Introducción

Los pueblos antiguos al utilizar registros estadísticos rudimentarios en los censos de población y de propiedades hicieron un tratamiento de la información obtenida al organizarla y presentarla en tablas con el propósito de llevar un mejor control.

En tiempos más recientes se introdujo el uso de las gráficas, con la ventaja de que estas permiten observar mejor las relaciones que se dan entre los datos y se percibe en forma más clara la información.

Las actividades que día a día realizan los habitantes de una comunidad o de una sociedad se llevan en registros, son ejemplos de estos los nacimientos, las defunciones, los casamientos, etcétera. El conteo y la medición de tales hechos generan una gran cantidad de información que se hace necesario ordenarla, clasificarla y analizarla para saber qué dicen de su comportamiento en un período de tiempo.

Las instituciones como el Instituto Nacional de Estadísticas (INE) dan a conocer esta información por medio de tablas y gráficas, para que la población tenga conocimiento de cuál ha sido su desarrollo. Esto hace resaltar la importancia del manejo de la información.

En esta unidad, el estudio de la presentación y tratamiento de la información incluye el uso de porcentajes, tablas, gráficas y otras formas de presentar la información registrada.

En el desarrollo de estos temas se utilizarán los instrumentos de geometría para la construcción de tablas y gráficas y la calculadora para abreviar tiempo en la obtención o comprobación de resultados.

¿Qué vamos a aprender?

Competencias	Objetivos	Contenido
1. Desarrollan los conceptos de estadística, población y muestra. 2. Recolectan y organizan datos estadísticos. 3. Interpretan y comunican información presentada en tablas y gráficos.	1. Explicar los conceptos estadísticos. 2. Clasificar mediante instrumentos sencillos datos estadísticos. 3. Elaborar tablas y gráficas sencillas con datos extraídos de su entorno. 4. Interpretar y comunicar información estadística organizada en tablas y gráficos sencillos.	1. Estadística, población y muestra 2. Organización de datos estadísticos 3. Tablas y gráficos
5. Calculan las medidas de tendencia central de datos en frecuencia simple y agrupada. 6. Calculan las medidas de dispersión para datos en frecuencia simple y agrupada. 7. Valoran la importancia de la estadística en la realidad.	5. Calcular las medidas de tendencia central de datos en frecuencia simple y agrupada. 6. Resolver ejercicios de medidas de dispersión para datos en frecuencia simple y agrupada. 7. Justificar la importancia de la estadística en la vida cotidiana.	4. Medidas de tendencia central de datos en frecuencia simple y agrupada 5. Medidas de dispersión para datos en frecuencia simple y agrupada

Mis conocimientos previos

Instrucciones

Realice las actividades que a continuación se le solicitan, luego comparta con sus compañeros y tutor los resultados.

1. Suponga que en una maquila se obtuvo la siguiente producción de camisetas durante 25 días de trabajo de un mes:

140	152	146	140	160
155	149	152	148	147
150	141	146	152	157
148	155	152	160	148
160	140	152	148	155

a. Ordene los siguientes datos de mayor a menor.

b. Registre los datos en la tabla que se muestra a continuación, marque con una rayita en la columna conteo el número de veces que se repiten los datos:

Producción de camisetas	Conteo
160	
157	
155	
152	
150	
149	
148	
147	
146	
141	
140	

c. Encuentre la oscilación o rango restando el dato mayor del menor:

2. Se preguntó la edad a 40 alumnos de un instituto, los datos fueron los siguientes:

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

a. Ordene los datos en forma decreciente.

b. Regístrelos en una tabla que contenga los datos ordenados y su conteo.

c. Las edades de 10 personas entrevistadas al azar fueron: 27, 14, 13, 16, 17, 18, 15, 14, 12 y 18 años. Ordenar los datos en forma creciente (mayor a menor) y en forma tabular (en una tabla).

Estadística

La palabra estadística procede del latín statisticum collegium (consejo de Estado) y de su derivado italiano statista (hombre de Estado o político). El término alemán statistik, que fue inicialmente introducido por Gottfried Achenwall (1749), designaba originalmente al análisis de datos del Estado, es decir, la ciencia del Estado (también llamada aritmética política, de su traducción directa del inglés). Fue hasta en el siglo XIX cuando el término estadística adquirió el significado de recolectar y clasificar datos. Este concepto fue introducido por el inglés John Sinclair.

En su origen la estadística estuvo asociada a datos, al ser utilizados por los gobiernos y cuerpos administrativos. La colección de datos acerca de Estados y localidades continúa ampliamente a través de los servicios de estadística nacional e internacional. En particular, los censos suministran información regular acerca de la población.

Desde los comienzos de la civilización han existido formas sencillas de estadística, pues ya se utilizaban representaciones gráficas y otros símbolos en pieles, rocas, palos de madera y paredes de cuevas para contar el número de personas, animales o ciertas cosas.

Hacia el año 3000 a. C., los babilónicos usaban pequeñas tablillas de arcilla para recopilar datos en tablas sobre la producción agrícola y de los géneros vendidos o cambiados mediante trueque. Los egipcios analizaban los datos de la población y la renta del país mucho antes de construir las pirámides en el siglo XI a. C. Los libros bíblicos de Números y Crónicas incluyen, en algunas partes, trabajos de estadística. El primero contiene dos censos de la población de Israel y el segundo describe el bienestar material de las diversas tribus judías. En China existían registros numéricos similares con anterioridad al año 2000 a. C. Los griegos clásicos realizaban censos cuya información se utilizaba hacia el 594 a. C. para cobrar impuestos.

La estadística es la ciencia fundamentada en la matemática que nos proporciona un conjunto de métodos y técnicas para la recolección, organización, presentación y análisis de datos, describiéndolos e interpretándolos para la toma de decisiones válidas en una determinada característica de una población.

Según se haga el estudio sobre todos los elementos de la población o sobre un grupo de ella, vamos a diferenciar dos tipos de estadística: descriptiva e inferencial.

Estadística descriptiva

Realiza el estudio sobre la población completa, observando una característica de la misma y calculando unos parámetros que den información global de toda la población. Se dedica a los métodos de recolección, descripción, visualización y resumen de datos originados a partir de los fenómenos en estudio. Los datos pueden ser resumidos numérica o gráficamente.

Ejemplos

- El nivel promedio de inteligencia obtenido mediante la prueba WAIS, resultó ser de 104 para el grupo 2 de estudiantes de Psicología de la UNAH.
- Durante los últimos dos días se han informado un total de doce homicidios.
- Según la encuesta del COHEP realizada en julio de 2013 sobre las preferencias electorales, se obtuvieron los siguientes datos:

  Xiomara Castro (Partido Libre): 28 %

  Salvador Nasralla (Partido Anticorrupción): 23 %

  Mauricio Villeda (Partido Liberal): 20 %

  Juan Orlando Hernández (Partido Nacional): 17 %

Estadística inferencial

Realiza el estudio descriptivo sobre un subconjunto de la población llamado muestra y, posteriormente, extiende los resultados obtenidos a toda la población. Se dedica a la generación de los modelos, inferencias y predicciones asociadas a los fenómenos en cuestión teniendo en cuenta lo aleatorio e incertidumbre en las observaciones. Se usa para modelar patrones en los datos y extraer inferencias acerca de la población de estudio.

Estas inferencias pueden tomar la forma de respuestas a preguntas sí/no (prueba de hipótesis), estimaciones de características numéricas (estimación), pronósticos de futuras observaciones, descripciones de asociación (corre lación) o modelamiento de relaciones entre variables (análisis de regresión).

Ejemplo

De acuerdo con una encuesta desarrollada por una empresa sobre telefonía residencial en el 2009, el gasto mensual promedio por cliente es de L.300.00.

Población

El concepto de población en estadística va más allá de lo que comúnmente se conoce como tal. Una población se precisa como un conjunto finito o infinito de personas u objetos que presentan características comunes, es un conjunto de los elementos que estamos estudiando acerca de los cuales intentamos sacar conclusiones.

El tamaño que tiene una población es un factor de suma importancia en el proceso de investigación estadística y en nuestro caso social; este tamaño viene dado por el número de elementos que constituyen la población. Cuando el número de elementos que integra la población es muy grande, se puede considerar a ésta como una población infinita, por ejemplo, el conjunto de todos los números primos.

Una población finita es aquella que está formada por un limitado número de elementos.

<u>Ejemplos</u>

- El número de habitantes de un barrio o colonia.
- Conjunto de valores del coeficiente intelectual de los alumnos de bachillerato de Educatodos.

Cuando se vaya a llevar a cabo alguna investigación, deben tenerse en cuenta algunas características esenciales al seleccionarse la población bajo estudio, tales como:

- Homogeneidad: que todos los miembros de la población tengan las mismas características según las variables que se vayan a considerar en la investigación.
- Tiempo: se refiere al período de tiempo donde se ubicaría la población. Determinar si el estudio es en tiempo presente o de años atrás o si se van a entrevistar personas de diferentes generaciones.
- Espacio: es el lugar donde se ubica la población de interés. Un estudio no puede ser muy abarcador y por falta de tiempo y recursos hay que limitarlo a un área o comunidad específica.
- Cantidad: se refiere al tamaño de la población, es sumamente importante porque ello determina o afecta al tamaño de la muestra que se vaya

a seleccionar, además que la falta de recursos y tiempo también puede limitar la extensión de la población que se vaya a investigar.

Muestra

Es una colección seleccionada de una población de interés. Existen diferentes técnicas para realizar el muestreo, las cuales dependerán de cada caso; algunas de ellas son:

- Muestreo aleatorio simple: todos los elementos de la población tienen igual posibilidad de ser escogidos y se eligen al azar.
- Muestreo sistemático: los elementos se seleccionan a un intervalo uniforme en una lista ordenada. Una preocupación de este muestreo es la existencia de factores cíclicos en el listado que pudieran dar lugar a un error.
- Muestreo estratificado: los elementos de la población son primeramente clasificados en grupos o estratos según una característica importante, luego de cada estrato se extrae una muestra aleatoria simple.
- Muestreo por conglomerado: los elementos de la población están subdivididos en grupos y se extraen aleatoriamente algunos de estos grupos completos.

Ejemplo

Identificar la población y la muestra en la siguiente situación:

En el Instituto Dr. Jesús Aguilar Paz se quiere saber la ocupación de los egresados de la última década. Para esto se convoca a una reunión de egresados y de los asistentes, se encuesta a diez egresados de cada año. Determinar la población y la muestra.

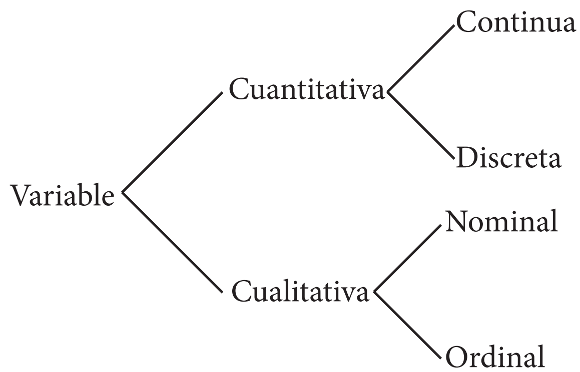
Solución

Población: Todos los egresados de la última década.
Muestra: Los 100 estudiantes seleccionados, 10 de cada promoción.

Al hacer un Estudio de una determinada población, observamos una característica o propiedad de sus elementos o individuos. Por ejemplo, con los alumnos de una clase se pueden estudiar el lugar de residencia, el número de hermanos, la estatura, etc. Cada una de estas características estudiadas se llama variable estadística o carácter estadístico.

Variable

Una variable estadística es cada una de las características o cualidades que poseen los individuos de una población. Dependiendo de la característica es posible distinguir varios tipos de variables:



- Variable cualitativa: es aquella característica que no podemos expresar con números y hay que expresarla con palabras. Por ejemplo, el lugar de residencia.
- Variable cuantitativa: es cualquier característica que se puede expresar con números. Por ejemplo, el número de hermanos o la estatura. Dentro de esta variable podemos distinguir dos tipos:
  - Variable cuantitativa discreta: es aquella variable que puede tomar únicamente un número finito de valores, por ejemplo, el número de hermanos.
  - Variable cuantitativa continúa: es aquella variable que puede tomar cualquier valor dentro de un intervalo real, por ejemplo, la estatura.
- Variable cualitativa ordinal: las variables son números, no magnitudes y en este caso los números indican orden: 1˚primero, 2˚segundo.
- Variable cualitativa nominal: presenta modalidades no numéricas que no admiten orden, por ejemplo, el estado civil con las siguientes modalidades: soltero, casado, separado y viudo.

<u>Ejemplo 1</u>

Justifique por qué la siguiente situación es una cualitativa ordinal:

Una comercializadora de helados saca una nueva marca al mercado en presentaciones de diferentes sabores. Para comercializarlos se ofrecen degustaciones a un grupo de personas, quienes luego de degustarlas deben clasificar en muy buenos, regulares y malos.

Solución

La situación anterior es cualitativa ordinal porque las variables indican un orden, en este caso, el orden está dado por la categoría según el gusto.

Ejemplo 2

Justifique por qué la siguiente situación es cualitativa nominal:

En una encuesta a un grupo de empleados de una empresa, se les pregunta sobre su ocupación y se dan las siguientes opciones de respuesta:

- Directivo
- Practicante
- Administrativo
- Funcionario
- Personal de seguridad

Solución

La solución anterior es cualitativa nominal, ya que las variables no indican orden.

Para hacer un estudio estadístico de una característica de una población, es necesario elegir dicha característica y después hacer un recuento. Uno de los primeros recuentos que hacemos en el aula con los alumnos es la elección del presidente o presidenta del curso. Este recuento puede resultar más o menos fácil dependiendo del número de alumnos.

Una vez realizado el recuento, hay que organizar los datos y expresarlos de forma simplificada para que su interpretación sea fácil y rápida. Esto se hace disponiendo los datos por columnas o filas, formando lo que llamamos una tabla estadística.

Recolección de datos

Es el momento en el cual el investigador se pone en contacto con los objetivos o los elementos sometidos a estudio, con el propósito de obtener los datos o respuestas de las variables consideradas.

Técnicas de recolección de datos

Las técnicas de recolección de datos son muy variadas y dependen de la naturaleza del objeto de estudio, de las posibilidades de contacto con los elementos de la investigación, del tamaño de la población y la muestra, así como de los recursos para obtener los datos.

Un instrumento de recolección de datos es cualquier recurso del cual se vale el investigador para extraer la información. Todo instrumento usado en la recolección de datos de una investigación científica debe poseer tres requisitos: validez, confiabilidad y objetividad.

Las técnicas de mayor uso son las siguientes:

a. Fuentes primarias o directas: son los datos obtenidos de primera mano por el propio investigador o, en el caso de búsqueda bibliográfica, por artículos científicos, monografías, tesis, libros o artículos de revistas especializadas originales, no interpretados.

b. Fuentes secundarias: consisten en resúmenes, compilaciones o listados de referencias, preparados con base a fuentes primarias. Es información ya procesada.

Fuentes primarias

- La observación
- Técnica documental
- El cuestionario
- La encuesta
- La entrevista
- Experimentos
- Fotografías

Fuentes secundarias

- Bibliotecas (libros y revistas)
- Documentos (actas y cartas, películas, diarios, periódicos)
- Censos
- Expedientes

La obtención de la información se puede realizar por diversos medios. Una forma es a través de una encuesta a un grupo de individuos, en la cual a cada uno se le hacen las mismas preguntas. Otra forma es a través de experimentos en donde la respuesta a la variable es el resultado del experimento. También pueden recolectarse los datos en forma directa, es decir, la información se extrae de alguna base de datos seleccionando una muestra de ellos. En cualquiera de estos casos contamos con una selección de información, llamada muestra, que se procede a analizar.

Registro de datos

En muchas actividades del género humano se requiere realizar encuestas o recopilar datos para posteriormente organizarlos y efectuar un análisis de la información obtenida, lo cual permitirá tomar decisiones que sirvan para evaluar los procesos, mejorar las áreas donde se detectan errores, etcétera.

Cuando se recaba determinada información, lo primero que procede es la organización de datos y su tabulación.

Ejemplo

En la ciudad de Comayagüela circulan una gran cantidad de vehículos en determinadas horas del día en la mayor parte de avenidas y calles.

La recopilación de esos datos, su agrupación y conteo, registro en tablas, sirve a las autoridades de la alcaldía municipal para decidir qué medidas se deben tomar en cuanto al sentido de las calles, colocación de semáforos, vigilancia, etc., para evitar grandes congestionamientos de vehículos, así como pérdida de tiempo y molestias a quienes transiten por la ciudad.

A continuación aparecen los datos que se obtuvieron contando el número de vehículos que cruzan cierta calle cada cinco minutos:

28	32	24	26	23	23	30	25	
34	30	32	25	25	20	25	37	
39	34	36	28	24	26	26	24	
24	21	20	22	28	29	26	27	27
26	31	33	31	28	38	35	28	27

Al ordenar los datos en forma decreciente y registrar el conteo en una tabla se obtiene:

TABULACIÓN	
Vehículos	Conteo
39	I
38	I
37	I
36	I
35	I
34	II
33	I
32	II
31	II
30	II
29	I
28	~~IIII~~
27	III
26	~~IIII~~
25	IIII
24	IIII
23	II
22	I
21	I
20	II

Organización y representación de datos

Distribución de datos no agrupados

Básicamente las técnicas que permiten organizar los datos, son la tabular y la gráfica. Es una de las formas más sencillas de presentarlos, generalmente se colocan los valores en forma ascendente o creciente (menor a mayor), también se acostumbra colocarlos en forma descendente o decreciente (mayor a menor), lo cual ofrece las siguientes ventajas:

a. Se descubren rápidamente los valores mínimos y máximos en los datos.
b. Se pueden dividir fácilmente los datos en secciones.
c. Se puede dar cuenta si algunos valores aparecen más de una vez en el arreglo.
d. Se puede observar la distancia entre los valores consecutivos de la tabla o arreglo.

¿Qué puede hacer una persona para organizar los números desordenados que recoge en sus investigaciones? ¿Cómo se las arregla para transformar esa masa de datos en un resumen fácil de entender? El primer paso en la solución de este problema es construir lo que se llama una distribución de frecuencias.

Ejemplo 1

Suponga que las edades de un grupo de 8 estudiantes de séptimo grado son: 12, 13, 14, 12, 14, 13, 12 y 13 años. Se pide ordenarlos en forma creciente y en la forma tabular.

Al ordenarlos en forma creciente quedan así: 12, 12, 12, 13, 13, 13, 14, 14.

Para ordenarlos de forma tabular se construye una tabla de la siguiente manera:

Edad	Conteo
12	III
13	III
14	II

Se puede observar que:

La edad de 12 años se repite 3 veces en el grupo.

La edad de 13 años se repite 3 veces en el grupo.

La edad de 14 años se repite 2 veces en el grupo.

El número de veces que aparece una observación o un mismo valor de la variable, se llama frecuencia y se representa con la letra f .

Considerando la definición anterior, los datos quedarían ordenados en forma tabular de la siguiente manera:

Edad x	Frecuencia f
12	3
13	3
14	2
Total	8

De este ejemplo se tiene que:

La frecuencia de 12 años es 3, entonces: f = 3.

La frecuencia de 13 años es 3, entonces: f = 3.

La frecuencia de 14 años es 2, entonces: f = 2.

Se dice que la distribución anterior es de frecuencia simple o no agrupada, lo cual significa que los valores de las variables (X), en el ejemplo anterior "la edad", no se combina para formar grupos, sino que cada valor de ella es un grupo en sí mismo. El siguiente ejemplo resume toda la información hasta el momento proporcionada.

<u>Ejemplo 2</u>

Los siguientes datos muestran las edades de cierto número de personas:

32	28	41	42	35	35	28
17	17	20	17	17	18	35
18	17	35	42	28	42	21
21	18	21	20	35	20	35

La variable es la edad.

Al ordenar los datos en forma creciente queda:

17	17	17	17	17	18	18
18	20	20	20	21	21	21
28	28	28	32	35	35	35
35	35	35	41	42	42	42

La ordenación es tediosa, pero es muy conveniente hacerla, porque luego permite más fácilmente obtener la tabulación o forma tabular de frecuencia simple, como se muestra en la siguiente tabla:

Edad (valores de las variables)	Frecuencia (número de personas)
17 años	5
18 años	3
20 años	3
21 años	3
28 años	3
32 años	1
35 años	6
41 años	1
42 años	3
Total	28

La oscilación de las edades es de 42 a 17 años, es decir, el rango es de 25.

En una serie de datos, la diferencia entre el valor máximo (V_{max}) y valor mínimo (V_{min}) de la variable se llama rango (Rg).

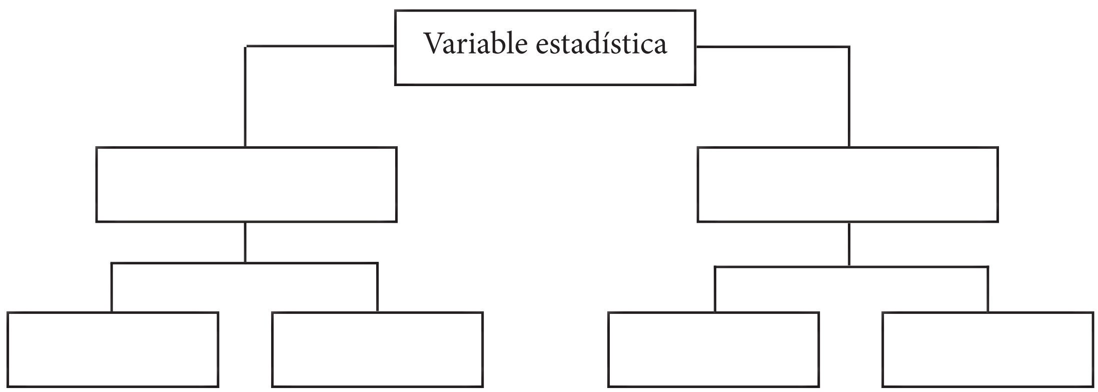
$$Rg = V_{max} - V_{min}$$

Para calcular el rango de la distribución anterior se aplica la fórmula del rango:

$Rg = V_{max} - V_{min} = 42 - 17 = 25$

Actividad 1

1. Complete el siguiente mapa conceptual:



2. Escriba en el espacio en blanco qué tipo de variables son las siguientes (cualitativas o cuantitativas):

- Bebida favorita ____________________
- Profesión que te gusta ____________________
- Número de goles de tu equipo favorito ____________________
- Número de alumnos de tu escuela ____________________
- CI de tus compañeros de clase ____________________

3. Las calificaciones de 50 alumnos de la clase de matemáticas al final del año fueron las siguientes:

68	84	75	82	68	90	62	88	76	93
73	79	88	73	60	93	71	50	85	55
61	75	75	87	74	62	95	76	63	50
66	78	82	75	94	77	69	74	68	60
99	78	89	61	75	95	60	79	53	54

a. Ordene los números en forma descendente.

b. Encuentre el rango.

c. Halle las notas de los 6 alumnos de mayor puntuación.

d. Ubique las notas de los 2 alumnos de menor puntuación.

e. ¿Cuántos alumnos obtuvieron calificaciones mayores que 80?

f. ¿Cuántos alumnos reprobaron?

Distribución de datos agrupados

En ciudades como San Pedro Sula o Tegucigalpa es común que haya una gran afluencia de vehículos circulando en determinadas horas y en ciertas calles. La recopilación de estos datos, su agrupación y conteo, registro en tablas, sirve a las autoridades de la Secretaría de Obras Públicas, Transporte y Vivienda (SOPTRAVI) para decidir qué medidas se deben tomar en cuanto al sentido de las calles: colocación de semáforos, vigilancia, etc.; además para evitar grandes congestionamientos de vehículos, así como pérdida de tiempo y molestias a quienes transiten por la ciudad. Con relación a lo anterior, considere el siguiente ejemplo:

A continuación aparecen los datos que se obtuvieron contando el número de vehículos que cruzan cierta calle cada cinco minutos:

28	32	24	26	23	23	30	25
34	30	32	25	25	20	25	37
39	34	36	28	24	26	26	24
24	20	22	28	29	26	27	27
26	31	33	31	28	38	28	27

Al ordenar los datos en forma descendente y registrar el conteo en una tabla se obtiene lo siguiente:

TABULACIÓN	
Vehículos	Conteo
39	I
38	I
37	I
36	I
34	II
33	I
32	II
31	II
30	II
29	I
28	~~IIII~~
27	III
26	~~IIII~~
25	IIII
24	IIII
23	II
22	I
20	II

Puede observarse que el mayor número de vehículos es 39 y el menor número es 20, de manera que la diferencia entre estos datos es 19, o sea que el rango es de 19:

$Rg = V_{max} - V_{min} = 39 - 20 = 19$

El rango es útil para determinar cómo se pueden agrupar los datos, es decir, cuántos grupos o intervalos de datos se tendrán de acuerdo con la cantidad de ellos (N). Comúnmente, los intervalos son grupos de 3, 5 o 7 datos, los cuales constituyen la amplitud del intervalo. Cabe señalar cada intervalo tiene dos extremos, el número menor se denomina límite inferior y el extremo mayor límite superior.

Si se escoge un intervalo que conste de 5 datos, es decir 39, 38, 37, 36, 35; se expresa como 35-39. De donde se observa que: 39 es el límite superior y 35 es el límite inferior.

Luego se debe determinar el número de intervalos y para esto se divide el rango entre el número de datos del intervalo, o sea: 19 ÷ 5 = 3.8.

Redondeando 3.8 a la unidad más próxima se tiene 4, lo cual indica que debe haber 4 intervalos cuya amplitud sea 5.

Los datos se presentan en una tabla que incluye el intervalo, el conteo se expresa con números y recibe el nombre de frecuencia.

TABULACIÓN		
Intervalo	Conteo	Frecuencia "f "
35-39	IIII	4
30-34	~~IIII~~ IIII	9
25-29	~~IIII~~ ~~IIII~~ IIII III	18
20-24	~~IIII~~ IIII	9
Total		40

Ahora bien, si se hubiera escogido un intervalo que constara de 3 datos, es decir 39, 38, 37; se expresaría como 37 - 39, de tal manera que 39 es el límite superior y 37 es límite inferior.

Para obtener el número de intervalos se divide el rango entre el número de datos del intervalo, o sea: 19 ÷ 3 = 6.3.

Como el número del intervalo debe ser entero y mayor que 6, esto indica que debe haber 7 intervalos con amplitud de 3.

Se realiza la tabulación que incluya el intervalo, el conteo y la frecuencia.

TABULACIÓN		
Intervalo	Conteo	Frecuencia "f "
37-39	III	3
34-36	III	3
31-33	~~IIII~~	5
28-30	~~IIII~~ III	8
25-27	~~IIII~~ ~~IIII~~ II	12
22-24	~~IIII~~ II	7
19-21	II	2
Total		40

Considerando lo anterior se aprecia que cuando hay un número considerable de datos, es conveniente agruparlos en grupos o intervalos, este arreglo facilita la comprensión de los mismos.

Actividad 2

Realice en su cuaderno lo que se le indica, con base en la siguiente información.

Un grupo de 40 alumnos presentó un examen de Ciencias Naturales y se obtuvieron los siguientes resultados:

70	48	43	39	65	67	28	36
56	62	33	45	40	66	63	58
43	39	70	68	49	29	30	40
67	36	25	12	23	26	68	25
40	50	57	39	29	44	65	41

a. Ordene los datos en forma descendente.

b. Determine su rango.

c. Con intervalos de 5 datos, determine el número de intervalos.

d. Realice una tabulación, incluyendo el intervalo, el conteo y la frecuencia.

Intervalos de clase, población y muestra

Cuando se realiza una encuesta o una investigación cualquiera se obtienen datos, los cuales es necesario, en primer término, organizar para posteriormente analizarlos y así encontrar la información que se busca. Véase el siguiente ejemplo:

En un centro de salud se quiere saber el peso de los pacientes de un año que acuden a consulta para valorar el grado de nutrición de los niños en esta edad que habitan en esta comunidad.

Al cabo de una semana se atendieron 20 niños, los cuales obtuvieron los siguientes pesos en libras:

José 16.0	Francisco 14.8	Martha 11.0	Roberto 18.0	Susana 17.8
Pablo 20.4	Laura 19.0	Iris 18.0	Isabel 13.0	María 14.0
Manuel 16.6	Marvin 14.8	Carmen 17.0	Martín 18.8	Adrian 18.0
Luis 18.0	Alejandro 19.6	Lorena 17.0	Trinidad 18.0	Héctor 16.0

Como los datos están en desorden, para facilitar su estudio se deben ordenar en forma creciente o decreciente, en este caso en forma creciente:

11.0 13.0 14.0 14.8 14.8 16.0 16.0 16.6 17.0 17.0 17.8 18.0
18.0 18.0 18.0 18.0 18.8 19.0 19.6 20.4

Posteriormente, se registran los datos en una tabla o una tabulación, la cual consta de tres columnas:

1. Nombre de la variable estudiada (en este caso el peso): en esta columna se anotan los datos sin repetirlos y en forma ordenada.
2. Conteo de los datos: aquí se registra el número de veces que se repite cada dato con una marca o rayita.
3. Frecuencia: en esta columna se anota el número que representa a las rayitas obtenidas en el conteo de los datos.

Peso	**Conteo**	**Frecuencia**
11.0	I	1
13.0	I	1
14.0	I	1
14.8	II	2
16.0	II	2
16.6	I	1
17.0	II	2
17.8	I	1
18.0	~~IIII~~	5
18.8	I	1
19.0	I	1
19.6	I	1
20.4	I	1
Total		20

Una vez hecha la tabulación, se observa que el peso mayor es de 20.4 libras y el menor es de 11.0 libras. De estos dos datos se obtiene el rango, el cual resulta de la diferencia entre ellos:

$$Rg = 20.4 - 11.0 = 9.4$$

El rango sirve para calcular los intervalos de clase, como se verá a continuación. Los intervalos son datos agrupados de acuerdo con la amplitud de intervalo que se elija. Para determinar el número de intervalos se divide el rango entre la amplitud del intervalo elegido.

En el ejemplo anterior se requiere una amplitud de intervalo de 2 libras, pues cada dos libras de peso es significativo en los niños de un año. Así, para obtener el número de clases o intervalos, se realiza la operación siguiente:

$$\text{Número de intervalos} = \frac{\text{Rango}}{\text{Amplitud del intervalo}} = \frac{9.4}{2} = 4.7 = 5 \text{ intervalos}$$

En este caso, el cociente de la división es 4.7, por lo cual se redondeó al entero más próximo, que es 5.

A continuación se definen los intervalos o clases tomando los límites inferior y superior considerando la amplitud del intervalo elegida. Por ejemplo, en el primer intervalo el límite inferior es 11.0 y se le suman 2, obteniéndose 13, que es el límite superior.

Posteriormente se determina el número de casos que están dentro de cada intervalo, es decir, las frecuencias de clase. Los datos agrupados se presentan en una tabla de frecuencia:

Peso (intervalos)	Frecuencia
11.0-13.0	2
13.1-15.1	3
15.2-17.2	5
17.3-19.3	8
19.4-21.4	2
Total	20

Las tablas de frecuencia facilitan el análisis de los resultados de una investigación, ya que permiten establecer relaciones entre los datos.

En el ejemplo anterior, los médicos observaron que la mayoría de los pacientes de un año está bien alimentados, pues se encontraron dentro del intervalos de 13 a 19 libras de peso; no obstante, 2 están bajo la línea de peso normal y 2 tienen sobrepeso.

En muchas ocasiones es necesario realizar investigaciones que estudien las características o valores de una población determinada, pero debido a las limitaciones de tiempo o de recursos no se trabaja con la totalidad de la población, sino con una parte de esta.

Por ejemplo, si se desea saber las preferencias deportivas de todos los alumnos de los centros de educación que actualmente están estudiando con Educatodos, entrevistar a todos los alumnos implicaría un gasto excesivo en tiempo y dinero y el análisis de los resultados sería dificultoso, porque la población a investigar es demasiado grande y está dispersa en todo el territorio nacional. En situaciones como esta es necesario utilizar un método estadístico llamado muestreo, que consiste en seleccionar una muestra que represente a toda la población; este recurso permite que hace la investigación sea menos costosa y que se realice en menor tiempo.

Se llama población al grupo o conjunto de individuos, características u objetos que se examinan. En el ejemplo anterior, la población serían todos los alumnos de Educatodos y debido a que es muy difícil examinar toda la población se elige una muestra, que es una pequeña parte de la población que se toma como representativa del conjunto.

Para tener validez, la muestra elegida debe cumplir con los siguientes requisitos:

a. Debe representar a la población, es decir, ha de pertenecer a esta y ser elegida al azar o en forma aleatoria, para que todos los elementos de la población tengan la misma probabilidad de ser considerados.

b. Debe ser confiable, o sea, que los resultados que se obtengan se puedan generalizar a toda la población con cierto grado de precisión.

c. Que sea práctica, es decir, que sea sencilla de llevar a cabo.

d. Que sea eficiente, debe proporcionar la mayor información a menor costo.

Existe un procedimiento para elegir una muestra al azar, este es llamado método de muestreo aleatorio, el cual se explica detalladamente en el siguiente ejemplo.

En una maquila trabajan 4500 obreros y se desea tomar una muestra de 45 personas. El primer paso es enumerarlos atendiendo al número de cifras que tiene la población total, como en este caso se tienen 4 cifras, el primer elemento se enumera con 0001.

Posteriormente se colocan en una urna 10 bolitas enumeradas del 0 al 9 y cada vez que se elija un obrero se extrae una esfera, se anota el dígito y se regresa a la urna, se revuelven las esferas y se repite el procedimiento tres veces más hasta obtener un número de cuatro dígitos. De esta forma todos los obreros tienen la misma posibilidad de aparecer en la muestra.

Se preguntará qué pasará cuando el número escogido de cuatro dígitos sea mayor que el de la población, por ejemplo si se obtiene 6456, en este caso se divide por el último número (4500) y se toma el residuo:

$$\begin{array}{r|l} 6456 & \underline{4500} \\ \underline{4500} & 1 \\ 1956 & \end{array}$$

En este caso se elegirá 1956, de esta forma no se desperdician los números mayores a la población.

Pero, de qué tamaño debe ser la muestra. Esta varía dependiendo del tamaño de la población y de las características propias de la investigación, pero en general se toma se toma el 1 % de la población para formar la muestra. El tamaño de la población se denota con la letra N y el de la muestra con la letra n.

Actividad 3

1. Responda en su cuaderno las siguientes interrogantes:

   a. ¿En qué casos se aplica el método del muestreo?
   b. ¿Cuáles son las ventajas de este método?
   c. ¿A qué se le llama población?
   d. ¿Qué es una muestra?
   e. ¿Qué requisitos debe cumplir una muestra para tener validez?
   f. ¿Cómo se denota el tamaño de una población? ¿Y el de la muestra?

2. Obtenga la población y la muestra de cada uno de los siguientes casos:
   a. Una empresa que fabrica 10,000 focos a la semana realiza un control de calidad de sus productos, probando 1,000 focos en el mismo lapso de tiempo.
   b. Se desea saber el porcentaje de calcetines defectuosos producidos en 5 días. Diariamente se producen 1,000 calcetines y se pretende examinar 10 calcetines diferentes por día en diferentes horas del día.
   c. Anote los datos de alguna posible investigación que se pueda hacer en su comunidad y defina la población y la muestra.
   d. Cómo tomaría una muestra de 20 alumnos del centro básico en que estudia, tomando en cuenta del 1º al 9º para realizar un estudio de las asignaturas preferidas.

Representación gráfica de los datos

Para lograr una mejor comprensión de los números arreglados en forma tabular, se utilizan los gráficos, los que destacan algunos hechos más claramente. Un gráfico para ser de utilidad real, debe ser simple y poner mayor énfasis en los rasgos significativos de los datos.

Un gráfico estadístico es la representación de un fenómeno por medio de figuras geométricas (puntos, líneas, rectángulos, círculos, etc.). Entre estos están los siguientes: barras simples, barras comparativas y gráfico circular, entre otros.

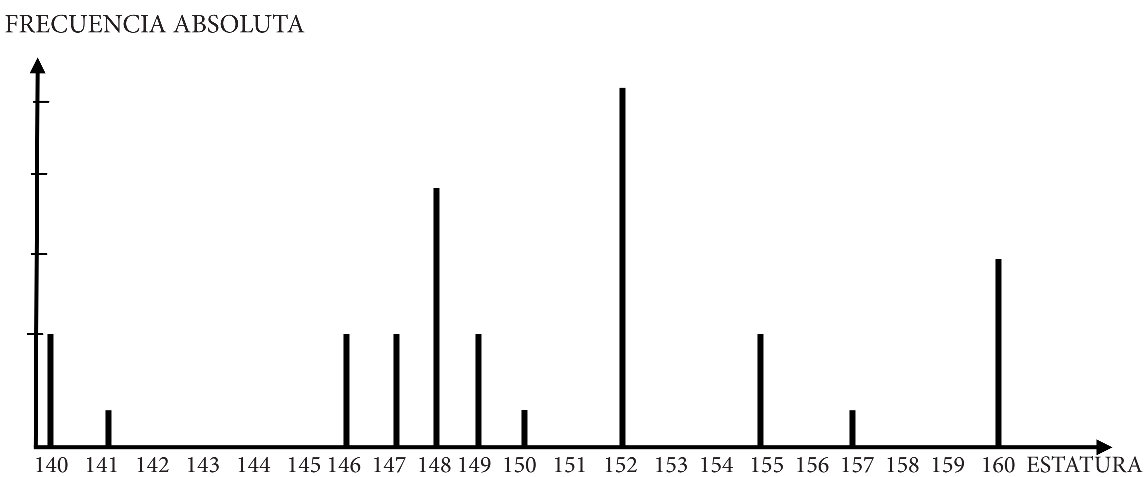
Gráfico de frecuencia absoluta

Cuando los datos estadísticos han sido registrados en una tabla, pueden representarse por medio de una gráfica. Para precisar esta idea considere el siguiente ejemplo:

Se han tabulado los datos referentes a la estatura en centímetros de los miembros de un equipo de futbol de un centro básico, integrado por 25 jugadores (titulares y reservas):

Estaturas (cm)	Conteo	Frecuencia absoluta
140	II	2
141	I	1
146	II	2
147	II	2
148	IIII	4
149	II	2
150	I	1
152	~~IIII~~	5
155	II	2
157	I	1
160	III	3
Total		25

Para elaborar la gráfica representativa de los datos anotados en la tabla, se trazan dos ejes, uno horizontal y otro vertical, que sean perpendiculares. Por lo general en el eje horizontal se anotan los datos (estaturas) y en el vertical las frecuencias.

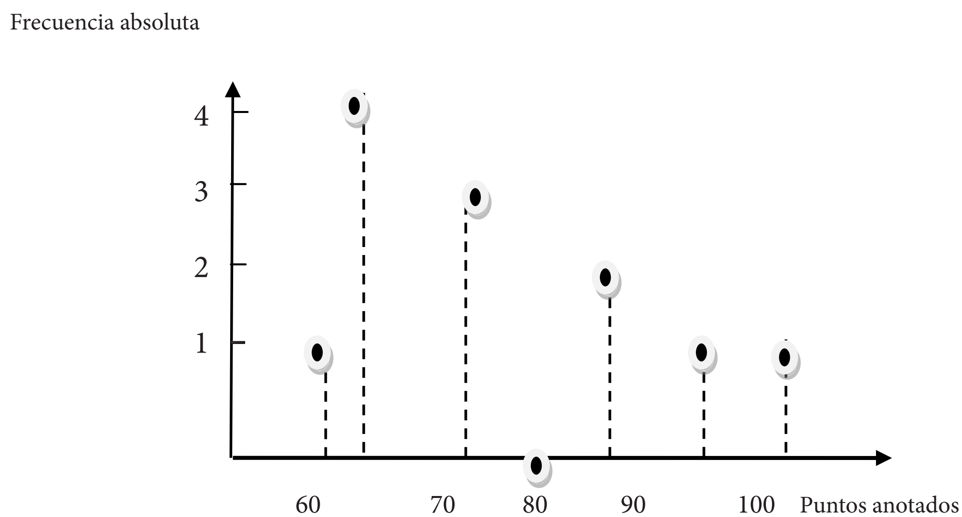


<u>Otro ejemplo</u>

El entrenador de un equipo de basquetbol ha registrado el número de puntos que cada uno de sus jugadores anotó en total durante la temporada que recientemente terminó y dicha información la concentró en la siguiente tabla:

Puntos anotados	Conteo	Frecuencia absoluta
62	I	1
64	IIII	4
72	III	3
79		
85	II	2
93	I	1
97	I	1
Total		12

La gráfica representativa de esta tabla se ve así:



Seguramente, el análisis de estas gráficas y otras que puedan elaborar los entrenadores, con características personales y del rendimiento de sus jugadores en la práctica de un deporte, permitirán optimizar el funcionamiento de los equipos a su cargo.

Conocer cómo se realiza la representación gráfica de las frecuencias absolutas, sirve de base para comprender cómo se elaboran e interpretan otras gráficas más específicas en cuanto a cualidades y cantidades.

Actividad 4

Realice en su cuaderno lo que se le pide a continuación:

Complete las tablas y elabore las gráficas que representen la frecuencia absoluta, correspondientes a cada una de las tablas que se dan a continuación:

Visitas	Conteo	Frecuencia absoluta
Museos	IIII	
Parques arqueológicos	III	
Empresas	III	
Zoológicos	I	
Talleres	II	
Oficinas públicas	IIII	
Total		

Peso (lbs.)	Conteo	Frecuencia absoluta
37.5	~~IIII~~	
40	~~IIII~~ I	
40.5	III	
41	~~IIII~~ III	
46	II	
51	IIII	
52.5	I	
Total		

Frecuencia relativa

Cuando se tiene una serie de datos debidamente ordenada y tabulada, no solamente es necesario conocer la frecuencia absoluta de los valores que se incluyen, o sea el número de veces que un dato aparece en el total considerado, sino que se requiere saber también cuál es la frecuencia relativa, es decir, el tanto por ciento de la aparición de ese dato con relación al conjunto de los datos.

Para que se aprecie el procedimiento que se sigue cuando es necesario obtener una frecuencia relativa, considere el siguiente ejemplo:

En un centro de salud de Comayagua se atendió durante la semana pasada a cierto número de pacientes con síntomas de diversos padecimientos, los cuales se enumeraron en la tabla que aparece a continuación:

Motivo de consulta	Conteo	Frecuencia absoluta
Gripe	~~IIII~~ IIII	9
Tos	~~IIII~~ I	6
Herida	IIII	4
Quemadura	~~IIII~~	5
Hepatitis	III	3
Diabetes	~~IIII~~ III	8
Dolor estómago	~~IIII~~ I	6
Quebradura	~~IIII~~ IIII	9
Total		50

La frecuencia relativa es un dato que se obtiene dividiendo la frecuencia absoluta entre el total y multiplicando el cociente por cien. Es decir:

$$\text{Frecuencia relativa} = \frac{\text{Frecuencia absoluta}}{\text{suma o total de las frcuencias absolutas}} \text{ x } 100$$

Realizando las operaciones según los datos de la tabla anterior, se obtiene:

$\frac{9}{50}$ (100) = 0.18 (100) = 18

$\frac{6}{50}$ (100) = 0.12 (100) = 12

$\frac{4}{50}$ (100) = 0.08 (100) = 8

$\frac{5}{50}$ (100) = 0.10 x 100 = 10

$\frac{3}{50}$ x 100 = 0.06 x 100 = 6

$\frac{8}{50}$ x 100 = 0.16 x 100 = 16

$\frac{6}{50}$ x 100 = 12

$\frac{9}{50}$ x 100 = 18

Estos resultados son tantos porcientos y la suma de ellos debe ser 100 %, ya que en la fórmula interviene como cociente 50, que es el total de los valores registrados en la tabla. Al representar la frecuencia con un tanto por ciento, se está haciendo referencia a la frecuencia relativa. A continuación se presenta una tabla en la cual se incluyen las frecuencias relativas.

Motivo de consulta	Conteo	Frecuencia absoluta	Frecuencia relativa (%)
Gripe	~~IIII~~ IIII	9	18
Tos	~~IIII~~ I	6	12
Herida	IIII	4	8
Quemadura	~~IIII~~	5	10
Hepatitis	III	3	6
Diabetes	~~IIII~~ III	8	16
Dolor estómago	~~IIII~~ I	6	12
Quebradura	~~IIII~~ IIII	9	18
Total		50	100 %

Es posible concluir, por tanto, que el 16 % de las consultas se dio a enfermos de diabetes y 12 % de los pacientes con dolor de estómago, etcétera.

Actividad 5

Complete la siguiente tabla anotando las frecuencias absoluta y relativa en las columnas correspondientes. Puede usar calculadora si lo desea.

Gastos	Conteo	Frecuencia absoluta	Frecuencia relativa (%)
Mantenimiento edificio	~~IIII~~ I		
Mantenimiento jardín	IIII		
Útiles de aseo	~~IIII~~ III		
Bonos	~~IIII~~ ~~IIII~~ III		
Total			

Gráfica de barras

Los datos pueden ser de dos tipos: cualitativos y cuantitativos. Los primeros hacen referencia a cualidad y los segundos a números o cantidades. Los datos cualitativos son medidas de características, de rasgos, de cualidades asociadas con la unidad de observación.

Una gráfica de barras es una representación gráfica de una tabla de frecuencias para datos cualitativos. Por lo general, en el eje horizontal se disponen los datos cualitativos y en el eje vertical las frecuencias de los datos. La elaboración de una gráfica es sencilla y para ello se debe tomar en cuenta lo siguiente:

- La línea base es una línea horizontal o vertical en el cual descansan o inician las barras; sirve para establecer comparaciones entre los datos cualitativos, representados por las barras, con una simple y rápida inspección.
- El ancho de las barras, todas las barras o rectángulos de un gráfico, tendrán la misma medida de ancho, siendo este arbitrario. El ancho de la barra generalmente depende del número de datos a representar con relación al espacio disponible para la construcción del gráfico.
- La separación entre las barras, el espacio entre barras o rectángulos, no debe ser menor que la mitad del ancho de la barra, ni mayor que el ancho de la misma, manteniendo siempre la misma distancia.
- Toda gráfica debe ir acompañada de su tabla de datos, con la suma de sus totales.

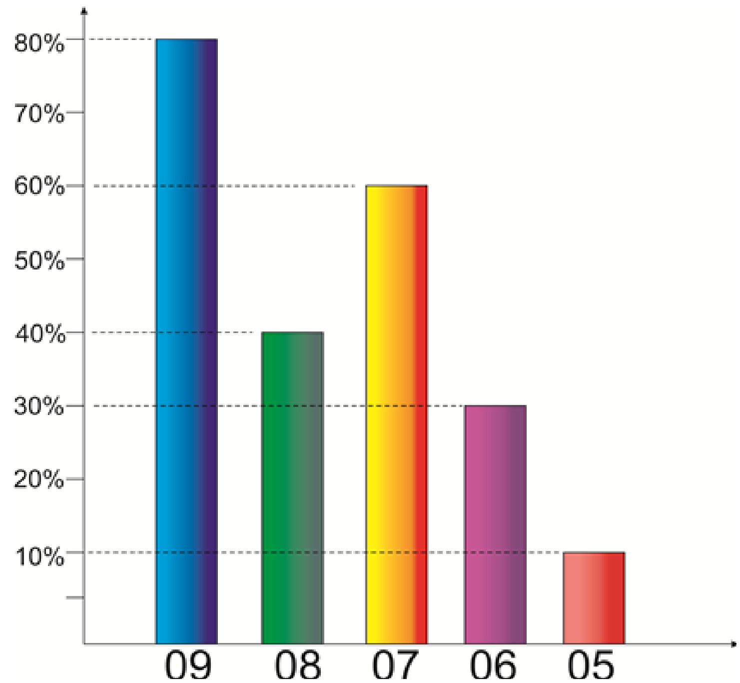
Ejemplo

El director de un centro básico realizó una encuesta para conocer las preferencias de los estudiantes de séptimo grado, con relación a las asignaturas que cursan. Los datos obtenidos son los siguientes:

Asignatura	Frecuencia absoluta
Matemáticas	(M) 9
Español	(E) 4
Educación Artística	(EA) 4
Ciencias Sociales	(CS) 8
Inglés	(ING) 4
Educación Física	(EF) 2
Ciencias Naturales	(CN) 11
Total	42

Para construir la gráfica de barras debe seguirse el siguiente procedimiento:

1. Se trazan dos ejes perpendiculares.
2. Se coloca la escala de valores o frecuencias sobre el eje vertical y los datos cualitativos en el eje horizontal.
3. Se trazan los rectángulos o barras del mismo ancho sobre el eje de los datos cualitativos, dejando un mismo espacio entre ellos. La longitud de cada barra representa el número de frecuencias.



Al analizar la gráfica de los 45 alumnos entrevistados, se observa que:

1. 9 alumnos prefieren Matemáticas, 4 prefieren Español, 4 prefieren Educación Artística, etcétera.

2. La asignatura que más prefieren los alumnos es Ciencias Naturales.

3. Las asignaturas que menos prefieren los alumnos son: Español, Inglés y Educación Artística.

4. Los alumnos que prefieren otra materia que no sea Matemáticas, son 36.

5. Los alumnos que prefieren Inglés más que otras materias son 4.

6. Al preguntar a un alumno sobre la materia de su preferencia, lo más probable es que conteste Ciencias Naturales y las menos probables son: Español, Inglés y Educación Artística.

La gráfica de barras es una forma objetiva de presentar los datos en estudio de un problema estadístico.

Actividad 6

Con base en la siguiente tabla, elabore una gráfica de barras y luego conteste las preguntas con relación a la gráfica:

Sabores de helados que prefiere un grupo de escolares de primer grado:

Sabor	Frecuencia
Chicle	13
Banano	5
Coco	10
Sandía	5
Chocolate	10
Fresa	2
Total	45

1. ¿Cuál sabor es el que gusta más?
2. ¿Cuál es el sabor que gusta menos?
3. Si usted vendiera helados, ¿de cuáles sabores tendría más?
4. Si llegara un grupo de escolares a comprar helados, ¿de cuál sabor es probable que venda menos?

Actividad 7

Escriba los datos de la siguiente investigación en una tabla y posteriormente elabore la gráfica de barras correspondiente:

Las temperaturas máximas registradas para el mes de abril durante 6 años en Tegucigalpa fueron de 29 °C en 1988, 28 °C en 1999, 29 °C en el 2000, 30 °C en el 2001, 32 °C en el 2002 y 34 °C en el 2003.

Gráfica de barras comparativas

Es muy común que se tengan que realizar comparaciones de las investigaciones realizadas, así como la presentación de los datos ya sea en forma tabular o gráfica, para la toma de decisiones o para formular planes a corto y largo plazo. Para concretar esta idea considere el siguiente caso:

Un médico realizó una investigación acerca de la libreta de vacunación en dos centros básicos de la misma comunidad y elaboró una tabla comparativa en la que registró los datos de los alumnos que carecen de algunas vacunas.

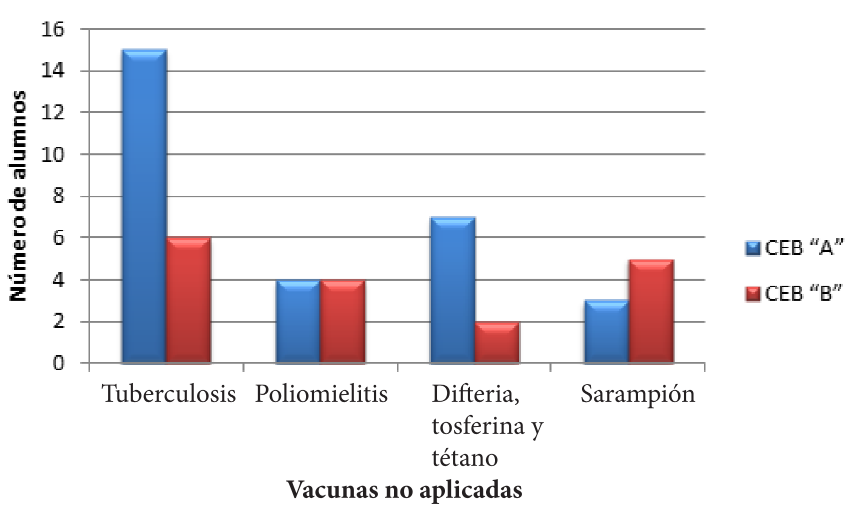
Resulta, además, que en ese lugar no existe un centro de salud y las vacunas se aplican cuando una brigada médica recorre diversas poblaciones de la región.

Vacunas no aplicadas	Centro A	Centro B
Tuberculosis	15	6
Poliomielitis	4	4
Difteria, tosferina y tétano	7	2
Sarampión	3	5
Total	28	18

Por medio de la tabla anterior se pueden determinar algunos aspectos como los siguientes:

1. El mayor número de alumnos que no han sido vacunados contra la tuberculosis se encuentra en el centro A.
2. El menor número de alumnos que no han sido vacunados contra difteria, tosferina y tétano se encuentra en el centro B.
3. Existe igual número de alumnos que no han sido vacunados en los centros A y B en lo que respecta a las vacunas de poliomielitis.

La representación gráfica de los datos de la investigación es la siguiente:



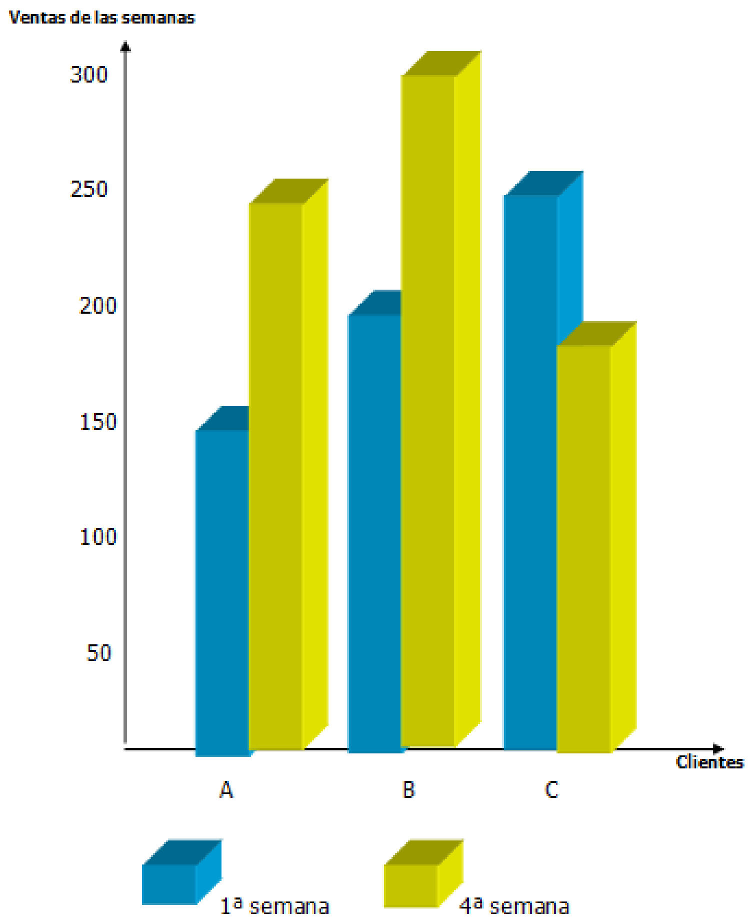
Esta información comparativa servirá de base para solicitar una campaña de vacunación y lograr que los alumnos que carecen de alguna vacuna sean inmunizados.

Así como en el ejemplo anterior se puede llegar a tomar una decisión con base en las tablas y gráficas elaboradas, existen casos en los cuales es necesario llevar un control en cuanto a compras o ventas. Véase el siguiente ejemplo:

Una empresa vende confites por caja y se necesita saber en qué semana, 1ª o 4ª del mes, se venden más cajas y cuál es el mejor cliente que se tiene, para este efecto se ha elaborado la siguiente tabla comparativa:

Clientes	Semanas	
	1ª	4ª
A	150	250
B	200	300
C	250	175

Una vez realizada la tabla, se procede a representar los datos en una gráfica de barras comparativas:

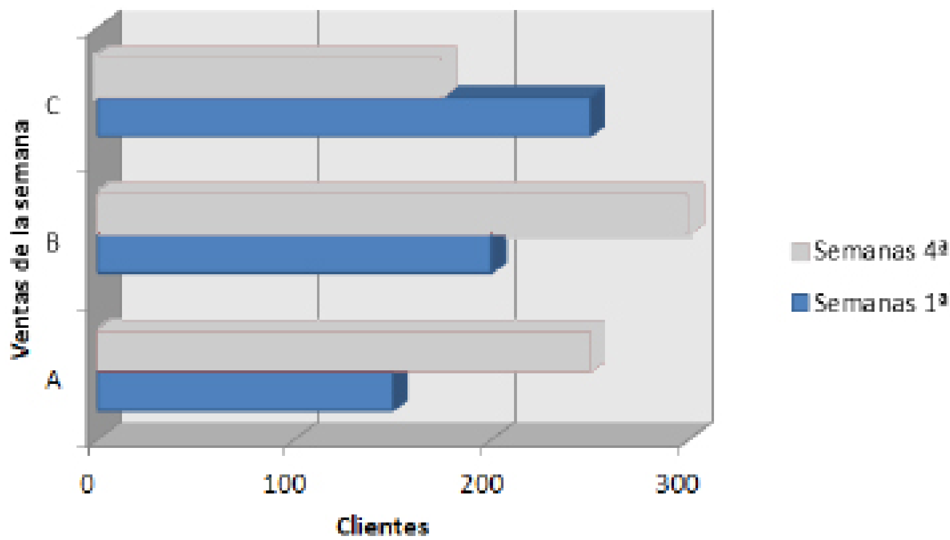


Como puede observarse, estas tablas y gráficas se hacen con la finalidad de conocer el volumen de ventas y así poder aumentarlas.

La ventaja de comparar tablas y gráficas radica en que se pueden tomar decisiones que sirvan para el desarrollo de medidas preventivas o correctivas en múltiples situaciones de importancia económica, política o social. Otra ventaja es que permite al lector apreciar y valorar la información de una manera clara y sencilla.

Por lo general, en el eje horizontal se disponen los datos cualitativos y en el eje vertical las frecuencias de los datos, pero también los datos se pueden ubicar de forma contraria.

Observe como quedaría la gráfica del ejemplo anterior al ubicar los datos cualitativos en el eje vertical y las frecuencias de los datos en el eje horizontal:



Actividad 8

Conteste en su cuaderno las siguientes interrogantes:

a. ¿Considera importante realizar investigaciones para posteriormente presentar los datos en forma tabular y gráfica? ¿Por qué?

b. ¿Realizar comparaciones de la información obtenida en una investigación le permite tomar decisiones? ¿Cómo lo ejemplificaría?

c. Elabore en su cuaderno dos gráficas comparativas de las ventas de prendas de vestir, realizadas en los meses de agosto y septiembre, por una tienda de ropa. Una gráfica con los datos cualitativos en el eje horizontal y los datos de las frecuencias en el eje vertical y la otra con los mismos datos, pero intercambiando los ejes.

VENTAS DE ROPA EN LOS MESES DE AGOSTO Y SEPTIEMBRE

Prendas de vestir	Agosto	Septiembre
Fajas	15	10
Pantalones	20	17
Camisas	30	8
Corbatas	9	2
Total	74	37

d. Escriba en su cuaderno dos conclusiones derivadas de las gráficas.

Gráfica circular o diagrama de sectores

Una gráfica circular es otro ingenioso medio de presentar un resumen visual de frecuencias de datos cualitativos, resulta muy útil para representar una distribución de frecuencias relativas.

En la práctica es frecuente encontrar situaciones o fenómenos estadísticos que hacen relación a la subdivisión de un total en sus partes componentes y porcentajes que cada una de ellas representa. Para representar tales situaciones se emplea la gráfica circular o de sectores.

El círculo completo representa el total todos los datos, es decir, el 100 %, al cual corresponden los 360º.

El procedimiento para representar datos en este tipo de gráficas se muestra con el siguiente ejemplo:

De 450 personas, 125 hablan, inglés; 100, francés; 75, alemán y el resto español. En este caso, lo primero que hay que hacer es averiguar la cantidad de personas que hablan español y luego construir una tabla con estos datos.

125 hablan inglés

100 hablan francés

\+ 75 hablan alemán

300

450 – 300 = 150

150 personas hablan español

Se tabulan los datos en una tabla de frecuencias absolutas:

Idioma	Frecuencia absoluta (personas)
Inglés	125
Francés	100
Alemán	75
Español	150
Total	450

Luego, se encuentran las frecuencias relativas de cada uno de los datos:

Personas que hablan inglés:

$\frac{125}{450}$ x 100 = 27.7 = 27.8 %

Personas que hablan francés:

$\frac{100}{450}$ x 100 = 22.2 %

Personas que hablan alemán:

$\frac{75}{450}$ x 100 = 16.6 = 16.7 %

Personas que hablan español:

$\frac{150}{450}$ x 100 = 33.3 %

Para hacer la gráfica circular se convierten los porcentajes a grados con la relación 1 % = 3.6º utilizando las proporciones, es decir:

Para 27.8 %

1 %: 3.6º = 27.8 %: x

En esta proporción la incógnita es un extremo, por lo tanto:

$X = \frac{3.6º \text{ x } 27.8\%}{1\%} = 100.08º \cong 100º$

En todas las proporciones el cociente será 1 %, por lo que para convertir el porcentaje a grados, solo es necesario efectuar la multiplicación entre ellos.

Para 22.2 %

22.2 x 3.6º = 79.92º ≅ 80º

Para 16.7 %

$16.7 \times 3.6^{o} = 60.12^{o} \cong 60^{o}$

Para 33.3 %

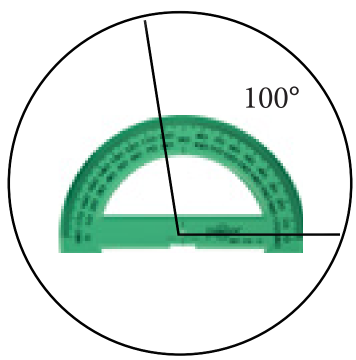
$33.3 \times 3.6^{o} = 119.88^{o} \cong 120^{o}$

Estos datos se organizan en una tabla, de la siguiente manera:

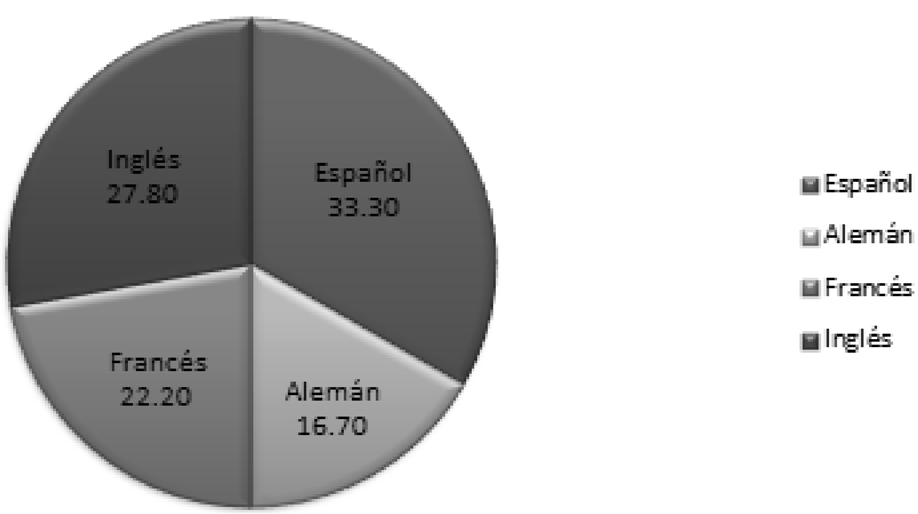
Idioma	Frecuencia absoluta (personas)	Frecuencia relativa (%)	Grados (º)
Inglés	125	27.8	100
Francés	100	22.2	80
Alemán	75	16.7	60
Español	150	33.3	120
Total	450	100.0	360

Observe que la suma de los grados que representan a las frecuencias relativas de los datos es 360º.

Estos datos se representan en un círculo de radio cualquiera, con la ayuda del transportador.



Quedando finalmente así:



Notas importantes:

1. Al 100 % del área del círculo le corresponden 360º por tanto, al 1 % le corresponde 3.6º.
2. Toda cantidad parcial debe expresarse en porcentaje.
3. Asignar a cada porcentaje parcial un sector circular de acuerdo al ángulo correspondiente a dicho porcentaje.
4. Usar el compás y el transportador para hacer el círculo y dibujar los ángulos respectivamente obtenidos.
5. Si no desea sombrear, puntear o rayar dentro del círculo, se sugiere escribir los indicadores en la parte superior o inferior derecha del gráfico.
6. La suma de los grados debe cuadrar a 360º y la de los porcentajes a 100 %.

Actividad 9

Realice los ejercicios que se le presenta a continuación:

Construya una gráfica circular de la siguiente tabla de datos obtenidos por un estudio de la clasificación de la red vial oficial:

Tipo de carretera	Kilómetros
Pavimentada	2845
Transitable todo el tiempo	9357
Transitable solo en verano	1484
Total	13,686

Actividad 10

Elabore una gráfica circular con la siguiente información y al terminar escriba 3 conclusiones sobre la gráfica:

Tipos de comida preferidas en el aula de clase: montucas, 13; enchiladas, 10; mondongo, 8; yuca con chicharrón, 5; nacatamales, 16; chop suey, 6.

Histogramas

Los histogramas son representaciones gráficas de datos cuando estos son cuantitativos, es decir, que hacen referencia a números o cantidades. Por ejemplo, el número de personas que integran una familia, las calificaciones obtenidas en un examen por un grupo de niños, la estatura de un cierto número de personas, etcétera.

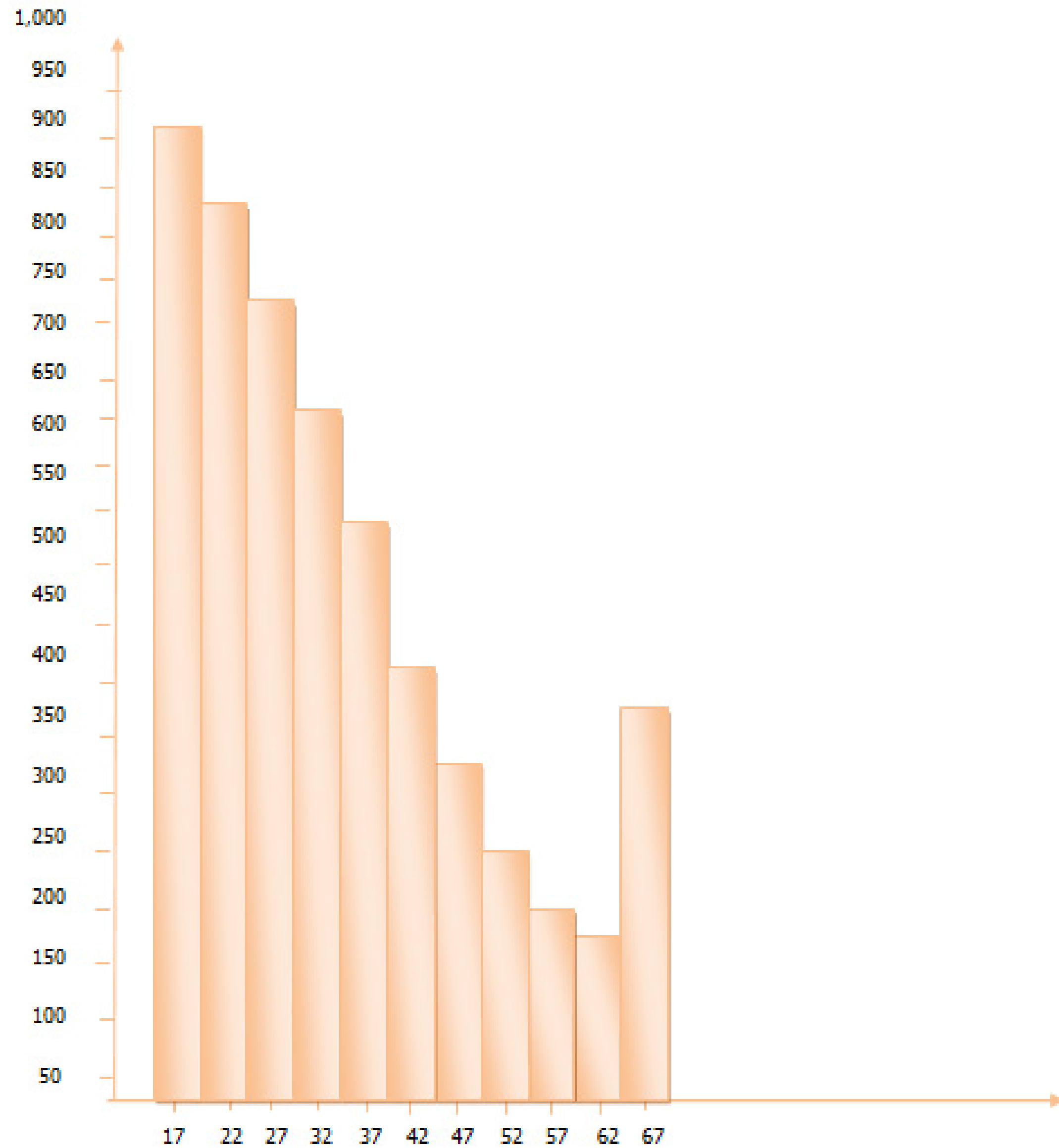
Para realizar un histograma es necesario agrupar los datos en intervalos, señalar la frecuencia y el punto medio que se obtiene sumando los límites del intervalo y dividiendo entre dos; asimismo, emplear dos líneas perpendiculares, un eje horizontal en donde se señalará la amplitud del intervalo y un eje vertical en el que se presentarán las frecuencias.

Ejemplo

En el censo de 1990, en un municipio de Honduras, se obtuvieron los siguientes datos sobre el número de personas que pueden leer y escribir, según su edad:

Intervalos	Frecuencia	Punto medio
15 - 19	966	17
20 - 24	886	22
25 - 29	765	27
30 - 34	645	32
35 - 39	528	37
40 - 44	400	42
45 - 49	316	47
50 - 54	253	52
55 - 59	202	57
60 - 64	169	62
65 - 69	400	67

Obsérvese el histograma que representa esos datos:



Como puede verse, la base de cada rectángulo corresponde a cinco unidades, es decir, a la amplitud del intervalo, y su altura equivale a la frecuencia de los datos.

Para determinar la longitud de la base en los rectángulos se señala la mitad del intervalo a la izquierda del punto medio y la otra mitad a la derecha del mismo. Por ejemplo, en el caso anterior se tiene que uno de los puntos medios es 17, entonces se hizo una marca de 2.5 unidades a derecha e izquierda del 17 y así con los demás puntos medios, por lo cual los rectángulos quedan contiguos.

A las marcas que señalan el inicio y el final de la base en cada rectángulo se les conoce como límites reales. Además, cabe mencionar que el inicio del eje horizontal empieza en 17, indica que en la distancia del original primer dato existen valores no señalados por convenir así para el trazo de la gráfica. Así, puede observarse que el histograma es una forma más clara de presentar los datos para su análisis.

Polígono de frecuencias

Cuando los datos a graficar son numéricos, se dice que son cuantitativos y un medio para expresarlos es la gráfica poligonal o polígono de frecuencias.

Una vez que se ha elaborado el histograma, el polígono de frecuencias es fácil de trazar; para ello, véase el siguiente planteamiento:

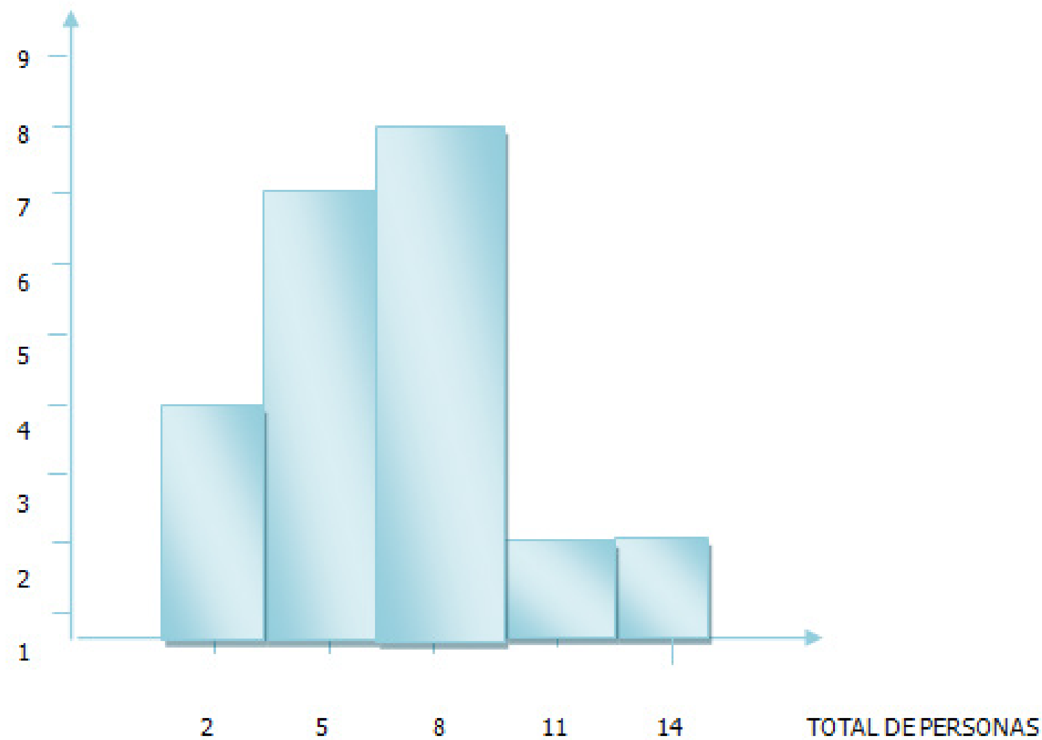
Al entrevistar a estudiantes de un grupo de 2° año de bachillerato y preguntarles cuál es el número de personas que viven en su casa, se obtuvieron los siguientes datos:

9	6	3	8	4	3	5
7	5	5	6	9	6	7
8	10	15	3	8	9	2

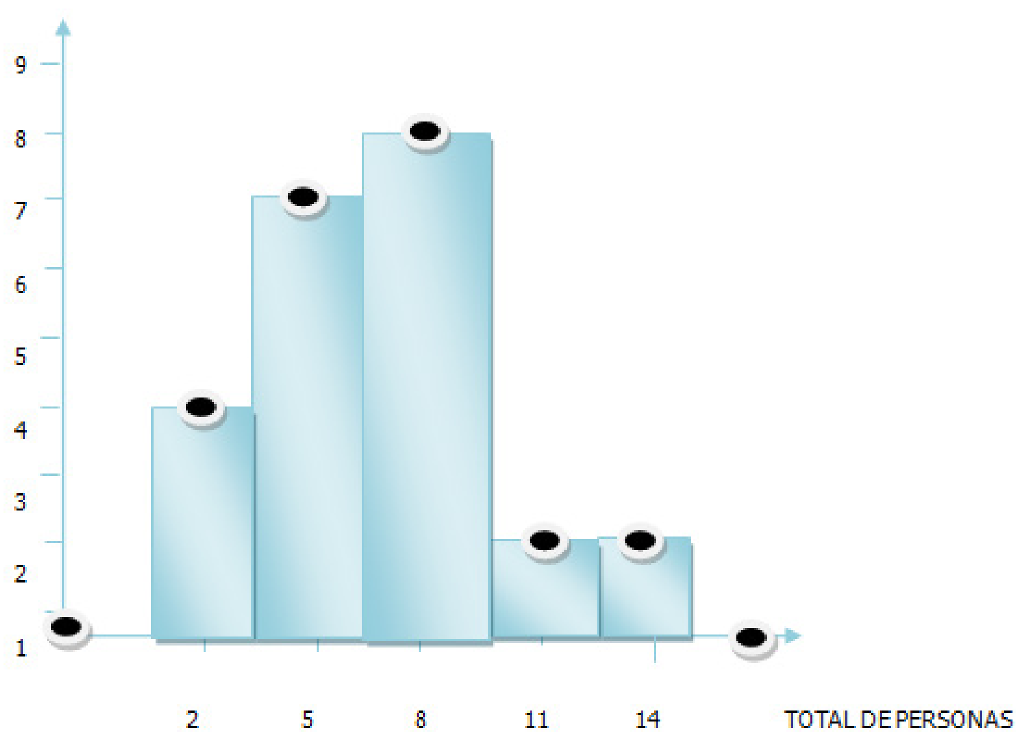
Con estos datos se realiza la tabulación correspondiente:

Intervalo	Frecuencia	Punto medio
1 - 3	4	2
4 - 6	7	5
7 - 9	8	8
10 – 12	1	11
13 - 15	1	14
Total	21	

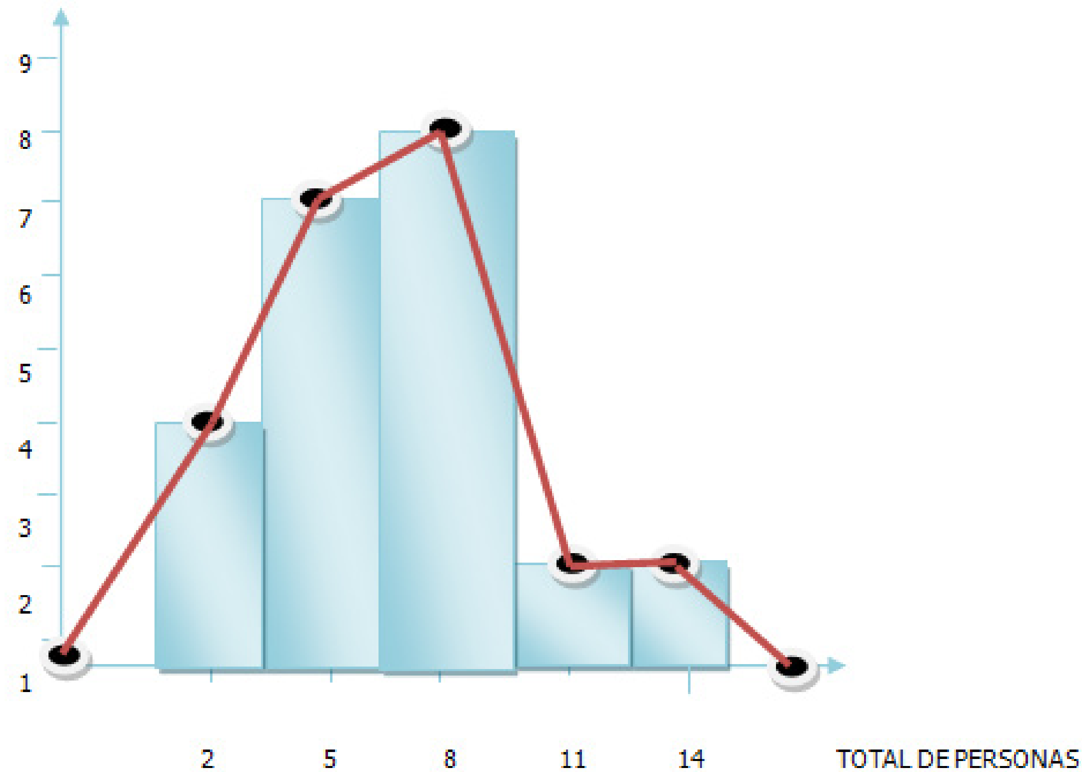
Con la tabulación es posible realizar el histograma:



Una vez trazado este, se localizan los puntos medios en la parte superior de cada rectángulo y se señala la medida de la mitad del intervalo a la izquierda de la primera barra del histograma y la mitad de la derecha del último, sobre el eje horizontal.



Y por último, se unen dichos puntos de donde se obtiene la poligonal:



Para trazar un polígono de frecuencias se unen con un poligonal los puntos medios de los techos de los rectángulos del histograma.

Actividad 11

Realice el ejercicio que se le presenta a continuación.

El peso de 37 estudiantes de un centro básico de Danlí está registrado en la siguiente tabla:

Peso en libras	Número de estudiantes
60 – 68	4
69 – 77	9
78 - 86	10
87 - 95	5
96 – 104	7
105 - 113	2
Total	37

Elabore un histograma y un polígono de frecuencias de la tabla anterior.

Sumatoria

La estadística generalmente es definida como la rama de las matemáticas que se ocupa de reunir, organizar y analizar datos numéricos para su presentación por medio de tablas y de representaciones gráficas y así ayudar a la toma de decisiones para resolver problemas. Por tal razón, es importante el manejo de datos numéricos.

En la estadística continuamente se están sumando datos diferentes y para abreviar esta operación se utiliza el símbolo Σ, que es la letra griega sigma, mayúscula, y se le denomina sumatoria.

Si se quiere expresar la suma de los cinco primeros números naturales, se tendría que escribir lo siguiente:

$$1 + 2 + 3 + 4 + 5 = 15$$

Si se utiliza el operador sumatoria, se escribiría así:

$$\sum_{i=1}^{5} x_i = 1 + 2 + 3 + 4 + 5 = 15$$

Esto se lee: la sumatoria de *i*, desde *i* = 1 a 5 de *x* sub *i* es igual a 15.

En general, cuando se escribe una expresión de la forma:

Esto se lee: sumatoria de *i*, desde *i = 1* hasta *n*, de *x sub-i*.

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + x_3 + \ldots + x_n$$

Indica que lo que se desea es la suma de todas las x_i (*x sub i*) desde *i=1* hasta *n*.

A partir de la expresión anterior, los siguientes ejemplos se pueden interpretar en lenguaje corriente, así:

$\sum_{i=1}^{4} x_i = x_1 + x_2 + x_3 + x_4$ sumar todos los valores consecutivos de x, desde 1 hasta 4

$\sum_{i=1}^{3} a_i^2 = a_1^2 + a_2^2 + a_3^2$ sumar todos los valores consecutivos de *a* al cuadrado, desde 1 a 3

Ejemplo 1

Si $x_1 = 2$, $x_2 = -5$, $x_3 = 3$, $x_4 = 1$ evaluar las siguientes sumatorias.

$\sum_{i=1}^{3} x_i$

Solución:

$x_1 + x_2 + x_3 = 2 + (-5) + 3 = 0$

$\sum_{i=1}^{3} x_i = 0$

Ejemplo 2

$\sum_{i=1}^{4} x_i^2$

$x_1^2 + x_2^2 + x_3^2 + x_4^2 = 2^2+(-5)^2+3^2+1^2 = 4 + 25 + 9 + 1 = 39$

$\sum_{i=1}^{4} x_i^2 = 39$

Actividad 12

Dados los siguientes valores:

$x_1 = 2$, $x_2 = 4$, $x_3 = 6$, $x_4 = 8$, $x_5 =10$

$y_1 = 3$, $y_2 = -2$, $y_3 = 5$, $y_4 = 1$

Calcule las siguientes sumatorias:

$\sum_{x=1}^{5} x_i =$

$\sum_{x=1}^{5} x_i^2 =$

$\sum_{y=1}^{4} y_i =$

$\sum_{y=1}^{4} y_i^2 =$

Medidas de tendencia central para datos no agrupados

Las medidas de tendencia central son valores numéricos que tienden a localizar el punto medio de un conjunto de datos. A menudo, se asocia el término promedio con este concepto, pero las medidas de tendencia central más comunes son tres: la media aritmética, la mediana y la moda.

Media aritmética

Los promedios tienen por objeto la representación de una serie de datos por un solo número. La condición fundamental para que el promedio pueda representar una serie es que exista una tendencia de los datos a concentrarse hacia un valor central. El promedio representa, en una primera aproximación, la tendencia central de la serie de datos.

Ejemplo

Las alturas en cm de 33 alumnos de un grado en un CEB de Olancho son:

153	154	159	162	157	156	160
161	160	163	161	164	158	158
151	161	154	164	158	158	158
158	163	158	160	161	161	
156	163	159	159	163	160	

Las alturas en cm de 29 alumnos del mismo grado en un CEB de El Paraíso son:

156	158	154	157	156	157
159	160	163	159	163	157
158	156	163	155	158	159
164	159	159	157	161	159
156	158	158	156	155	

¿Cuáles de los alumnos son más altos? ¿Los de Olancho o los de El Paraíso?

Para resolver este problema se procede de la siguiente manera: sumar las alturas de todos los alumnos de cada curso y dividir por el número de ellos. De este modo se obtiene la media aritmética de las alturas de los alumnos de dichos cursos y se denota con la letra *M*.

Para Olancho resulta: $M = \frac{5263}{33} = 159.4$ cm

Para El Paraíso resulta: $M = \frac{4590}{29} = 158.2$

La altura media del grado de Olancho es mayor que la de El Paraíso.

La media aritmética representa la tendencia central de la distribución.

A causa de su fácil cálculo y de su antiguo uso, la media aritmética es la más conocida y la más empleada en los promedios.

La media aritmética se obtiene sumando todos los datos y dividiendo la suma por el número de ellos.

Generalmente, la media aritmética es conocida como promedio de una colección de datos.
Para encontrarla en una colección de *n* números ($x_1, x_2, x_3 \dots\ x_n$), solo hay que sumar todos los números (esta idea se expresa como $\sum x$) y dividir la suma entre el número de datos *n*; lo anterior se expresa como:

Media aritmética = $\bar{x} = \dfrac{x_1 + x_2 + x_3 \dots + x_n}{n}$

$\bar{x} = \dfrac{\sum x}{n}$

$\bar{x}$ = media aritmética

$\sum x$ = sumatoria de datos

n = número de datos

Ejemplo

1. ¿Cuál es la media aritmética de los números 23, 19, 7, 32, 4 y 14?

   Los datos se ordenan de menor a mayor o viceversa, se emplea la fórmula, se sustituyen las variables por los valores numéricos y se realizan las operaciones indicadas.

   4, 7, 14, 19, 23, 32

   $n = 6$, ya que es el número de datos ordenados.

   $$\bar{x} = \frac{\sum x}{n} = \frac{4+7+14+19+23+32}{6} = \frac{99}{6} = 16.5$$

   $$\bar{x} = 16.5$$

   La media aritmética o promedio de dichos números es 16.5, lo cual indica que es un valor representativo de la serie de números dados.

   La desventaja de la media aritmética es que si alguno de los valores es extremadamente grande o extremadamente pequeño, la media no es el promedio apropiado para representar la serie de datos.

Actividad 13

Sin utilizar la fórmula (mentalmente) obtenga la media aritmética de los datos que se dan y posteriormente escríbala:

a. 4, 6, 6, 7, 15

b. 5, 5, 8, 12, 12, 12

c. 1, 2, 4, 4, 5, 5, 6, 7, 11

Actividad 14

Resuelva lo siguiente:

a. La duración en minutos de las llamadas de teléfono hechas en una oficina en cierto día fueron:

   3, 7, 10, 2, 4, 8, 11, 6, 5, 3, 2, 4, 8, 15, 10, 8, 7, 4, 9 y 14

   Calcule la duración media de las llamadas.

b. Juan obtuvo en matemáticas los siguientes resultados: I parcial, 75 %; II parcial, 80 %; III parcial, 60 % y IV parcial, 90 %. Calcule el promedio de las notas.

c. El peso de un grupo de niños del tercer grado es el siguiente: 75 libras, 80 libras, 78 libras, 88 libras, 86 libras. Calcule el peso promedio.

Mediana

Cuando una serie de datos contiene uno o dos valores muy grandes o muy pequeños, la media aritmética no es representativa. El valor central en tales problemas puede ser mejor descrito usando una medida de tendencia central llamada mediana.

Si se tiene una colección de datos y estos se ordenan en forma creciente o decreciente, se presentan dos casos para determinar la mediana:

a. Si el número de datos es impar, la mediana es el dato que queda justo a la mitad de la colección de ellas.

b. Si el número de datos es par, la mediana es la media aritmética de los datos que quedan a la mitad de la colección de ellas.

Con base en lo anterior, la mediana es el valor que divide a una colección de datos exactamente, de manera que queden la misma cantidad de datos arriba y abajo o a la derecha e izquierda de este valor.

Para determinar el lugar que ocupa la mediana dentro de la colección de datos, se aplica la siguiente relación:

$$\text{Lugar} = \frac{\text{número de datos} + 1}{2}$$

<u>Ejemplos</u>

Determinar la mediana en las siguientes series de números:

a. 11, 17, 18, 12, 3, 7, 6

Se ordena en forma creciente, en este caso: 3, 6, 7, 11, 12, 17, 18.

Para determinar el lugar en donde se encuentra la mediana se aplica la relación:

$$\text{Lugar} = \frac{\text{número de datos} + 1}{2} = \frac{7+1}{2} = \frac{8}{2} = 4^{\circ}\ \text{lugar}$$

Esto indica que la mediana es el valor del dato que ocupa el cuarto lugar.

Como el número de los datos es impar, esto indica que la mediana es el valor que se encuentra exactamente a la mitad de los datos, en este caso es 11 y a su vez ocupa el cuarto lugar en los datos, esto es: 3, 6, 7, <u>11</u>, 12, 17, 18.

b. 9, 7, 1, 4, 10, 2, 6, 11

Se ordenan los números de mayor a menor, en este caso: 11, 10, 9, 7, 6, 4, 2, 1

El lugar que ocupará la mediana será:

$$\text{Lugar} = \frac{\text{número de datos} + 1}{2} = \frac{8+1}{2} = \frac{9}{2} = 4.5$$

Esto indica que la mediana estará entre los números que ocupen el cuarto y el quinto lugar, para este caso, 7 y 6.

Como el número de datos es par, se obtiene la media aritmética de estos valores y esta a su vez será la mediana:

$$\bar{x} = \frac{7+6}{2} = \frac{13}{2} = 6.5$$

Incluyendo este valor en los datos iniciales, se tiene que:

11, 10, 9, 7, 6.5, 6, 4, 2, 1

Obsérvese que la mediana quedó exactamente a la mitad de los datos iniciales.

Propiedades para la mediana

1. Es única.
2. Es simple.
3. Los valores extremos no tienen efectos importantes sobre la mediana, lo que sí ocurre con la media.
4. La notación más usual que se utiliza para representar a la mediana es $\overline{X}$, *Md* o *Me*.

Actividad 15

1. Se tienen las siguientes edades tomadas de un grupo de 10 madres de familia de niños de primer grado de un centro básico y se desea conocer cuál sería su media y su mediana de las edades:

   25, 27, 35, 28, 30, 24, 25, 29, 32, 37
2. La altura de seis amigos es de 160 cm, 155 cm, 172 cm, 180 cm, 175 cm y 184 cm, ¿cuál es la mediana?
3. Hallar la mediana de las edades de los alumnos de 8º grado de un centro básico:

   12, 11, 13, 12, 11, 13, 14, 15, 12, 14, 12, 15, 13, 14, 11, 12, 12, 11, 14, 13, 14.

Moda

La moda es una medida muy natural para describir un conjunto de datos, su concepto se adquiere fácilmente: es la altura más corriente, es la velocidad más común, etc. Además tiene la ventaja de que no se ve afectada por la presencia de valores altos o bajos.

La moda es la medida que se relaciona con la frecuencia con que se presenta el dato o los datos con mayor incidencia, con lo que se considera la posibilidad de que exista más de una moda para un conjunto de datos. Las notaciones más frecuentes son las siguientes: Mo y $\hat{x}$. Esta medida se puede aparecer tanto para datos cualitativos como cuantitativos.

Se dice que cuando un conjunto de datos tiene una moda, la muestra es unimodal; cuando tiene dos modas, bimodal; cuando la muestra contiene más de un dato repetido se dice que es multimodal y un último caso es cuando ningún dato tiene una frecuencia, en dicho caso se dice que la muestra es amodal.

<u>Ejemplos</u>

1. Determinar la moda del siguiente conjunto de datos:

    a. 1, 2, 3, 3, 4, 5, 6, 7, 7, 3, 1, 9, 3

    La moda de este conjunto de datos es igual a 3 y se considera unimodal.

    b. 1, 2, 3, 4, 4, 5, 2, 1, 3, 4, 2, -3, 4, 6, 3, 3

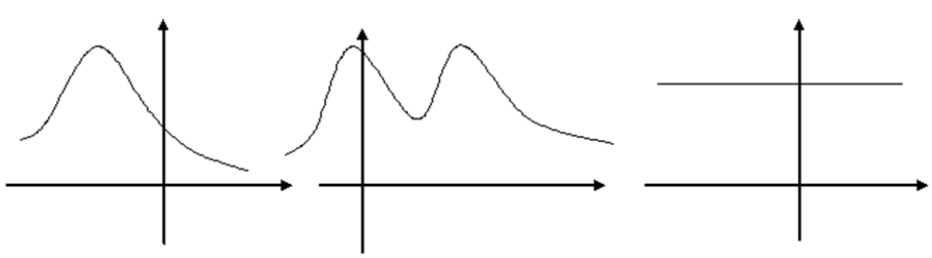
    Las modas de este conjunto de datos son 3 y 4, ya que ambas tienen la más alta frecuencia, por lo que la muestra es bimodal.

    c. - 1, 2, 3, 4, 5, 6, 7, 8, 9

    La muestra no contiene ningún dato repetido, por lo que se considera que la muestra es amodal.

    Gráficamente eso se puede reflejar mediante el análisis de un histograma de frecuencias:

a. Unimodal b. Bimodal c. Amodal



Grafica de 3 distribuciones diferentes a. Unimodal, b. Bimodal y c. Amodal

De lo anterior se concluye que se requiere un número suficiente de observaciones para que se manifieste o se defina claramente el tipo de moda de la muestra.

Actividad 16

Realice una encuesta con los alumnos de séptimo, octavo y noveno grados de su centro de estudio sobre la edad. Con estos datos determine la media, la mediana y la moda de cada grado.

Medidas de dispersión para datos no agrupados

Para comprender el concepto de varianza, supóngase que tenemos los datos siguientes, de los cuales queremos saber qué tan dispersos están con respecto a su media:

2, 3, 4, 5, 6 con media = 20/5 = 4

Si se toma la suma de diferencias de cada valor respecto a su media y las sumamos se tiene:

(-2) + (-1) + (0) +(1) +(2) = 0

Por lo que tomando diferencias simples no es posible determinar la dispersión de los datos. Si ahora se toman esas mismas diferencias al cuadrado y las sumamos se tiene:

4 + 1 + 0 + 1 + 4 = 10

Dispersión o varianza de los datos

El grado en que los datos numéricos tienden a extenderse alrededor de un valor medio se llama varianza o dispersión de los datos. Se utilizan distintas medidas de dispersión o variación y una de las más empleadas es la varianza, que es una medida que ayuda a comprender la variabilidad de los datos, qué tan distanciados están de la media. Existen dos tipos de desviación típica:

1. Varianza poblacional (*σ2*): se obtiene dividiendo el valor anterior entre *n = 5*, es decir, el promedio de la suma de las diferencias al cuadrado, tomando *n* datos:

$$\sigma^2 = \sum \frac{(xi - \bar{x})^2}{n}$$

2. Varianza poblacional para muestras (s^2): se obtiene dividiendo el valor anterior entre *n - 1 = 4*, o sea el promedio de la suma de las diferencias al cuadrado, tomando *n -1* datos:

$$s^2 = \sum \frac{(xi - \bar{x})^2}{n-1}$$

Desviación estándar es la raíz cuadrada de la varianza:

Para el caso de una población: $\sigma = \sqrt{\sum \frac{(xi - \bar{x})^2}{n}}$

Para el caso de una muestra: $s = \sqrt{\sum \frac{(xi - \bar{x})^2}{n-1}}$

<u>Ejemplo 1</u>

La resistencia al rompimiento de dos muestras de botellas es la siguiente:

Muestra 1:	230	250	245	258	265	240
Muestra 2:	190	228	305	240	265	260

Calcule la desviación estándar para ambas muestras.

Media de la muestra 1

$\bar{x} = 248$

$\sum(x_i - \bar{x}^2)^2 = 790$

n - 1 = 5

$s = \sqrt{\frac{790}{5}} = 12.56$

Media de la muestra 2

$\bar{x} = 248$

$\sum(x_i - \bar{x}^2)^2 = 7510$

n-1 = 5

$s = \sqrt{\frac{7510}{5}} = 38.75$

Aunque la media en ambas muestras es la misma la desviación estándar (s), es menor en la muestra 1, por lo cual se deduce que es la que presenta menor variabilidad.

<u>Ejemplo 2</u>

Se desea hacer un estudio estadístico de la variación del calibre de una película de polietileno, para esto es necesario tomar una muestra y calcular la media, mediana y la desviación estándar.

Se realizan 14 observaciones arrojando los siguientes datos (en milésimas de pulgada):

2.11, 3.8, 4.0, 4.0, 3.1, 2.9, 2.5, 3.6, 2.0, 2.4, 2.8, 2.6, 2.9, 3.0

1. Cálculo de la media:

   $\sum x_i = 41.71$

   n = 14

   $$\frac{x_i}{n} = \frac{41.71}{14} = 2.98$$

   Mediana: ordenando los datos de mayor a menor se obtiene:

   2.0, 2.1, 2.4, 2.5, 2.6, 2.8, 2.9, 2.9, 3.0, 3.1, 3.6, 3.8, 4.0, 4.0

   Como el número de observaciones es impar, la mediana es el promedio aritmético de los dos números que están al centro, es decir, en las posiciones 7 y 8, respectivamente.

   Esto es:

   $$\frac{2.9 + 2.9}{2} = 2.9$$

2. Desviación estándar:

   $\sum(x_i - \bar{x}^2)^2 = 5.63$

   n - 1 = 13

   $$s = \sqrt{\frac{5.63}{13}} = 0\ .65$$

Actividad 17

Hallar la varianza y desviación estándar de las siguientes distribuciones de datos:

a. 12, 6, 7, 3, 15, 10, 18, 5

b. 9, 3, 8, 8, 9, 8, 9, 18

Medidas de tendencia central de datos agrupados

La media

Se calcula utilizando la siguiente fórmula:

$$\bar{X} = \frac{\sum f \cdot M}{n}$$

Donde:

$\bar{X}$ es la media.

f es la frecuencia o número de observaciones en cada clase.

M es el punto medio de cada clase, se determina como el valor medio entre los límites de clase.

n es el tamaño de la muestra o la suma de todas las frecuencias de las clases.

Ejemplo 1

Una línea aérea analiza el movimiento de pasajeros en los días festivos para poder ofrecer ofertas y aumentar sus ventas. Estos datos se ven reflejados en la siguiente tabla:

Pasajeros	Días (f)
100-109	2
90-99	8
80-89	12
70-79	18
60-69	7
50-59	3
	n=50

Dada la siguiente tabla de resultados, calcule la media de los datos.

Solución

Se encuentran las marcas de clase y se multiplican por cada frecuencia:

Pasajeros	Días (f)	Marca de clase (M)	f·M
100-109	2	104.5	209.0
90-99	8	94.5	756.0
80-89	12	84.5	1014.0
70-79	18	74.5	1341.0
60-69	7	64.5	451.5
50-59	3	54.5	163.5
	n=50		∑ f·x=3935

Se calcula la media:

$$\bar{X} = \frac{\Sigma f \cdot x}{n} = \frac{3935}{50} = 78.7 \text{ pasajeros}$$

La mediana

La mediana se calcula con la fórmula:

$$Med = L_{md} + \left[\frac{\frac{n}{2} - f_a}{f_{md}}\right](c)$$

Donde:

L_{md} es el límite inferior de la clase de la mediana cuya $f \geq n/2$.

f_a es la frecuencia acumulada de la clase que antecede a la clase de la mediana, la frecuencia acumulada (f_a) que es la suma de las frecuencias anteriores a cada intervalo.

f_{md} es la frecuencia de la clase de la mediana.

c es la amplitud del intervalo de clase de la mediana que es la diferencia entre dos límites reales de clase.

<u>Ejemplo 2</u>

Calcular la mediana de la distribución de datos del ejemplo 1.

Observe que se añadió a la tabla la columna de la frecuencia acumulada (f_a):

Pasajeros	Días (f)	Marca de clase (M)	f · M	Frecuencia acumulada (f_a)
100-109	2	104.5	209.0	2
90-99	8	94.5	756.0	10
80-89	12	84.5	1014.0	22
70-79	18	74.5	1341.0	40
60-69	7	64.5	451.5	47
50-59	3	54.5	163.5	50
	n=50		∑ f·x=3935	

Solución

Primero se identifica la clase donde se encuentra la mediana cuya $f \geq \frac{n}{2}$, en este caso $\frac{n}{2} = \frac{50}{2} = 25$ y pertenece a la clase de 70 a 79 con punto central de clase = 74.5, el límite inferior es 70 $L_{md} = 70$.

La clase mediana es de 70 a 79 entonces su frecuencia es de 18(f_{md}=18), por tanto, la frecuencia acumulada anterior a la clase mediana es 10 (f_a = 10).

c es la amplitud del intervalo de la clase mediana que es 10, porque 79.5 – 69.5 = 10.

Con estos elementos se calcula la mediana, así:

$$Med = L_{md} + \left[\frac{\frac{n}{2} - f_a}{f_{md}}\right](c) = 70 + \left[\frac{\frac{50}{2} - 10}{18}\right](10) = 78.33 \text{ pasajeros}$$

Moda

La moda se calcula con la fórmula:

$$M_o = L_{mo} + \left[\frac{\Delta_1}{\Delta_2 + \Delta_1}\right](c)$$

Donde:

L_{mo} es el límite inferior de la clase modal con la frecuencia más alta (70).

Δ_1 es la diferencia entre la frecuencia de la clase modal y la clase que la antecede (18 – 7 = 11).

Δ_2 es la diferencia entre la frecuencia de la clase modal y la clase que le sigue (18 – 12 = 6).

c es la amplitud del intervalo de la clase modal (80 – 70 = 10).

Ejemplo 3

Calcular la moda de la distribución de datos del ejemplo 1.

La frecuencia más alta es 18, entonces es la clase modal que posee la frecuencia más alta, en este caso la clase 70 a 79, por tanto, el límite modal es 70 ($L_{mo} = 70$).

La frecuencia de la clase que antecede a la clase modal es 7, entonces:

$\Delta_1 = 18 - 7 = 11$.

La frecuencia de la clase modal que le sigue a la clase modal es 12, entonces:

$\Delta_2 = 18 - 12 = 6$.

c es la amplitud del intervalo de la clase modal que es 10, porque 79.5 – 69.5 = 10.

Con estos elementos se calcula la mediana, así:

$$M_o = L_{mo} + \left[\frac{\Delta_1}{\Delta_2 + \Delta_1}\right](c) = 70 + \left[\frac{11}{6+11}\right](10) = 76.47$$

Actividad 18

La altura de estudiantes de un instituto de secundaria está tabulada en la siguiente tabla. Encuentre la media, la moda y la mediana de la altura de los estudiantes:

Altura (pulgadas)	Frecuencia (f)
60-62	5
63-65	18
66-68	42
69-71	27
72-74	8

Actividad 19

Encuentre la media, la moda y la mediana de los siguientes datos:

Clase	Frecuencia (f)
11-15	3
16-20	7
21-25	16
26-30	12
31-35	9
36-40	5
41-45	2

Medidas de dispersión de datos agrupados

Varianza y desviación estándar

Para calcular la varianza y la desviación estándar se utilizan las siguientes fórmulas:

$$s^2 = \frac{\sum fM^2 - n\overline{X}^2}{n-1}$$

$$s = \sqrt{s^2}$$

Donde:

f es la frecuencia.

M es la marca de clase.

$\bar{X}$ es la media de los datos.

n es el total de datos.

Ejemplo 1

Calcule la varianza y desviación estándar de la siguiente distribución de datos:

Pasajeros	Días (f)
100-109	2
90-99	8
80-89	12
70-79	18
60-69	7
50-59	3
	n=50

Solución

Se colocan en la tabla los siguientes elementos M, M^2, $f \cdot M$, $f \cdot M^2$, se realizan las sumas , se calcula la media, luego la varianza y la desviación estándar:

Pasajeros	Días (f)	Marca de clase(M)	f·M	M^2	$f \cdot M^2$
100-109	2	104.5	209.0	10920.25	21840.50
90-99	8	94.5	756.0	8330.25	71442.00
80-89	12	84.5	1014.0	7140.25	85683.00
70-79	18	74.5	1341.0	5550.25	99904.50
60-69	7	64.5	451.5	4160.25	29121.75
50-59	3	54.5	163.5	2790.25	8910.75
	n=50	$\Sigma f \cdot x = 3935$			$\Sigma f \cdot M^2 = 316902.5$

Para los datos anteriores se tiene:

$$\bar{X} = \frac{\sum f \cdot x}{n} = \frac{3935}{50} = 78.7 \text{ pasajeros}$$

$$s^2 = \frac{316902.50 - 50(78.7)^2}{49} = 147.31 \text{ pasajeros}$$

$$s = 12.14 \text{ pasajeros}$$

Con esta información, la línea aérea puede tomar sus decisiones.

Actividad 20

La altura de estudiantes de un instituto de secundaria está tabulada en la siguiente tabla. Encuentre la varianza y la desviación estándar de los datos que representan la altura de los estudiantes:

Altura (pulgadas)	Frecuencia (f)
60-62	5
63-65	18
66-68	42
69-71	27
72-74	8

$s^2=8.5275$

$s=2.92$

Actividad 21

Encuentre la varianza y la desviación estándar de los siguientes datos:

Clase	Frecuencia (f)
11-15	3
16-20	7
21-25	16
26-30	12
31-35	9
36-40	5
41-45	2

Autoevaluación

A continuación se le presenta una serie de ejercicios de evaluación que usted debe responder y confrontar con su propio aprendizaje. Recuerde que debe ser muy honesto, ya que usted es el responsable de su aprendizaje.

1. Construya un gráfico de barras con la información de la siguiente tabla:

Categoría	Frecuencia
Dieta severa	9
Miedo a engordar	3
Hiperactividad	4
Laxantes	5
Ropa holgada	6

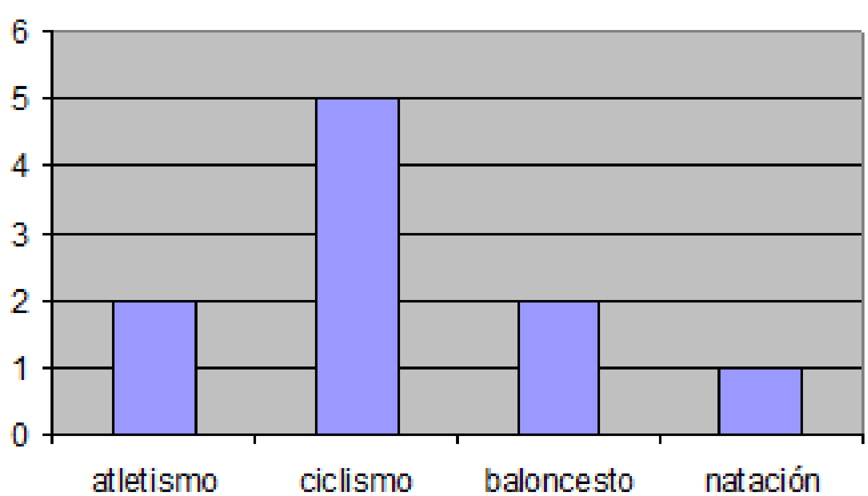
2. La siguiente tabla refleja las calificaciones de 30 alumnos en un examen de matemáticas:

Nota	2	4	5	6	7	8	9	10
No. alumnos	2	5	8	7	2	3	2	1

   a. ¿Cuántos alumnos aprobaron? ¿Cuántos alumnos sacaron como máximo un 7? ¿Cuántos sacaron como mínimo un 6?

   b. Calcule la nota media, la moda, la mediana, varianza y desviación estándar.

3. A partir de la siguiente gráfica estadística de gustos deportivos:



a. Calcule la tabla de frecuencias.

b. ¿A qué porcentaje de las personas no le gusta el ciclismo?

4. La altura en cm de 25 alumnos es la siguiente:

167	159	168	165	150	170	172	158	163	156
151	173	175	164	153	158	157	164	169	163
160	159	158	174	164					

Elabore una tabla que represente estos resultados con sus frecuencias absolutas, tome intervalos de amplitud 5 cm comenzando por 150. Calcule la media, la moda, la mediana, la varianza y la desviación estándar de los datos.

5. Elabore un diagrama de sectores que represente la procedencia de los extranjeros residentes en España, en diciembre de 1999, recogidos en la siguiente tabla:

Procedencia	Extranjeros residentes en España
Europa	353.556
América	166.709
Asia	66.340
África	213.012
Oceanía	1.013
Desconocida	699

6. Se ha aplicado un test a los empleados de una fábrica, obteniéndose la siguiente tabla:

Clase	Frecuencia (f)
38-44	7
44- 50	8
50- 56	15
56- 62	25
62- 68	18
68 74	9
74- 80	6

Dibuje el histograma y el polígono de frecuencias de la distribución.

Actividad metacognitiva

Con base a lo que ha aprendido, responda lo siguiente:

1. ¿Qué ventajas tiene organizar los datos?

2. ¿Por qué es importante la estadística en la vida cotidiana?

3. ¿Considera que adquirió nuevos aprendizajes al estudiar los temas de esta unidad? ¿Puede mencionar cuáles?

4. ¿Qué contenidos de los estudiados considera importantes para su aplicación en su vida habitual? ¿Por qué?

5. ¿En qué temas de esta unidad tuvo mayor dificultad?

Glosario

Gráfico estadístico: representación de un fenómeno por medio de figuras geométricas.

Frecuencia absoluta: número de veces que aparece cada modalidad o dato (resultado del recuento). La frecuencia total, de todas las modalidades juntas, se representa por n.

Distribución de frecuencias: agrupación de datos según la frecuencia de los valores.

Intervalo: conjunto de los valores que toma una magnitud entre dos límites dados.

Bibliografía

Alea, V.; Maqueda, I.; Muñoz, C. y Viladomiu, N. (2001). *Estadística para las Ciencias Sociales (Cuestiones tipo test)*. Madrid: AC.

Carrascal Arranz, U. (2007). *Estadística descriptiva con Microsoft® Excel 2007.* Madrid: RA-MA.

Montero Lorenzo, J. M. (2007). *Problemas resueltos de estadística descriptiva para ciencias sociales.* Madrid: Thomson.

Mures Quintana, M. J. (2003). *Problemas de Estadística Descriptiva Aplicada a las Ciencias Sociales.* Madrid: Pearson- Prentice Hall.

Tomeo, V. y Uña, I. (2003). *Lecciones de estadística descriptiva. Curso teórico-práctico.* Madrid: Thomson.