallos.

Ejemplo

A continuación se muestra un diagrama de tallo y hoja, correspondiente al salario, en dólares, de un grupo de trabajadores de una empresa.

Tallo	Hojas			
4	00	25		
5	31	42		
6	28	36	99	
7	18	69	87	
8	17	25	33	95
9	25	58	77	96
10	53	76		
11	01	69		
12	00	79		
13	33	70	95	
14	59	61		
15	89			
16	12	51		
17	16			
18	65			
19	30			
20				
21				
22	00			

Del diagrama anterior se puede desprender información como la siguiente.

- El trabajador que tiene menor salario gana $ 400.
- El trabajador que tiene mayor salario percibe $ 2 200.
- La mayoría de los trabajadores percibe entre $ 400 y $ 1 600, ya que entre esos valores se presenta la mayor frecuencia de salarios.

BECU_M1_B4_P144_173.indd 153 4/22/14 11:58 AM

Actividades

Identifica variables cualitativas y cuantitativas.

1. Indica en cada caso si la variable dada es cualitativa o cuantitativa.

a. Número de personas que integran un grupo familiar en una parroquia de Quito.

b. Sueldo de los empleados de una empresa.

c. Color de ojos de los estudiantes de un curso.

d. Nivel de escolaridad de los integrantes de un grupo familiar.

Diferencia entre variables discretas y continuas.

2. Señala cuáles de las siguientes son variables discretas y cuáles son continuas.

a. Temperaturas diarias medidas en una ciudad.

b. Ingresos de los ejecutivos de un banco.

c. Longitudes de 100 clavos producidos por una empresa.

d. Número de estudiantes en una sala de clases.

Identifica a que clase pertenecen las variables.

3. Clasifica cada una de las siguientes variables, según su clase: cuantitativa (discreta o continua) o cualitativa (nominal u ordinal).

a. Distancia recorrida por un automóvil desde Latacunga a Ambato.

b. Tiempo que se requiere para responder una encuesta.

c. Cantidad de llamadas que llegan a una central telefónica.

d. Periódicos vendidos en un kiosco.

e. Color de pelo de los integrantes de una familia.

f. Número de acciones transadas en la bolsa en un día determinado.

Determina las características de un problema estadístico.

4. Se realizó una encuesta para conocer la preferencia del jefe(a) de hogar de un barrio, por algún tipo de supermercado. Entre los 100 jefes(as) de hogar entrevistados, 30 prefirieron el supermercado «Triunfo».

a. ¿Cuál es la muestra de esta encuesta?

b. ¿Cuál es el tamaño de la muestra?

c. ¿Cuál es la población?

d. ¿Cuál es la variable de estudio?

e. ¿Cuál es la proporción, dentro de la muestra, de jefes (as) de hogar que prefirieron el supermercado «Triunfo»?

f. ¿Cuántos jefes(as) de hogar prefirieron otro supermercado?

Completa tablas estadísticas y analiza su información.

5. En una guardería y prebásica, las edades de los niños se han representado en la siguiente tabla. Observa, y luego, responde.

Edad (años)	Números de niños
1	6
2	8
3	7
4	3

a. ¿Cuántos niños mayores de 3 años hay en el jardín?

b. ¿Qué porcentaje de niños tienen un año de edad?

c. ¿Cuántos niños tienen 2 años o menos?

d. ¿Cuántos niños tienen más de 1 año?

e. ¿Qué porcentaje de niños tiene más de 2 años?

f. ¿Cuántos niños hay en total en la institución?

BECU_M1_B4_P144_173.indd 154 4/22/14 11:58 AM

6. Los siguientes datos corresponden a la duración, en horas, del uso continuo de 50 ampolletas iguales, que serán sometidas a un control de calidad. Observa y luego responde.

480	496	724	780	801
570	802	795	886	714
775	712	683	830	560
826	560	794	676	760
890	590	750	489	725
666	746	668	880	570
830	452	810	720	680
680	660	490	895	660
720	793	870	715	708
710	679	762	793	751

a. ¿Cuál es la población de estudio?

b. ¿Cuál es la muestra del estudio?

c. ¿Cuál es la variable de estudio?

d. ¿Qué porcentaje de ampolletas tiene una duración de al menos 590 horas?

e. Grafica una tabla de distribución de frecuencias, considerando la frecuencia absoluta, la frecuencia relativa, la relativa porcentual y las respectivas frecuencias acumuladas.

f. ¿Qué porcentaje de ampolletas tiene una duración de 660 horas o más?

g. Grafica en Excel un histograma para representar el tiempo de duración de las ampolletas.

7. Las siguientes cantidades son las tarifas, en dólares, que una empresa de mensajería cobró por entregar paquetes pequeños la tarde del jueves pasado.

4,03	3,07	4,02	5,57	4,63	3,82	4,30
4,07	6,04	3,62	3,89	5,02	5,46	5,24
3,59	7,86	2,93	3,70	4,15	3,10	4,91

a. Elabora un diagrama de tallo y hojas.

b. Construye una distribución de frecuencias.

8. De los partidos de fútbol jugados en un mes en cierto campeonato, se contabilizó el número de goles marcados en cada partido. Observa y, luego responde.

Número de goles	Frecuencia
0	3
1	5
2	7
3	7
4	6
5	4
6	3

a. ¿Cuál es la población de estudio?

b. ¿Cuál es la muestra del estudio?

c. ¿Cuál es el tamaño de la muestra?

d. ¿Cuál es la variable de estudio?

e. Elabora una tabla de distribución de frecuencias considerando la frecuencia absoluta, la frecuencia relativa, la relativa porcentual y las respectivas frecuencias acumuladas.

f. Construye, utilizando Excel, un gráfico circular a partir de los datos de la tabla.

g. Obtén al menos tres conclusiones a partir del gráfico.

9. A continuación, se proporcionan las edades de 50 bailarines que asistieron a una audición para participar en una comedia musical.

21	22	18	23	19	19	21	21	22	21
21	21	19	20	19	20	19	21	19	21
18	19	22	24	24	19	19	20	19	20
20	22	20	20	20	19	21	19	19	19
20	21	22	19	19	21	18	21	20	17

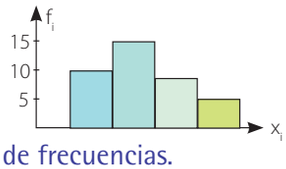
a. Elabora una distribución de frecuencias donde los intervalos sean:

17 años -18 años
19 años
20 años
21 años
Más de 21 años

BECU_M1_B4_P144_173.indd 155 4/22/14 11:58 AM

Actividades

10. Elige la respuesta correcta. El siguiente diagrama es un:



a. polígono de frecuencias.

b. histograma.

c. diagrama de barras.

d. polígono de frecuencias acumuladas.

11. Se realizó un estudio de seguridad en 34 ciudades de un país. Este arrojó los siguientes resultados a la pregunta: «En qué lugar se siente más seguro».

	Casa	Lugar de trabajo	Lugares públicos	Calle
Muy seguro	52%	41%	41%	13%
Muy inseguro	47%	30%	55%	86%
No responde	1%	29%	4%	1%

a. Construye un gráfico circular para cada uno de los lugares. ¿Qué puedes concluir?

b. Elabora un histograma que muestre las diferencias entre los cuatro sitios. ¿A qué crees que se deba esta divergencia?

c. Responde: Si la muestra de la encuesta anterior fuera de 402 personas, ¿cuántas corresponderían a cada categoría?

12. La tabla siguiente indica la edad de los 40 socios de un club.

Edad	15	16	17	18	19
Nº de personas	5	8	2	20	5

Haz el histograma o el diagrama de barras correspondiente.

13. En una encuesta, realizada a los estudiantes, con la pregunta «¿Debería haber más actividades extraescolares?», las respuestas fueron: De acuerdo, 55,8%; En desacuerdo, 17,2%; Indiferente, 27%. Señala cómo representarías estas respuestas.

a. Con un gráfico de sectores

b. Con un polígono de frecuencias

c. Con un histograma de frecuencias

d. Con un diagrama de barras

14. En una inspección oftalmológica a los empleados de una empresa, un oculista anotó las dioptrías del ojo derecho en esta tabla.

Dioptrías	0	0,5	1	1,5	2
f_i	42	30		24	
h_i	0,35		0,15		0,05

a. Completa el cuadro.

b. Grafica el diagrama de sectores.

c. Dibuja el polígono de frecuencias relativas.

15. Se ha lanzado una moneda 40 veces y se han obtenido los siguientes resultados: x, x, c, x, c, c, c, c, x, x, c, x, x, x, c, x, c, x, x, x, x, c, c, x, x, c, x, c, x, x, c, c, x, x, x, x, c, x.

a. Obtén la tabla de frecuencias.

b. Responde: ¿Qué gráfico es más adecuado para representar los datos?

16. La siguiente información corresponde al número de pacientes atendidos en una sala de urgencias el pasado mes de febrero.

Domingo	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado
						① 15
② 21	③ 10	④ 8	⑤ 17	⑥ 16	⑦ 9	⑧ 8
⑨ 21	⑩ 9	⑪ 25	⑫ 32	⑬ 16	⑭ 28	⑮ 17
⑯ 20	⑰ 12	⑱ 8	⑲ 15	⑳ 12	㉑ 32	㉒ 15
㉓ 12	㉔ 14	㉕ 18	㉖ 29	㉗ 22	㉘ 10	

a. Construye un gráfico de serie de tiempo.

b. ¿Cuáles son las conclusiones que se pueden obtener de estos datos?

BECU_M1_B4_P144_173.indd 156 4/22/14 11:58 AM

17. Para una investigación se midió la estatura, en centímetros, de 35 mujeres y 35 hombres. Los datos son los siguientes.

Hombres		
172	159	171
174	157	155
166	167	174
171	179	158
156	152	157
167	153	178
169	159	170
152	174	154
160	159	153
154	168	163
169	159	180
171	155	

Mujeres		
146	136	148
138	140	144
153	142	164
162	136	159
162	144	165
164	145	143
135	152	164
147	146	137
161	146	159
142	162	164
139	146	159
136	149	

a. Elabora un diagrama de tallo y hojas para cada muestra.

b. Compara los diagramas anteriores y presenta una conclusión con respecto a los datos.

c. Construye una distribución de frecuencias para las estaturas de los hombres y una distribución de frecuencias para las estaturas de las mujeres.

18. El siguiente diagrama de tallo y hojas corresponde al peso, en libras, de 34 estudiantes de un colegio.

Diagrama de tallo y hojas para el pesos de 34 estudiantes

13	2 5 7
14	2 3 5 8
15	0 4 4 5 7 8
16	1 2 2 5 7 8
17	0 0 6 6 7
18	3 4 6 8
19	0 1 5 5
20	5
21	5

Tallo: decenas y centenas

Hojas: unidades

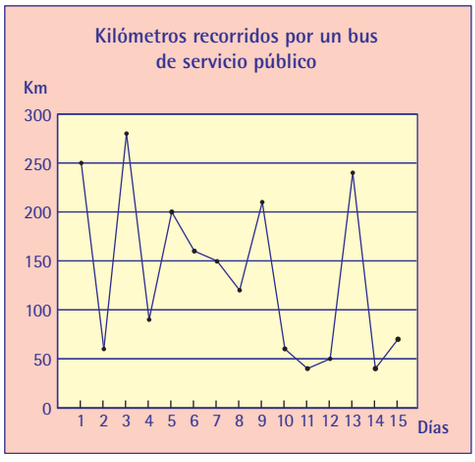
a. Construye una distribución de frecuencias con los datos recogidos sobre el peso de los 34 estudiantes.

b. Determina, según tu criterio, los porcentajes más significativos.

19. La siguiente información corresponde a la cantidad de kilómetros recorridos por un bus de servicio público durante los últimos 15 días.

a. A partir de la gráfica, elabora un diagrama de tallo y hojas del kilometraje recorrido por el bus.

b. Responde: ¿Cuántos días el recorrido del bus está entre 100 km y 200 km?



BECU_M1_B4_P144_173.indd 157 4/22/14 11:58 AM

Medidas de tendencia central

Destreza con criterio de desempeño:
Calcular las **medidas de tendencia central** y de dispersión para diferentes tipos de datos. **(P)**

Conocimientos previos

Responde estas preguntas:

a. ¿Cómo obtienes el promedio de tus notas?

b. ¿Qué significa estar a la moda?

*Las medidas de tendencia central son parámetros estadísticos que indican valores cuyo objetivo es resumir la información para un conjunto de datos, es decir, son representantes de un conjunto de datos. Las medidas de tendencia central más conocidas son: la **media aritmética**, la **mediana** y la **moda**.*

MEDIDAS DE TENDENCIA CENTRAL PARA DATOS NO AGRUPADOS

Media aritmética

Es el valor numérico que corresponde al cociente de la suma de todos los datos y el número total de observaciones (promedio). Se denota como $\overline{X}$.

Es decir,

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n}$$

n: número de elementos de la muestra

Toma en cuenta

- La media aritmética es única y fácil de calcular e interpretar, sin embargo, se ve afectada si existen valores extremos.
- La mediana es única y no se ve afectada si existen valores extremos.

Ejemplo

Si se considera el número de hijos de 7 familias con los siguientes resultados: 1, 2, 2, 4, 5, 5, 6, la media aritmética del número de hijos es 3,6, ya que

$$\overline{X} = \frac{1 + 2 + 2 + 4 + 5 + 5 + 6}{7} = \frac{25}{7} \approx 3{,}6.$$

Observar que la media aritmética no coincide con ningún valor de los datos, ya que estos corresponden a números enteros positivos (número de hijos), en estos casos, el valor de la media se puede aproximar, es decir $\overline{X} \approx 4$.

Mediana

Se define como el valor central de un conjunto de datos ordenados de manera creciente o decreciente. En el caso de que el número de datos sea par, la mediana corresponde a la media aritmética de los dos valores centrales. Se denota como **Me**.

Ejemplo

Las notas de 6 alumnos de un curso de botánica son las siguientes: 2,0; 3,0; 3,0; 4,0; 4,0; 5,0. El número de observaciones es par, por lo tanto, la mediana es 3,5 pues $\frac{3{,}0 + 4{,}0}{2} = 3{,}5$.

Observa

Es posible que no exista moda o que exista más de una, como ocurre en el siguiente ejemplo.

La moda de:
1-1-2-2-3-4-5-6-7 es 1 y 2.

Moda

La moda de un conjunto de observaciones corresponde a aquel dato que tiene la mayor frecuencia. Se denota como **Mo**. En el ejemplo anterior, hay dos modas: 3,0 y 4,0.

BECU_M1_B4_P144_173.indd 158 4/22/14 11:59 AM

Observa

Ejemplos para datos agrupados.

Observa la siguiente tabla de frecuencias.

Intervalo	f_i
40 – 45	12
46 – 51	15
52 – 57	4

La marca de clase para cada intervalo es 42,5; 48,5 y 54,5, respectivamente, además n = 31. Luego, la media aritmética está dada por:

$$\overline{X} = \frac{42{,}5 \cdot 12 + 48{,}5 \cdot 15 + 54{,}5 \cdot 4}{31}$$

$$\overline{X} \approx 46{,}95$$

Además, la mediana se ubica en el intervalo 46 – 51, ya que ahí se encuentra el 50% del total de los datos, por lo tanto n = 31; t = 6; $f_{i-1} = 12$; fmediana = 15; $L_i = 46$. Luego,

$$M_e = \frac{46 + \frac{31}{2} - 12}{15} \cdot 6 = 47{,}4.$$

Por otro lado, $d_1 = 15 - 12 = 3$; $d_2 = 15 - 4 = 11$. $L_i = 46$. Luego,

$$M_o = \frac{46 + 3}{3 + 11} \cdot 6 \approx 47,29$$

Las fórmulas correspondientes a la mediana y a la moda solo son válidas cuando *t* el ancho del intervalo es constante.

MEDIDAS DE TENDENCIA CENTRAL PARA DATOS AGRUPADOS

Media aritmética

La media para datos agrupados se calcula multiplicando la marca de clase de cada intervalo (x_i), con sus respectivas frecuencias absolutas (f_i). Después se suman los resultados obtenidos y este total se divide para el número total de datos (n). Este proceso está representado en la siguiente fórmula.

$$\overline{X} = \frac{\sum_{i=1}^{k} x_i f_i}{n}$$

Mediana

Una manera aproximada de calcular la mediana para datos agrupados es mediante la siguiente expresión.

$$Me = L_i + \frac{\frac{n}{2} - F_{i-1}}{f_i} \cdot a$$

Donde:

L_i: límite inferior del intervalo donde se encuentra la mediana.

n: número total de elementos de la muestra, o bien, la frecuencia total.

a: amplitud de los intervalos.

F_{i-1}: frecuencia acumulada anterior al intervalo en el cual se encuentra la mediana.

f_i: frecuencia del intervalo en el cual se encuentra la mediana.

Moda

La moda para datos agrupados está dada por la siguiente expresión.

$$Mo = L_i + \frac{d_1}{d_1 + d_2} \cdot a$$

Donde:

d_1: diferencia de la frecuencia del intervalo modal (intervalo con mayor frecuencia absoluta) y la frecuencia de la clase anterior.

d_2: diferencia de la frecuencia del intervalo modal y la frecuencia de la clase posterior.

a: ancho de los intervalos.

L_i: límite inferior de la clase modal.

Si se calcula la moda para datos agrupados, el resultado corresponde a una aproximación de esta.

BECU_M1_B4_P144_173.indd 159 4/22/14 2:10 PM

Medidas de dispersión

Destreza con criterio de desempeño:
Calcular las **medidas de tendencia central de dispersión** para diferentes tipos de datos. **(P)**

*Las medidas de dispersión son parámetros estadísticos que indican cuánto se alejan los datos respecto de la media aritmética, es decir, señalan la **variabilidad de los datos**.*

Las medidas de dispersión más utilizadas son el **rango**, la **desviación media** y la **desviación estándar** o **típica**.

Conocimientos previos

Los siguientes datos corresponden a las estaturas en centímetros de un grupo de estudiantes de primer año de bachillerato. Indica cuál es el valor de la estatura mayor, cuál es el menor valor, y qué diferencia hay entre estos dos valores.

165, 170, 168, 162, 165, 165, 169, 168, 170, 171, 167, 166, 165, 162, 172, 168, 169, 170, 164, 168

RANGO

Indica la dispersión entre los valores extremos de una variable. Se calcula como la diferencia entre el mayor y el menor valor de la variable.

Se denota como **R**, es decir,

$$R = x_n - x_1$$

Donde:
$\mathbf{x_n}$: estadístico de orden n, es decir, el mayor valor de la variable.
$\mathbf{x_1}$: estadístico de orden 1, es decir, el menor valor de la variable.

DESVIACIÓN MEDIA

- **La desviación media de una observación (d)**, con respecto a la media ($\overline{\mathbf{X}}$), se define como la diferencia entre ellas. Es decir,

$$d = x - \overline{X}$$

Toma en cuenta

La suma de las desviaciones de todos los datos con respecto a la media aritmética es siempre cero.

- **La desviación media de un conjunto de datos (DM)** es la media aritmética de los valores absolutos de las desviaciones de cada dato respecto a la media ($\overline{\mathbf{X}}$). Es decir,

$$DM = \frac{\sum_{i=1}^{n} |x_i - \overline{X}|}{n}$$

Donde:
$\mathbf{x_i}$: valores de la variable.
n: número total de datos.

Actividades

Determina el rango y la desviación media.

1. Calcula el rango, y la desviación media a partir de los datos de edad de las personas que están en una función determinada de cine, de acuerdo con la siguiente distribución.

x_i	10	15	20	25	30	35
f_i	2	12	27	41	30	7

BECU_M1_B4_P144_173.indd 160 4/22/14 11:59 AM

Recuerda

Desviación estándar para datos agrupados
Está representada con la expresión:

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

Donde $\mathbf{f_i}$ es la frecuencia de cada valor de la variable y $\mathbf{x_i}$ es la marca de clase.

DESVIACIÓN ESTÁNDAR O TÍPICA

La desviación estándar mide el grado de dispersión de los datos con respecto a la media. Se denota como **σ** para la población, o bien **s** para una muestra. Está dada por la siguiente expresión.

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \overline{X})^2}{n}}$$

Mientras menor sea la desviación estándar, los datos son más homogéneos, es decir, a menor dispersión mayor homogeneidad, y viceversa.

Ejemplo

Las notas obtenidas por un alumno en una asignatura son las siguientes: 2,0; 3,9; 5,0; 5,9; 6,2.

Obtener el rango, desviación media para cada nota y la desviación estándar de las notas obtenidas.

Se tiene que el rango es: R = 6,2 – 2,0 = 4,2, lo que indica que las notas son bastante dispersas, ya que la amplitud entre ambos valores es «grande».

Para calcular la desviación media, primero se debe calcular la media aritmética. Se tiene: $\overline{X} = \frac{22,0 + 3,9 + 5,0 + 5,9 + 6,2}{5} = 4,6$. Con esta media aritmética, se pueden obtener las desviaciones medias para cada nota.

Nota	Desviación media
2,0	–2,6
3,9	–0,7
5,0	0,4
5,9	1,3
6,2	1,6

Estos valores indican el mayor o menor alejamiento de las notas respecto a $\overline{X}$.

Para la desviación estándar, se tiene:

$$\sigma = \sqrt{\frac{(2,0 - 4,6)^2 + (3,9 - 4,6)^2 + (5,0 - 4,6)^2 + (5,9 - 4,6)^2 + (6,2 - 4,6)^2}{5}}$$

$$\sigma = \sqrt{\frac{6,76 + 0,49 + 0,16 + 1,69 + 2,56}{5}} \approx 1,53$$ — Grado de dispersión respecto a $\overline{X}$.

Luego, la desviación estándar para las notas obtenidas es aproximadamente 1,53.

Trabajo cooperativo

Hallen la desviación estándar con los datos de longitud en centímetros de truchas pescadas en el río Pita, según la siguiente distribución.

f_i	Longitud (cm)
26	[0,10[
35	[10,20[
46	[20,30[

BECU_M1_B4_P144_173.indd 161 4/22/14 11:59 AM

VARIANZA

Es otro parámetro utilizado para medir la dispersión de los valores de una variable respecto a la media. Corresponde a la media aritmética de los cuadrados de las desviaciones, respecto a la media. Está dada por la siguiente expresión.

$$\sigma = \frac{\sum_{i=1}^{n} \left(x_i - \overline{X}\right)^2}{n}$$

Dado que la varianza corresponde al cuadrado de la desviación estándar, está expresada en unidades cuadradas.

Toma en cuenta

Valor de la varianza

La varianza corresponde al cuadrado de la desviación estándar y, por lo tanto, su valor es siempre positivo.

COEFICIENTE DE VARIACIÓN

Permite determinar la razón existente entre la desviación estándar (σ) y la media. Se denota como **CV**. El coeficiente de variación se calcula mediante la siguiente expresión.

$$CV = \frac{\sigma}{\overline{X}}$$

El **CV** no tiene unidades de medida, por lo que permite la comparación de variables, sin importar sus magnitudes ni lo que estas representan.

Recuerda

Al calcular la covarianza de una variable respecto a sí misma, se obtiene la varianza.

CORRELACIÓN

El análisis de la correlación es apropiado cuando se necesita conocer el grado de asociación entre dos variables.

Covarianza

La **covarianza** (**cov**(**x, y**)) de dos variables es un indicador de la relación entre ellas. Este parámetro puede utilizarse para medir la relación entre ambas solo si están expresadas en la misma escala o unidad de medida. Se obtiene a partir de la fórmula a continuación.

$$\text{cov}(x, y) = \frac{\sum_{i=1}^{n} \left(x_i - \overline{X}\right)\left(y_i - \overline{Y}\right)}{n}$$

Coeficiente de correlación de Pearson

La correlación o grado de asociación entre dos variables se mide utilizando el **coeficiente de relación de Pearson**. Este coeficiente mide el grado de asociación lineal entre dos variables. Se denota como **r** y su valor fluctúa en el intervalo [–1, 1].

Este coeficiente se calcula mediante la siguiente expresión.

$$r = \frac{\text{cov}(x, y)}{\sigma_x \cdot \sigma_y}$$

BECU_M1_B4_P144_173.indd 162 4/22/14 11:59 AM

Toma en cuenta

Karl Pearson (1857-1936)

No solo se destacó por sus contribuciones a la estadística, sino también por sus valiosos aportes a la antropología, biometría y genética.

Donde:

σ_x: desviación estándar de la variable x.

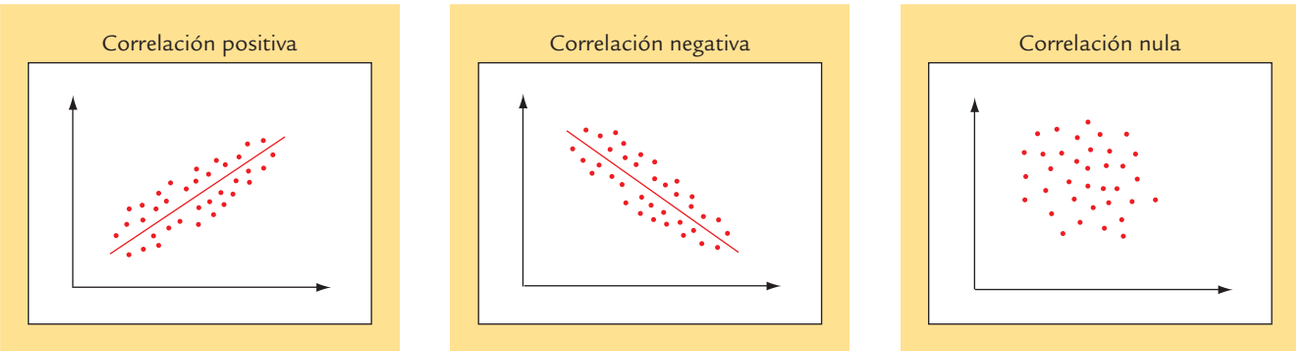
σ_y: desviación estándar de la variable y.

Análisis del coeficiente de correlación

Según el valor del coeficiente de correlación (r), se presentan estas opciones.

- Si **r** es **positivo**, la relación lineal entre las variables es directa. Se dice que la correlación es positiva.
- Si **r** es **negativo**, la relación lineal entre las variables es inversa. Se dice que la correlación es negativa.
- Si **r = 0**, no existe relación lineal entre las variables. Se dice que la correlación es nula.
- Si **r = 1**, existe una relación de dependencia total directa entre las variables, es decir, si una de ellas aumenta (o disminuye), la otra aumenta (o disminuye) en igual proporción.
- Si **r = –1**, existe una relación de dependencia total inversa entre las variables, es decir, si una de ellas aumenta (o disminuye), la otra disminuye (o aumenta) en igual proporción.

Representación gráfica de la correlación



Correlación perfecta

Si el coeficiente de correlación toma valores extremos, es decir, **r = 1** o **r = –1**, se dice que la correlación es perfecta.

Si **r = 1**, la correlación es máxima directa.

Si **r = –1**, la correlación es máxima inversa.

En ambos casos, todos los puntos pueden representarse gráficamente, en una misma recta.

BECU_M1_B4_P144_173.indd 163 4/22/14 11:59 AM

Medidas de localización

Destreza con criterio de desempeño:
Calcular las **medidas localización** para diferentes tipos de datos. **(P)**

*Una **medida de localización** nos indica el lugar donde se ubica un valor de la variable dentro de un conjunto ordenado de valores. Las medidas de localización más utilizadas son cuartiles, deciles y percentiles.*

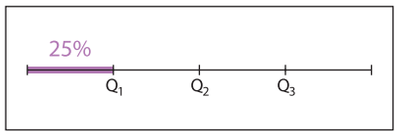
Conocimientos previos

Determina la mediana del siguiente grupo de datos. Recuerda que este valor divide en la mitad a la distribución de datos.

3, 5, 6, 7, 3, 5, 4, 6, 8, 7, 9, 1 3, 5, 6, 0 5, 6, 7, 5, 7, 8,9 2 3,3 3, 9 0, 4, 4, 5, 6

CUARTILES

Son tres valores que dividen al conjunto de observaciones ordenadas en cuatro partes iguales. Por lo tanto, el **primer cuartil** ($\mathbf{Q_1}$) es el valor por debajo del cual, o en el cual, se ubica el 25% de todos los valores; el **segundo cuartil** ($\mathbf{Q_2}$) es el valor por debajo del cual se ubica el 50% de todos los valores y el **tercer cuartil** ($\mathbf{Q_3}$) es el valor por debajo del cual se ubica el 75% de todos los valores. Gráficamente, se los representa de la siguiente manera.



Para determinar cada cuartil, se utilizan las siguientes expresiones:

$$Q_1 = L_i + \frac{\frac{n}{4} - F_{i-1}}{f_i \cdot a} \qquad Q_2 = L_2 + \frac{\frac{n}{2} - F_{i-1}}{f_i \cdot a} \qquad Q_3 = L_3 + \frac{\frac{3n}{4} - F_{i-1}}{f_i} \cdot a$$

$$Q_i = L_i + \frac{i\left(\frac{n}{10}\right) - F_{i-1}}{f_i} \cdot a$$

Toma en cuenta

- El segundo cuartil coincide con la mediana.
- El quinto decil coincide con la mediana.
- Los deciles se denotan como $\mathbf{D_1}$, $\mathbf{D_2}$, ..., $\mathbf{D_9}$.

Donde:

i = 1, 2, 3.

$\mathbf{L_i}$: límite inferior del intervalo que contiene al cuartil.

$\mathbf{F_{i-1}}$: frecuencia acumulada del intervalo anterior que contiene al cuartil.

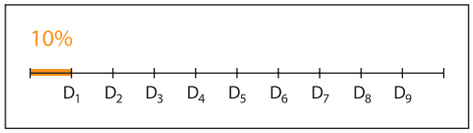
a: ancho del intervalo donde está el cuartil.

$\mathbf{f_i}$: frecuencia absoluta del intervalo que contiene al cuartil.

n: tamaño de la muestra.

DECILES

Los deciles corresponden a nueve valores que dividen al conjunto de observaciones ordenadas en diez partes iguales. Gráficamente, se representan así.



BECU_M1_B4_P144_173.indd 164 4/22/14 11:59 AM

Recuerda

Para calcular medidas de localización se utilizan las frecuencias acumuladas. Si la variable es continua, se toma como valor la marca de clase.

Para determinar el i-ésimo decil, se utiliza la siguiente expresión.

$$D_i = L_i + \frac{i\left(\frac{n}{10}\right) - F_{i-1}}{f_i} \cdot a$$

Donde:

i = 1, 2, 3, ..., 9.

$\mathbf{L_i}$: límite inferior del intervalo que contiene al decil.

n: número total de observaciones.

$\mathbf{F_{i-1}}$: frecuencia acumulada del intervalo que antecede al decil.

$\mathbf{f_i}$: frecuencia absoluta del intervalo al que pertenece el decil.

a: longitud del intervalo que contiene al decil.

PERCENTILES

Corresponden a 99 valores que dividen al conjunto de observaciones, ordenadas en cien partes iguales.

Para determinar el i-ésimo percentil se utiliza la siguiente expresión:

Donde:

i = 1, 2, 3, ..., 99.

$$P_i = L_i + \frac{i\left(\frac{n}{100}\right) - F_{i-1}}{f_i} \cdot a$$

$\mathbf{L_i}$: límite inferior del intervalo que contiene al percentil.

n: número total de observaciones.

$\mathbf{F_{i-1}}$: frecuencia absoluta acumulada del intervalo que antecede al percentil.

$\mathbf{f_i}$: frecuencia absoluta del intervalo al que pertenece el percentil.

a: longitud del intervalo que contiene al percentil.

Ejemplo

A partir de la siguiente tabla de frecuencias, determinar P_{90}.

Intervalo	f_i	F_i
60 – 69	10	10
70 – 79	20	30
80 – 89	20	50

Solución

El 90% de los valores de la tabla se ubican en el intervalo 80 – 89.

Se tiene: L_i = 80 (límite inferior del intervalo 80 - 89); n = 50 (tamaño de la muestra); f_i = 50 (frecuencia absoluta del intervalo 80 – 89); F_{i-1} = 30 (frecuencia acumulada del intervalo 70 - 79); a = 10 (amplitud del intervalo).

Remplazando los valores anteriores en la fórmula respectiva, se tiene:

$$P_{90} = 80 + \frac{90\left(\frac{50}{100}\right) - 30}{20} \cdot 10 = 87{,}5.$$

Luego, P_{90} = 87,5. Es decir, el 90% de los datos es menor o igual que 87,5.

Observa

Los percentiles se denotan como $\mathbf{P_1}$, $\mathbf{P_2}$, ..., $\mathbf{P_{99}}$.

Esta medida es una de las más utilizadas cuando el objetivo es la clasificación de personas respecto a alguna característica, por ejemplo, el peso o la estatura.

BECU_M1_B4_P144_173.indd 165 4/22/14 11:59 AM

Diagrama de caja

Destreza con criterio de desempeño:

Reconocer en diferentes **diagramas estadísticos** (tallo y hojas, polígonos de frecuencia, gráfico de barras, caja y bigotes, histogramas, etc.) la información que estos proporcionan. **(C)**

Conocimientos previos

En una población de 25 familias, se ha asignado a la variable *x* al número de autos que tiene la familia, y se han obtenido los siguientes datos: 0, 1, 2, 3, 1, 0, 1, 1, 1, 4, 3, 2, 2, 1, 1, 2, 2, 1, 1, 1, 2, 1, 3, 2 y 1.

Determina los cuartiles.

*Un **diagrama de caja** (box - plot) es una representación gráfica que se construye a partir de los cuartiles de un conjunto de valores de una variable. Además, en este tipo de gráfico se indican otros elementos de la distribución, tales como: rango, mediana, etc. En general, este gráfico se utiliza para comparar las distribuciones de diferentes grupos.*

CONSTRUCCIÓN DE UN DIAGRAMA DE CAJA

Para su construir este diagrama es necesario conocer los siguientes datos para cada variable.

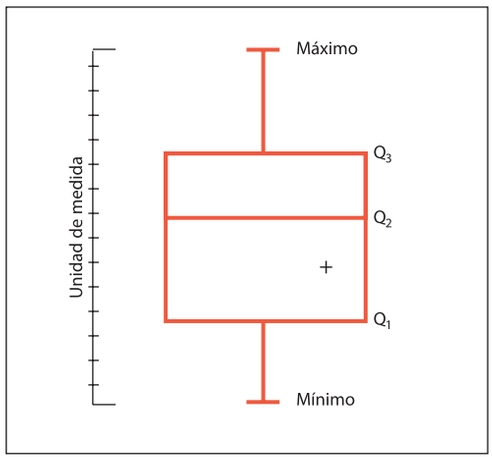
- Valor mínimo.
- Valor máximo.
- Primer cuartil.
- Segundo cuartil o mediana.
- Tercer cuartil.
- Media aritmética de los valores de la variable (este dato no es imprescindible para el diagrama, pero, en caso de conocerlo, se incluye en la gráfica).

En un diagrama de caja, se puede observar lo siguiente (ver figura).

- Las líneas que sobresalen del rectángulo, indican el valor mínimo y máximo de los valores de la variable.
- Los extremos inferior y superior del rectángulo indican el primer y tercer cuartil, respectivamente, mientras que la línea horizontal (o vertical) que divide al rectángulo indica la mediana (segundo cuartil).
- Para indicar la media de los valores (si se conoce) de la variable se utiliza un signo **+**.

Glosario

box-plot. Diagramas de cajas también conocidos como gráficos de cajón con bigotes.



BECU_M1_B4_P144_173.indd 166 4/22/14 11:59 AM

Recuerda

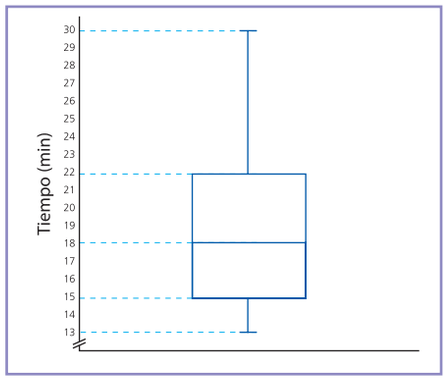
Sir Francis Galton (1822-1911)

Destacado científico cuyos aportes más importantes los realizó en el campo de la estadística. Se le atribuye la creación del cálculo correlacional.

Ejemplos

1. Un restaurante seleccionó una muestra de 30 entregas a domicilio durante un mes, obteniendo la siguiente información.

- Tiempo mínimo de entrega: 13 minutos.
- Tiempo máximo de entrega: 30 minutos.
- Q_1: 15 minutos.
- Mediana del tiempo de entrega: 18 minutos.
- Q_3: 22 minutos.



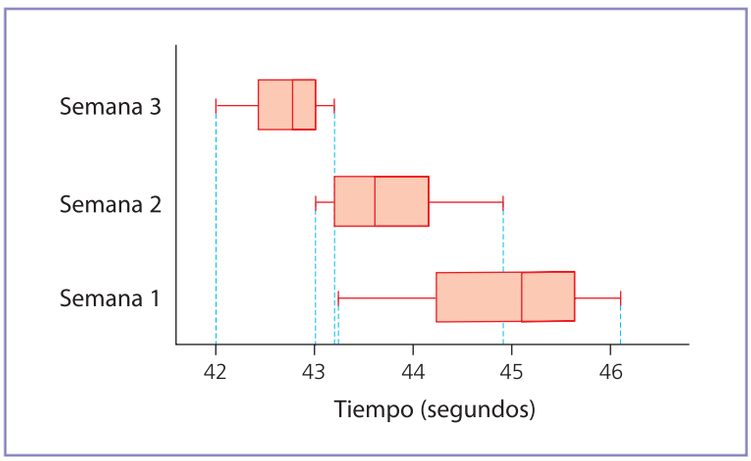
Solución

Con la información anterior se construyó el diagrama de caja. A partir de este, se puede observar que el tiempo de entrega de este restaurante está concentrado bajo el tercel cuartil, es decir, sus tiempos de entrega son, frecuentemente, bajos (demoran poco tiempo en las entregas).

2. Se registraron los tiempos obtenidos por una nadadora en 50 m pecho, durante su entrenamiento en tres semanas diferentes, con el objetivo de averiguar si la práctica favorecía el logro de mejores marcas.

Solución

A partir de los tiempos obtenidos en cada semana, se construyó el siguiente diagrama:



A partir del gráfico, se puede concluir que, a medida que transcurrieron los días de entrenamiento, la diferencia entre la máxima y mínima marca de cada semana es menor. Además, existe un desplazamiento de la gráfica hacia la izquierda, lo que muestra que los tiempos obtenidos han mejorado respecto de la semana anterior y, por ende, la práctica está dando resultados.

Toma en cuenta

Posición de un diagrama

Un diagrama de cajas se puede construir de manera vertical, o bien, horizontal.

BECU_M1_B4_P144_173.indd 167 4/22/14 11:59 AM

Actividades

Elabora un diagrama de tallo y hoja.

1. Dibuja un diagrama de tallo y hojas para el siguiente conjunto de datos. Luego, responde las preguntas.

5,0	2,8	2,5	6,8
6,2	5,7	5,8	3,7
3,5	3,3	2,2	2,7
5,1	5,5	4,3	4,0
3,7	3,9	3,4	4,3
2,7	3,2	4,9	7,0
4,3	4,8	4,0	4,2
5,7	4,0	4,6	5,3
4,5	5,0	6,0	4,1
5,3	4,5	4,8	5,3

a. ¿Cuál es la media de los datos?

b. ¿Cuál es el rango de los datos?

c. ¿Cómo es la dispersión de los datos?

Resuelve problemas de medidas de tendencia central.

2. Resuelve los siguientes problemas.

a. Andrés se entrena para participar en una carrera de 100 m planos obteniendo los siguientes tiempos medidos en segundos: 12,9; 13,1; 12,4; 13,2 y un quinto tiempo que no recuerda. Si el promedio de los tiempos fue 12,88 s, calcula el tiempo faltante.

b. Tres hombres de negocios tienen igual promedio de ganancias durante un año. ¿Significa esto que han recibido mensualmente las mismas ganancias? ¿Por qué?

c. El mayor de cuatro hermanos recibe $ 30 de mesada, mientras que sus hermanos reciben $ 15, $ 18 y $ 20. ¿La media aritmética de las mesadas es un valor representativo de ellas?

d. Las estaturas en centímetros de diez estudiantes son:159, 168, 173, 168, 173, 159, 165, 167, 173 y 182. Encuentra la estatura media, la mediana y la moda.

Analiza una tabla de datos no agrupados, y agrupados.

3. La siguiente distribución de frecuencias corresponde a los salarios de los empleados de una fábrica. Observa los datos y responde.

Salario ($)	Frecuencia
350 - 354,9	7
355 – 359,9	18
360 – 364,9	32
365 – 369,9	45
370 – 374,9	52
375 – 379,9	28
380 – 384,9	16
385 – 389,9	8

a. Calcula el tamaño de los intervalos.

b. ¿Cuál es el límite inferior del séptimo intervalo?

c. ¿Cuál es el límite superior del segundo intervalo?

d. Escribe en orden las marcas de clase de los intervalos.

e. Agrega a esta tabla la frecuencia acumulada, la frecuencia relativa y la frecuencia relativa porcentual.

f. Grafica en Excel un histograma de frecuencias.

4. Se lanzó un dado cierta cantidad de veces y, con los valores obtenidos, se elaboró la siguiente tabla de frecuencias.

Resultado	f_i
1	5
2	2
3	4
4	x
5	4
6	7

a. Si la media aritmética de los resultados es 3,8, ¿cuál es el número total de lanzamientos?

b. Determina el valor de la mediana.

BECU_M1_B4_P144_173.indd 168 4/22/14 11:59 AM

Analiza un problema mediante medidas de tendencia central y gráficos.

5. En un centro hospitalario se ha estudiado el número de días que han demorado ciertos pacientes en sentir mejoría con el consumo de un nuevo medicamento. Los resultados obtenidos fueron los siguientes.

Días mejoría	Frecuencia
0	100
1	150
2	200
3	600
4	450
5	600

a. Calcula la media aritmética, la mediana y la moda.

b. Calcula el rango de los datos.

c. Construye un pictograma para representar las frecuencias anteriores.

d. A partir de los parámetros estadísticos obtenidos y del gráfico, obtén al menos tres conclusiones.

6. En 15 días de trabajo se contabilizó el tiempo de espera (en minutos) de locomoción colectiva para desplazarse desde el hogar hasta el trabajo. Los tiempos registrados son los siguientes:

10	1	13	9	5
9	2	10	3	8
6	17	2	10	15

a. Determina la media aritmética, la mediana y la moda de los tiempos.

b. ¿Cuál de las medidas de tendencia central anteriores es más apropiada para representar el tiempo de espera? Justifica tu respuesta.

c. Mediante un gráfico estadístico, representa la información.

d. A partir del gráfico, obtén dos conclusiones.

Determina la mediana de una serie de datos.

7. Determina la mediana de los datos presentados en cada caso.

a.

3,0	4,5	6,0	14,0
8,2	9,0	12,0	

b.

12,0	15,3	17,0	21,0
25,1	26,1	28,0	28,0
28,0	30,0		

Analiza situaciones problema.

8. Responde: Qué ocurre con la mediana y la desviación estándar si:

a. en un conjunto de datos, cada observación se multiplica por 2.

b. en un conjunto de datos, se le suma 5 unidades a cada observación.

Resuelve problemas de medidas de tendencia central.

9. Resuelve los siguientes ejercicios.

a. Encuentra la media de 25 números sabiendo que la media de 7 de ellos es 3,6 y la media de los otros 18 números es 5,1.

b. La media de los números 3, 7, 8, 10 y x es 6. Determina el valor de x.

c. La media de los pesos de 5 deportistas es de 76 kg. Los pesos de 4 de ellos son 72 kg, 74 kg, 75 kg y 81 kg. ¿Cuál es el peso del quinto deportista?

d. Los números 3, 5, 7, 8 y C se encuentran ordenados en forma creciente. Determina el valor de C si la media es igual a la mediana.

e. Una fábrica produce pastelillos que deben pesar 85 g. Analizadas 400 unidades elegidas al azar, se obtuvo la siguiente información.

Edad (años)	x_i	f_i	F_i
[25, 35[	30	5	5
[35, 45[	40	6	11
[45, 55[	50	8	19
[55, 65[	60	6	25

- Calcula el D_2 y el Q_2.
- Calcula los P_{15} y el P_{75}.

BECU_M1_B4_P144_173.indd 169 4/22/14 11:59 AM

Actividades

Analiza información mediante las medidas de localización.

10. La siguiente tabla corresponde a la distribución de los puntajes de un grupo de estudiantes que rendirán un examen de Matemáticas para la universidad.

Puntaje	Nº de alumnos	%
250 – 299	123	0,1
300 – 349	2 910	1,4
350 – 399	2 389	1,1
400 – 449	10 015	4,7
450 – 499	25 223	4,7
500 – 549	25 999	12,3
550 – 599	37 209	17,6
600 – 649	39 411	18,6
650 – 699	28 645	13,5
700 – 749	21 367	10,1
750 – 799	11 1897	5,3
800 – 849	4 247	2,0
850 – 899	1 597	0,8
900 – 950	1 197	0,6
Total	211 519	100,0

a. ¿Cuántos estudiantes rindieron la prueba?

b. ¿Qué porcentaje de estudiantes obtuvo puntajes menores que los del intervalo 300 – 349?

c. ¿Qué porcentaje de estudiantes obtuvo más de 800 puntos?

d. Calcula la media aritmética de esta distribución de puntajes.

e. Calcula Q_1, Q_2 y Q_3.

f. Calcula P_{10} y P_{90}.

g. Calcula la desviación estándar de esta distribución de puntajes.

h. Construye, utilizando Excel, una curva de distribución de frecuencias.

i. ¿A qué distribución se asemeja la forma de esta curva?

Comprende situaciones de la vida cotidiana a través de la interpretación de datos estadísticos.

11. Se desea estudiar el número de errores de impresión de un libro. Para tal efecto, se selecciona una muestra de 50 páginas. Los resultados fueron los siguientes.

Nº de errores	f_i
0	25
1	20
2	3
3	1
4	1

a. ¿Cuál es el número medio de errores por página?

b. ¿Cuál es la mediana de errores por página?

c. Si el libro tiene 500 páginas, ¿cuántos errores se esperaría que tuviera el libro?

12. Se tiene un conjunto de 5 números consecutivos. La mediana de ellos es N.

a. Obtén la media del conjunto de datos.

b. Encuentra la media de los cuadrados de estos números.

c. Encuentra P_{10}.

d. Responde: ¿Cómo se puede interpretar la información anterior?

13. Resuelve los siguientes ejercicios.

a. Calcula el percentil 65 de los siguientes datos. Luego, interpreta el resultado obtenido.

Valores	f_i
2	12
4	10
6	8
8	7
10	5
12	8
14	10

BECU_M1_B4_P144_173.indd 170 4/22/14 11:59 AM

b. Ordena los siguientes datos de menor a mayor. Luego, calcula el rango de los datos.

3,22	2,93	3,01	4,48
5,06	4,31	2,98	3,07

c. Responde: ¿Qué significa que una persona haya obtenido un puntaje superior al noveno decil (D_9) en un cuestionario de interés científico?

d. La estatura del estudiante más alto de un curso es 1,92 m y la del menor es 1,68 m. Calcula el rango de la estatura de este grupo.

e. Se realizó un estudio estadístico con el objetivo de averiguar la antigüedad de los clientes de una empresa. A partir de los análisis, se pudo determinar que la antigüedad media de un cliente de esta empresa es de 4 años. En cambio, la mediana arrojó un valor de 2,5 años.

i) ¿Cómo se puede interpretar esta información?

ii) ¿Cómo es la desviación estándar de los datos obtenidos?

f. Calcula la desviación estándar de los siguientes datos: 3, 5, 6, 7, 10, 12, 15, 18.

g. Un grupo de deportistas obtuvo las siguientes marcas en atletismo (expresada en horas): 2,60; 2,40; 2,30; 2,00; 2,10. Calcula e interpreta la desviación media de los datos.

h. La siguiente tabla de distribución muestra la estaura de los miembros de un curso de educación media de cierto colegio. Calcula e interpreta, la desviación estándar de estos datos.

Estuaturas	Nº de alumnos (f_i)
[150 - 155[	3
[155 - 160[	7
[160 - 165[	6
[165 -170[	4
[170 - 175[	5

i. Se tienen dos variables X y Y. Ambas presentan la misma media, pero con desviación estándar de 4 y 9, respectivamente. ¿Para cuál de las dos variables la media es la medida más representativa?

j. La media de una variable Z es 10 unidades y la desviación estándar es cero. ¿Qué se puede afirmar respecto de la variable Z?

14. Se lanzó un dado 25 veces, obteniendo los siguientes resultados:

5	3	2	6	5
1	2	3	2	1
5	1	5	2	4
5	6	1	2	4
4	2	2	4	3

a. Si se grafica esta información con un polígono de frecuencias, ¿a qué se asemejaría la curva obtenida? Justifica tu respuesta.

b. Calcula P_{30}.

c. Obtén al menos dos conclusiones a partir de la información anterior.

15. En una prueba, dos cursos A y B obtuvieron los siguientes resultados.

Curso A → Media: 5,1 y s: 0,8
Curso B → Media: 5,2 y s: 0,2
De acuerdo con la información anterior, responde estas preguntas:

a. ¿En qué curso las notas fueron más homogéneas?

b. Es la media de las notas obtenidas por cada curso un valor representativo?

c. Un estudiante del curso A obtuvo un 6,5 y uno del curso B un 6,2. ¿A cuál de los dos estudiantes le fue mejor en la prueba, dentro de su curso?

BECU_M1_B4_P144_173.indd 171 4/22/14 11:59 AM