Universidad de Puerto Rico en Bayamón
Departamento de Matemáticas

MÓDULO 10:

INTRODUCCIÓN A LA ESTADÍSTICA

(7mo – 9no)

Preparado por:
Prof. Adalberto Agosto
Catedrático Auxiliar, Departamento de Matemáticas
Universidad de Puerto Rico en Bayamón

junio 2010
PRE-PRUEBA

Resuelva cada uno de los siguientes ejercicios y escoja la mejor alternativa.

1. La rama de la estadística que trata sobre la estimación, predicción y toma de decisiones a base de los resultados obtenidos de muestras es la estadística:

   a. descriptiva
   b. inferencial
   c. subjetiva
   d. probabilística

2. La variable, *cantidad de sangre que pasa por el corazón de un humano cada hora*, es:

   a. Cualitativa
   b. Cuantitativa continua
   c. Cuantitativa discreta
   d. Ninguna de las anteriores

3. Un maestro quiere hacer un estudio sobre todos los estudiantes que ha tenido durante sus veinte años en dicha profesión. Él decide seleccionar al azar el 10% de los estudiantes de cada uno de los grupos de los diferentes años académicos que ha tenido. Este método de muestreo se le conoce como:

   a. Estratificado
   b. Por conglomerado
   c. Sistemático
   d. Por tómbola

Considere la siguiente distribución de frecuencias para contestar las preguntas 4 y 5.

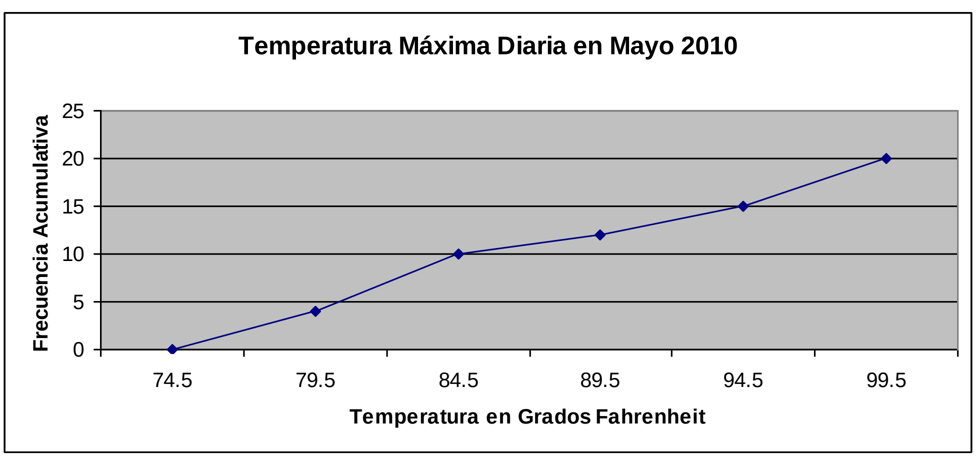
Edades de Ganadores de la Loto

Edad	Frecuencias
21 – 30	3
31 – 40	5
41 – 50	8
51 – 60	3
61 – 70	1

4. El ancho de clases de esta distribución es:
   a. 6
   b. 9
   c. 10
   d. 49

5. La marca de clase correspondiente a la segunda clase es
   a. 31
   b. 35.5
   c. 40
   d. 5

Considere la siguiente gráfica para contestar las preguntas 6 y 7.



6. La gráfica anterior se conoce como

a. ojiva
b. polígono de frecuencias
c. histograma
d. tallo y hojas

7. Los valores 74.5, 79.5, 84.5, 89.5, 94.5 y 99.5 en la gráfica anterior son:

a. límites de clases
b. marcas de clases
c. frecuencias de clases
d. fronteras de clases

8. La mediana en el siguiente conjunto de datos {4, 5, 3, 8, 3, 9, 2} es:

a. 3
b. 4
c. 5.5
d. 8

9. La moda de los datos resumidos en la siguiente distribución de frecuencias es:

Número de Hermanos	3	4	6	7	9	10	14
Frecuencia	1	2	2	3	5	2	1

a. 2
b. 5
c. 9
d. 14

10. La medida de tendencia central que más se afecta con los datos extremos o aislados es:

a. la media
b. el recorrido medio
c. la moda
d. la mediana

OBJETIVOS

Luego de finalizar el estudio de este módulo estarás capacitado para

1. distinguir entre parámetros y estadísticos.
2. identificar variables cualitativas y cuantitativas en la población.
3. distinguir entre variables discretas y continuas.
4. discutir los fundamentos en la conducción de métodos estadísticos.
5. describir la importancia de la estadística en el proceso de la toma de decisiones.
6. distinguir el proceso de muestreo utilizado en un estudio de investigación.
7. construir una distribución de frecuencias, frecuencias relativas y frecuencias acumulativas.
8. determinar el ancho de clase, límites de clase y marcas de clase para una distribución de frecuencias dada.
9. construir las gráficas de barras, circulares y pictóricas para presentar datos cualitativos.
10. construir histogramas, polígonos de frecuencias y ojivas para representar datos cuantitativos.
11. ordenar datos utilizando diagramas de tallo y hojas.
12. determinar la media, mediana, moda y recorrido medio para un conjunto de datos.
13. determinar la clase modal de una distribución de frecuencias.
14. aproximar la media y la mediana de datos agrupados en una distribución de frecuencias.
15. discutir las ventajas y desventajas de cada una de las medidas de tendencia central.

JUSTIFICACIÓN

La Estadística es por sí misma auxiliar de todas las demás ciencias. Los mercados, la medicina, la ingeniería, los gobiernos, etc. se nombran entre los más destacados clientes de ésta. La ausencia de ésta conllevaría a un caos generalizado, dejando a los administradores y ejecutivos sin información vital a la hora de tomar decisiones en tiempos de incertidumbre.

La Estadística que conocemos hoy en día debe gran parte de su realización a los trabajos matemáticos de aquellas personas que desarrollaron la teoría de las probabilidades.

Este módulo ha sido diseñado con el propósito de desarrollar en usted los conocimientos básicos acerca de conceptos estadísticos así como las destrezas relacionadas a la ordenación de datos, construcción de distribuciones de frecuencias y gráficas y la computación de las medidas de tendencia central relacionadas a los datos.

INTRODUCCIÓN A LA ESTADÍSTICA

CONCEPTOS BÁSICOS

Es común en nuestros días escuchar sobre la estadística y lo que los estudios investigativos reflejan utilizando estadística. Los hogares, los gobiernos y empresas se apoyan mucho en datos estadísticos para dirigir sus acciones. Pero, ¿qué es estadística? ***Estadística*** es la ciencia que estudia los métodos para la recopilación, clasificación, presentación e interpretación de datos para llegar a conclusiones con respecto a una población. Llamamos ***población*** al conjunto total de los elementos que serán estudiados. Una de las metas más importante de la materia de la estadística aprender sobre un grupo grande de individuos estudiando algunos de ellos.

La estadística se divide en dos ramas: la estadística descriptiva y la estadística inferencial. La ***estadística descriptiva*** estudia las técnicas de recopilación, clasificación y presentación de los datos. Por otro lado, la ***estadística inferencial*** estudia los métodos que se utilizan para analizar e interpretar los valores, las tablas, las graficas y cualquier otro resultado obtenido mediante la estadística descriptiva con el fin de llegar a conclusiones con respecto a una población de interés.

En muchas ocasiones es difícil o imposible hallar el valor exacto de una variable de interés en una población. Por ejemplo, ¿cuál es el número exacto de habitantes de Puerto Rico que son positivos al virus del SIDA hoy? Nadie lo sabe y es imposible saberlo con todas las limitaciones de tiempo, espacio, disponibilidad y efectividad de las pruebas, etc.. Este número es un ejemplo de un parámetro. Un ***parámetro*** es una característica de la población asociada a la variable de interés. En muchas ocasiones, el valor de un parámetro es estimado con un estadístico. Un ***estadístico*** es una característica de una muestra de la población asociada a la variable de interés, donde una ***muestra*** es un subconjunto de los elementos de la población.

Los estudios de investigación basados en la estadística pueden clasificarse en dos categorías: censo y muestral. En un ***estudio tipo censo*** se analizan los datos de la variable de interés de todos los elementos de la población. Sin embargo, en un ***estudio muestral*** se analizan los datos de la variable de interés de algunos elementos de la población (los de la muestra).

Es importante conocer cómo utilizar datos muestrales para llegar a inferencias sobre poblaciones. Veremos que es de suma importancia que la muestra sea representativa de la población, de lo contrario, son inútiles. Por lo tanto, los datos muestrales deben recopilarse de forma adecuada, así como a través de un proceso de selección aleatorio.

TIPOS DE DATOS

Un ***dato*** es el valor de la variable de interés asociado con un elemento de la población o muestra. Cuando hablamos de ***datos*** nos referimos al conjunto de valores de la variable de interés obtenidos de la población o muestra. Algunos datos consisten en números y otros son no numéricos. Los ***datos cuantitativos*** consisten en números que resultan de conteos o mediciones. Un ***dato cuantitativo discreto*** es un número que resulta de un conteo mientras que un ***dato cuantitativo continuo*** es un número que resulta de una medición dentro de una escala continua (sin huecos, saltos o interrupciones). Por otro lado, los ***datos cualitativos*** son categorías o atributos que describen a un elemento de la población.

Ejemplo 1: Un investigador desea determinar el número promedio de hermanos que tienen los estudiantes matriculados en la Universidad de Puerto Rico en Bayamón (UPRB) en cierto semestre. Para conseguirlo, decide seleccionar al azar 500 estudiantes de la lista oficial de los 4,500 estudiantes matriculados. Al conducir su encuesta, encuentra que el número promedio de hermanos de los estudiantes de la UPRB es 2.25. Determine:

a. Población
b. Muestra
c. Variable de Interés
d. Parámetro
e. Estadístico
f. Tipo de Datos

Respuesta:

a. Población – los 4,500 estudiantes matriculados en ese semestre
b. Muestra – 500 estudiantes seleccionados al azar
c. Variable de Interés – número de hermanos
d. Parámetro – número promedio de hermanos de todos los estudiantes matriculados en la UPRB en ese semestre
e. Estadístico – 2.25 hermanos por estudiante matriculado en la UPRB en ese semestre
f. Tipo de Datos – cuantitativo discreto

Ejemplo 2: Se lleva a cabo un estudio para determinar cuál es el tiempo promedio que tardan en salir de sus casas a trabajar, desde que se levantan de sus camas, las mujeres de San Juan que trabajan fuera de sus hogares. La investigadora selecciona 100 mujeres al azar de todas las mujeres de San Juan que trabajan fuera de sus hogares. A base de la muestra concluye que las mujeres de San Juan que trabajan fuera de sus hogares tardan un promedio de 90 minutos en salir de sus casas a trabajar desde que se levantan de sus camas. Determine:

a. Población
b. Muestra
c. Variable de Interés
d. Parámetro
e. Estadístico
f. Tipo de Datos

Respuesta:

a. Población – todas las mujeres de San Juan que trabajan fuera de sus hogares
b. Muestra – 100 mujeres de San Juan seleccionadas al azar que trabajan fuera de sus hogares
c. Variable de Interés – tiempo promedio que tardan las mujeres de San Juan que trabajan fuera de sus hogares en salir de sus casas a trabajar desde que se levantan de sus camas
d. Parámetro – tiempo promedio que tardan en salir de sus casas a trabajar desde que se levantan de sus camas todas las mujeres de San Juan que trabajan fuera de sus hogares
e. Estadístico – 90 minutos
f. Tipo de Datos – cuantitativo continuo

Ejemplo 3: Para cada una de las siguientes variables de interés indique si es cualitativa, cuantitativa discreta o cuantitativa continua según su tipo de datos:

a. nacionalidad de una persona

b. número de empleados masculinos

c. estatura de los estudiantes de noveno grado

d. estado civil de los maestros

e. número de millas recorridas en un viaje

f. número máximo de millas de garantía de un auto

Respuesta:

a. nacionalidad de una persona - cualitativa

b. número de empleados masculinos - cuantitativa discreta

c. estatura de los estudiantes de noveno grado - cuantitativa continua

d. estado civil de los maestros - cualitativa

e. número de millas recorridas en un viaje - cuantitativa continua

f. número máximo de millas de garantía de un auto - cuantitativa discreta

Ejercicios 1:

1. El 37 % de los 3,600 estudiantes de cierta universidad del área metropolitana de Puerto Rico provienen del área isla. Un investigador no tiene forma de conseguir este porcentaje y decide estimarlo utilizando una muestra representativa de 400 estudiantes. Al finalizar su estudio, él encuentra que 34 % de los estudiantes encuestados provienen de área isla.

   a. ¿Cuál es la población de interés?

b. Determine

   i. variable de interés

   ii. muestra

   iii. parámetro

   iv. estadístico

   v. tipo de datos

2. Para cada una de las siguientes variables de interés indique si es cualitativa, cuantitativa discreta o cuantitativa continua según su tipo de datos:

   a. peso en libras de una persona

   b. color de auto

   c. cantidad de dinero en una cuenta de ahorro

   d. deporte favorito

   e. créditos matriculados por semestre en una universidad

   f. temperatura actual en grados Fahrenheit

MÉTODOS DE MUESTREO

El objetivo principal de un estudio investigativo utilizando estadística es poder llegar a conclusiones o generalizaciones válidas para una población de acuerdo a los resultados obtenidos en muestras representativas de esa población. Decimos que una ***muestra*** es ***representativa*** de la población si tiene las mismas características relevantes que la población. Obviamente, la validez de las conclusiones dependerá de la muestra seleccionada. A continuación discutiremos algunos métodos de muestreo.

Muestreo Aleatorio - Es el proceso mediante el cual se selecciona una muestra de tamaño n de una población de tamaño N de forma tal que cada elemento de la población tiene igual probabilidad de ser seleccionado en la muestra.

Ejemplo 4: El número total de estudiantes de la Escuela Intermedia Miguel Meléndez Muñoz es 600. Al seleccionar una muestra del 10% de los estudiantes lo hacemos escribiendo en un papel el nombre de cada estudiante y echamos todos los papeles en una tómbola. Luego sacamos al azar 60 papeles que indicarán los nombres de los estudiantes seleccionados como muestra.

Notemos que el tamaño de la población es 600, el tamaño de la muestra es 60 y cada estudiante tiene igual probabilidad de ser seleccionado en la muestra.

Muestreo Probabilístico - Es el proceso mediante el cual se selecciona una muestra de una población de forma tal que cada elemento de la población tiene una probabilidad conocida de ser seleccionado en la muestra aunque pueda no ser la misma para cada elemento.

Ejemplo 5: El número total de estudiantes de la Escuela Intermedia Salvador Brau es 550. Al seleccionar una muestra de 55 estudiantes lo hacemos de la siguiente manera. Primero dividimos los 550 estudiantes en 55 grupos de 11 estudiantes cada uno donde cada estudiante pertenezca a un solo grupo. Luego asignamos al azar un número distinto del 2 al 12 a cada integrante de cada grupo. Finalmente lanzamos dos dados de seis caras cada uno y los estudiantes que tengan el número igual a la suma de las caras de ambos dados serán los seleccionados como muestra.

Notemos que el tamaño de la población es 550, el tamaño de la muestra es 55 y cada estudiante tiene una probabilidad conocida de ser seleccionado en la muestra, pero no todos con igual probabilidad. La próxima tabla nos presenta las probabilidades de cada número de acuerdo a las distintas posibilidades de obtenerlo.

PROBABILIDADES PARA LA SUMA DE DOS DADOS LANZADOS AL AZAR

Suma	Posibilidades	Probabilidad
2	(1, 1)	$\frac{1}{36}$
3	(1, 2) y (2, 1)	$\frac{2}{36}$
4	(1, 3) , (3, 1) y (2, 2)	$\frac{3}{36}$
5	(1, 4) , (4, 1) , (2, 3) y (3, 2)	$\frac{4}{36}$
6	(1, 5) , (5, 1) , (2, 4) , (4, 2) y (3, 3)	$\frac{5}{36}$
7	(1, 6) , (6, 1) , (2, 5) , (5, 2) , (3, 4) y (4, 3)	$\frac{6}{36}$
8	(2, 6) , (6, 2) , (3, 5) , (5, 3) y (4, 4)	$\frac{5}{36}$
9	(3, 6) , (6, 3) , (4, 5) y (5, 4)	$\frac{4}{36}$
10	(4, 6) , (6, 4) y (5, 5)	$\frac{3}{36}$
11	(5, 6) y (6, 5)	$\frac{2}{36}$
12	(6, 6)	$\frac{1}{36}$

Muestreo Sistemático - Es el proceso mediante el cual se selecciona una muestra de una población donde primero seleccionamos al azar un punto de partida de una lista o un registro de la población. Luego seleccionamos al azar un número k. Finalmente, tomamos cada k-ésimo elemento de la lista comenzando desde el punto de partida hasta tomar el número total de datos necesarios. Este punto de partida es arbitrario, por lo que puede ser tomado como parte de la muestra si se decide hacerlo.

Ejemplo 6: Una compañía con 1,000 empleados desea estimar la estatura promedio de sus empleados. El investigador decide seleccionar 80 empleados al azar como muestra. Para hacerlo busca los expedientes enumerados (desde 1 al 1,000) de todos los empleados y selecciona un número al azar como punto de partida. Suponga que se obtuvo el número 3 y el investigador decide tomarlo como parte de la muestra. Luego selecciona cada duodécimo expediente a partir del 3. Es decir, la muestra constará de los empleados correspondientes a los expedientes 3, 15, 27, 39, 51,…, 951.

Muestreo Estratificado - Es el proceso mediante el cual se selecciona una muestra de una población donde primero subdividimos la población en al menos dos subgrupos (estratos) diferentes de manera tal que los individuos que pertenezcan a un mismo grupo compartan las mismas características (como género, lugar de residencia, edad, ingresos, etc.). Luego seleccionamos una muestra al azar de cada subgrupo.

Ejemplo 7: Un estudiante universitario desea estimar qué por ciento de los estudiantes están a favor de cierta medida tomada por la administración de la universidad en la cual estudia. El estudiante divide toda la población estudiantil de la universidad en ocho subgrupos divididos por género y año de clasificación (1ro, 2do, 3ro o 4to). Luego toma como muestra el 10% de los integrantes de cada grupo seleccionados al azar. Los resultados son resumidos en la siguiente tabla.

MUESTRA DIVIDIDA POR GÉNERO Y AÑO DE CLASIFICACIÓN

Estrato	Cantidad de Estudiantes	Muestra (10%)
Femenino y 1er año	400	40
Masculino y 1er año	350	35
Femenino y 2do año	300	30
Masculino y 2do año	200	20
Femenino y 3er año	280	28
Masculino y 3er año	190	19
Femenino y 4to año	420	42
Masculino y 4to año	310	31

La muestra constará de 225 estudiantes de una población de 2,250.

Muestreo por Conglomerado - Es el proceso mediante el cual se selecciona una muestra de una población donde primero subdividimos la población en al menos dos subgrupos (estratos) diferentes de manera tal que los individuos que pertenezcan a un mismo grupo compartan las mismas características. Luego seleccionamos la muestra tomando aleatoriamente uno o varios grupos completos de los subdivididos.

Ejemplo 8: En el ejemplo anterior tenemos toda la población estudiantil dividida en ocho grupos. Seleccionamos al azar dos de estos grupos para componer la muestra. Un posible resultado será todos los estudiantes masculinos de segundo año unidos a las féminas de tercer año y la muestra constará de 480 estudiantes de una población de 2,250.

Muestreo por Conveniencia - En este método para obtener la muestra tomamos los datos de los elementos que son fácilmente accesibles. Este muestreo es regularmente utilizado cuando el tiempo o los recursos son pocos a la hora de realizar un estudio investigativo.

Ejemplo 9: Un reportero de un canal de televisión llega a una manifestación de protesta y entrevista a cada individuo que pasa frente a él para saber cuáles son los motivos principales de la protesta.

Notemos que la muestra es aleatoria en tanto y cuanto el reportero no discrimine sobre a quién entrevista y a quién no. De hecho, el ejemplo nos indica que el reportero entrevista a todo el que pasa cerca de él pues son los elementos fácilmente accesibles.

En general, el método o combinación de métodos utilizados a la hora de recopilar una muestra no es lo más importante, sino que la muestra sea representativa de la población. Para lograrlo la muestra debe ser aleatoria.

Ejercicios 2: Escoge la respuesta correcta:

1. La población de cierta ciudad se divide en 37 % mujeres mayores de 18 años, 20% mujeres menores de 18 años, 22% hombres mayores de 18 años y 21% hombres menores de 18 años. Si tomamos una muestra aleatoria del 15% de todos los grupos, esto es un ejemplo de muestreo

   a. sistemático   b. estratificado

   c. por conglomerado   d. por conveniencia

2. En una línea de producción de tornillos se toma el número cuarenta de cada cien tornillos producidos como muestra para verificar la calidad de la producción. Esto es un ejemplo de muestreo

a. sistemático b. estratificado

c. por conglomerado d. por conveniencia

3. En un estudio con fines sociológicos se desea saber el número promedio de hermanos que tienen los estudiantes de cierta escuela elemental. En esta escuela hay un solo grupo de cada uno de los grados de kinder a sexto. Para tomar la muestra identificamos un papelito por grupo con el nombre del grado y los echamos todos en una tómbola. La muestra constará de todos los estudiantes que cursan los dos grados seleccionados al azar de la tómbola. Esto es un ejemplo de muestreo

a. sistemático b. estratificado

c. por conglomerado d. por conveniencia

4. Una profesora universitaria desea saber cuál es el número promedio de los estudiantes fumadores en la universidad para la cual trabaja. Ella decide tomar como muestra todos los estudiantes que están matriculados en sus secciones. Esto es un ejemplo de muestreo

a. sistemático b. estratificado

c. por conglomerado d. por conveniencia

DISTRIBUCIONES DE FRECUECIAS

Al trabajar con grandes cantidades de datos, en muchas ocasiones es útil organizarlos o resumirlos en una tabla llamada distribución de frecuencias. Definimos una ***distribución de frecuencias*** como una lista de valores de datos de manera individual o por grupos de intervalos junto con sus respectivas frecuencias (o conteos).

A continuación presentamos una tabla, la cual es una distribución de frecuencias que resume las puntuaciones obtenidas por 31 estudiantes en un primer examen. Llamamos ***clases*** a los valores de los datos (individuales o grupales) ubicadas en la primera columna de la tabla. La ***frecuencia*** de una clase en particular es el número de datos que caen en esa clase. Por ejemplo, la frecuencia de la segunda clase de la tabla es 4, lo que indica que 4 estudiantes obtuvieron una puntuación entre 51 y 60 en el examen.

PUNTUACIONES EN EL PRIMER EXAMEN

Puntuación	Frecuencia
41 – 50	2
51 – 60	4
61 – 70	8
71 – 80	9
81 – 90	5
91 – 100	3

Antes de describir la manera en que se construyen las distribuciones de frecuencias les presentamos las siguientes definiciones de algunos términos relacionados con ellas.

Definiciones:

- Los ***límites de clases inferiores*** son los valores más pequeños que pueden pertenecer a las diferentes clases. (Los límites de clase inferiores de la tabla anterior son 41, 51, 61, 71, 81 y 91).
- Los ***límites de clases superiores*** son los valores más grandes que pueden pertenecer a las diferentes clases. (Los límites de clase superiores de la tabla anterior son 50, 60, 70, 80, 90 y 100).
- Los ***fronteras de clases*** son los valores que se utilizan para separar las clases, pero sin los espacios creados por los límites de clase. Para estos fines utilizamos los puntos o valores medios de cada espacio. Por ejemplo, las fronteras entre clases de la tabla anterior son 50.5, 60.5, 70.5, 80.5 y 90.5. Siguiendo el mismo patrón (restar 0.5 a los límites inferiores de las clases y sumar 0.5 a los límites superiores de las clases), la frontera de la primera clase es 40.5 y la frontera de la última clase es 100.5.
- Las ***marcas de clase*** son los puntos medios de las clases. Estos se calculan sumando el límite inferior con el límite superior, y dividiendo luego entre dos. (Las marcas de clase de la tabla anterior son 45.5, 55.5, 65.5, 75.5, 85.5 y 95.5)
- El ***ancho de clase*** es la diferencia absoluta entre dos límites de clases inferiores (o superiores) consecutivos. (El ancho de clase de la tabla anterior es 10).

Las distribuciones de frecuencias son útiles para resumir grandes conjuntos de datos y tener una base para construir gráficas, como los histogramas, que estudiaremos más adelante. Al construir una distribución de frecuencias debemos seguir las siguientes reglas:

1. El número de clases debe ser entre 5 y 20 inclusive.
2. Todas las clases deben tener el mismo ancho.

3. No debe haber intersección entre las clases. Cada dato debe pertenecer a una sola clase.
4. Ni la primera ni la última clase deben estar vacías.
5. Debe haber suficientes clases para que cada dato pertenezca a una clase.

Para construir una distribución de frecuencias debemos seguir el siguiente procedimiento básico:

1. Escoja el número de clases que desea entre 5 y 20.
2. Calcule el ancho de clase.

   $$\boxed{\text{ancho de clase} \approx \frac{\text{dato mayor} - \text{dato menor}}{\text{número de clases}}}$$

   Redondee el resultado para obtener un número más adecuado. Es posible que necesite cambiar el número de clases para obtener un ancho de clases apropiado para la distribución de los datos. Por ejemplo, no es apropiado tener pocas clases con muchos datos y muchas clases vacías, y tampoco es apropiado tener muchas clases con pocos datos. Sin embargo, cualquier cambio debe siempre cumplir con las reglas antes expuestas.
3. Escoja un número para el límite inferior de la primera clase. Puede escoger el dato menor o un valor conveniente un poco menor que éste.
4. Sume el ancho de clase al límite inferior de la primera clase para conseguir el límite inferior de la segunda clase. Sume el ancho de clase al límite inferior de la segunda clase para conseguir el límite inferior de la tercera clase, y así sucesivamente.
5. Calcule el límite superior de cada clase recordando mantener consistencia del ancho de clase (diferencia absoluta entre dos límites de clases inferiores o superiores consecutivos es la misma).
6. Anote los límites de clases en la primera columna de una tabla. Halle la cantidad de datos que caen en cada clase y anótela en la segunda columna (frecuencia). Póngale un título apropiando a la tabla de acuerdo con la procedencia de los datos.

DISTRIBUCIÓN DE FRECUECIAS RELATIVAS

Una variante importante de la distribución de frecuencias utiliza las ***frecuencias relativas***, las cuales se obtienen dividiendo cada frecuencia de clase entre el número total de datos. Las frecuencias relativas se pueden expresar de forma decimal, de forma fraccionaria o en por ciento. La suma total de las frecuencias relativas debe ser 1 o 100%, lo que aplique.

Una tabla de distribución de frecuencias relativas para la distribución de puntuaciones en el primer examen de la tabla discutida en la página 18 es la siguiente:

PUNTUACIONES EN EL PRIMER EXAMEN

Puntuación	Frecuencia	Frecuencia Relativa
41 – 50	2	$\frac{2}{31}$
51 – 60	4	$\frac{4}{31}$
61 – 70	8	$\frac{8}{31}$
71 – 80	9	$\frac{9}{31}$
81 – 90	5	$\frac{5}{31}$
91 – 100	3	$\frac{3}{31}$

Note que la suma de todas las frecuencias relativas es 1.

DISTRIBUCIÓN DE FRECUENCIAS ACUMULATIVAS

Otra variante importante de la distribución de frecuencias utiliza las *frecuencias acumulativas.* La ***frecuencia acumulativa*** de una clase se obtiene sumando la frecuencia de esta clase con las frecuencias de las clases anteriores. La frecuencia acumulativa de la última clase debe ser igual al número total de datos.

La tabla de distribución de frecuencias acumulativas para la distribución de puntuaciones de la tabla anterior es la siguiente:

PUNTUACIONES EN EL PRIMER EXAMEN

Puntuación	Frecuencia	Frecuencia Acumulativa
41 – 50	2	2
51 – 60	4	6
61 – 70	8	14
71 – 80	9	23
81 – 90	5	28
91 – 100	3	31

Ejemplo 10: Los siguientes datos representan las estaturas en pulgadas de 40 estudiantes universitarios.

68	69	60	58	72	62	65	76	62	63
70	60	67	68	64	61	61	64	63	70
62	69	68	68	65	68	59	66	71	78
62	64	69	70	65	68	62	69	73	58

Construya una distribución de frecuencias y determine las frecuencias relativas y acumuladas.

Respuesta:

Paso 1: Comience seleccionando el número total de clases como por ejemplo, 7.

Paso 2: Calcule el ancho de clase.

$$\text{ancho de clase} \approx \frac{78-58}{7} \approx 2.86$$

Redondee a 3, ya que es un número más conveniente.

Paso 3: Elegimos un punto de partida de 58 para el límite inferior de la primera clase.

Paso 4: Calculamos los límites inferiores de las clases, los cuales son 58, 61, 64, 67, 70, 73 y 76.

Paso 5: Calculamos los límites superiores de las clases, los cuales son 60, 63, 66, 69, 72, 75 y 78.

Paso 6: Construimos la tabla:

ESTATURA EN PULGADAS DE ESTUDIANTES UNIVERSITARIOS

Puntuación	Frecuencia	Frecuencia Relativa	Frecuencia Acumulativa
58 – 60	5	0.2	5
61 – 63	9	0.225	14
64 – 66	7	0.175	21
67 – 69	11	.275	32
70 – 72	5	.2	37
73 – 75	1	.025	38
76 – 78	2	.05	40

Ejercicio 3: Los siguientes datos representan el número de años de servicio de 30 maestros de una escuela pública de Bayamón.

13	22	5	15	12	32	21	3	16	13
8	8	16	18	28	10	25	12	23	22
9	19	26	4	15	20	28	30	25	23

Construya una distribución de frecuencias de 5 clases y determine las frecuencias relativas y acumuladas.

GRÁFICAS

Una de las mejores formas de resumir y presentar un conjunto de datos es mediante una representación gráfica. Aunque existe una gran variedad de tipos de graficas, la más apropiada está determinada por el tipo de dato y la percepción visual que se quiera presentar. A continuación les presentamos una descripción de algunas de las gráficas más utilizadas.

GRÁFICAS CIRCULARES

Las gráficas circulares generalmente se utilizan para visualizar datos cualitativos. Una gráfica circular consta de un círculo dividido por sectores como si fueran rebanadas de un pastel. Cada categoría o clase es representada por un sector proporcional en área a su frecuencia relativa.

Ejemplo 11: Se les preguntó a 50 estudiantes de la Escuela de Medicina del Recinto de Ciencias Médicas de la Universidad de Puerto Rico (UPR) cuál es su especialidad. Los resultados se resumen en la siguiente tabla. Construya una gráfica circular para estos datos.

ESPECIALIDAD DE ESTUDIANTES DE MEDICINA DE LA UPR

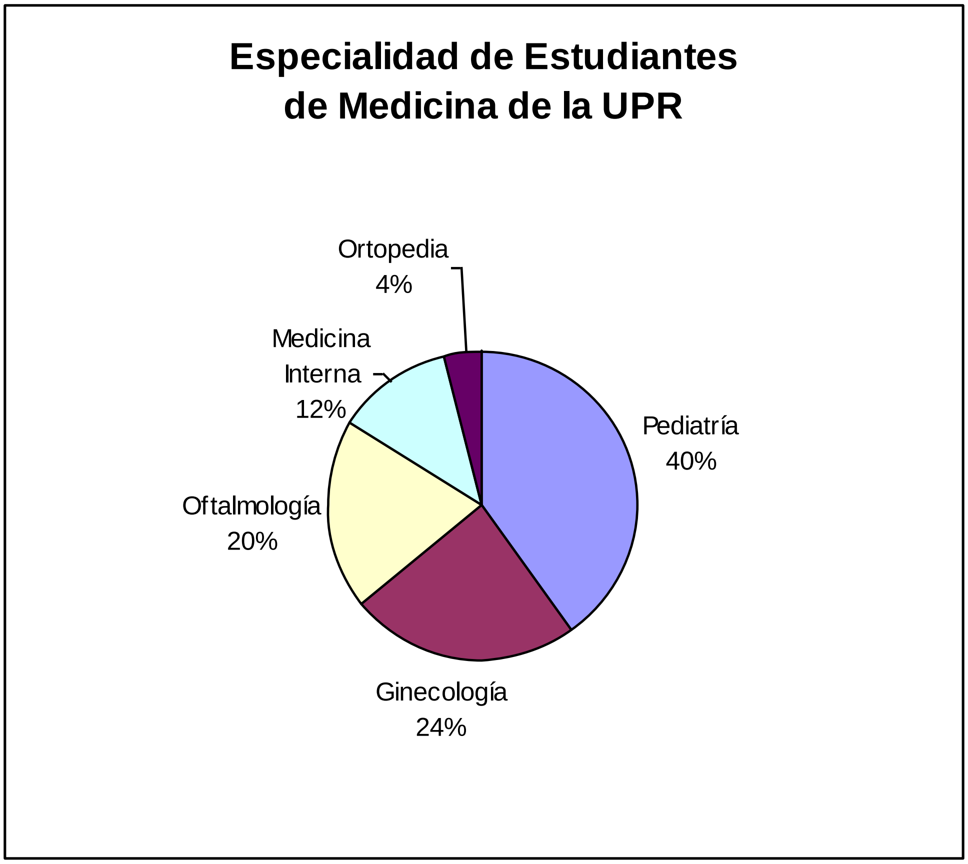
Especialidad	Frecuencia
Pediatría	20
Ginecología	12
Oftalmología	10
Medicina Interna	6
Ortopedia	2

Respuesta: Primero calculamos la frecuencia relativa de cada clase en forma de por ciento. Esto lo hacemos dividiendo la frecuencia de cada clase entre el número total de datos (50) y luego multiplicamos por 100.

ESPECIALIDAD DE ESTUDIANTES DE MEDICINA DE LA UPR

Especialidad	Frecuencia	Frecuencia Relativa
Pediatría	20	40%
Ginecología	12	24%
Oftalmología	10	20%
Medicina Interna	6	12%
Ortopedia	2	4%

Luego construimos la gráfica, la cual quedaría de esta forma:



GRÁFICAS DE BARRAS

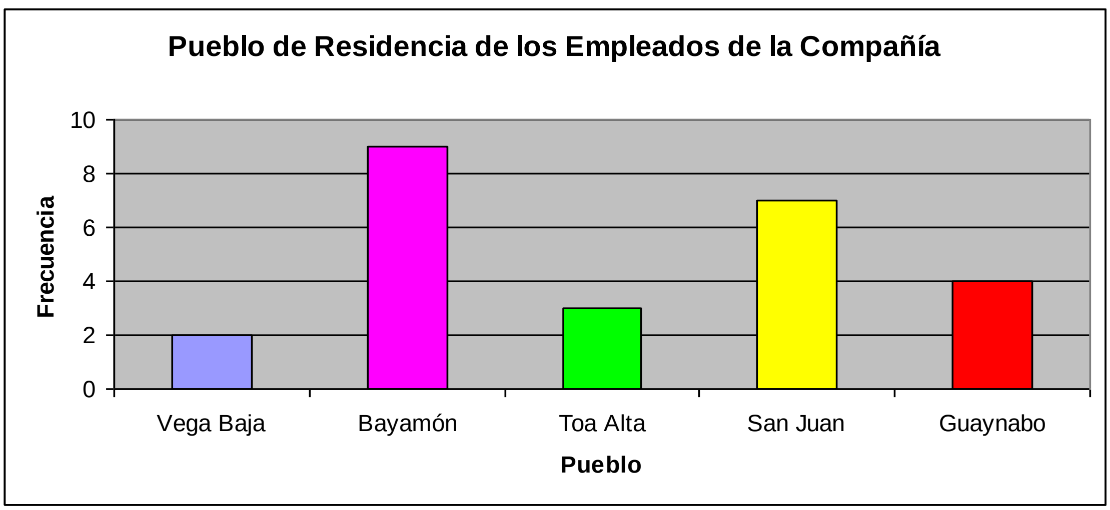
Las gráficas de barras se utilizan para visualizar datos cualitativos. Las barras pueden ser de forma vertical u horizontal. Una gráfica de barras consta de un cuadrante donde uno de sus ejes define los diferentes valores cualitativos de los datos y el otro la frecuencia de cada valor. Cada categoría o clase es representada por un rectángulo proporcional en longitud a su frecuencia.

Ejemplo 12: Se les preguntó a 25 empleados de una compañía en qué pueblo estaba su residencia principal. Los resultados se resumen en la siguiente tabla. Construya una gráfica de barras para estos datos.

PUEBLO DE RESIDENCIA DE LOS EMPLEADOS DE LA COMPAÑÍA

Pueblo	Frecuencia
Vega Baja	2
Bayamón	9
Toa Alta	3
San Juan	7
Guaynabo	4

Respuesta: Para construir la gráfica primero decidimos si las barras son de forma horizontal o vertical, lo que generalmente depende del espacio disponible para la gráfica. Luego construimos la gráfica, la cual quedaría de esta forma:



No olvide identificar los ejes y ponerle un título apropiado a la gráfica.

GRÁFICAS PICTÓRICAS

Las gráficas pictóricas generalmente se utilizan para visualizar datos cualitativos. Hay muchas variantes de gráficas pictóricas. Una de las variantes más utilizadas tiene figuras o dibujos que están asociados a una cierta cantidad de los datos que representa. Cada categoría o clase es representada por estas figuras en proporción a su frecuencia.

Ejemplo 13: La siguiente tabla resume el número de reclamaciones por accidentes de auto recibidas por una compañía de seguros durante el primer cuatrimestre del año 2010. Construya una gráfica pictórica para estos datos.

RECLAMACIONES RECIBIDAS PARA EL AÑO 2010

Mes	Frecuencia
Enero	200
Febrero	100
Marzo	150
Abril	250

Respuesta: La gráfica pictórica pudiera ser la siguiente:

HISTOGRAMAS DE FRECUENCIAS

Los histogramas de frecuencias son gráficas de barras que se utilizan para representar datos cuantitativos en la mayoría de los casos para distribuciones de frecuencias agrupadas. La escala horizontal representa las clases utilizando fronteras por lo que las barras quedan de manera adyacentes (sin huecos o espacios entre sí). Por otro lado, la escala vertical representa las frecuencias o las frecuencias relativas.

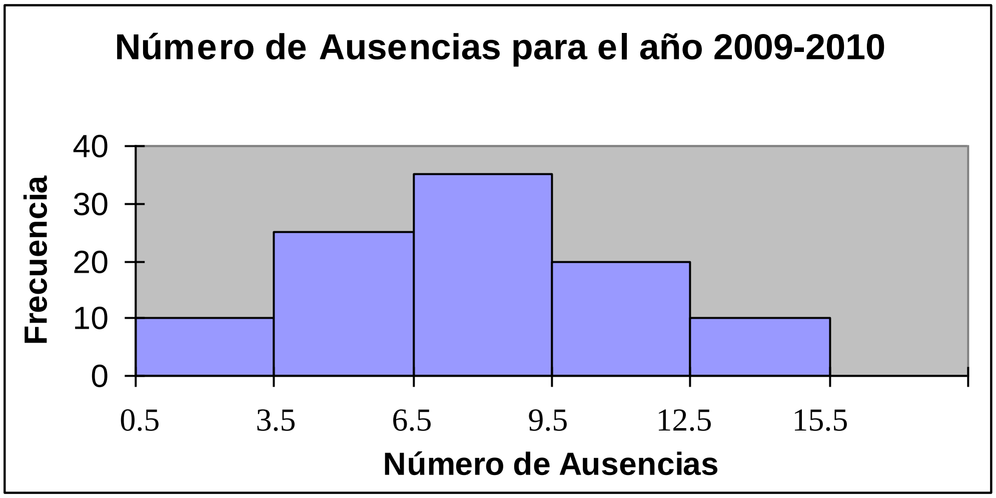
El primer paso para la construcción de un histograma es la creación de una tabla de distribución de frecuencias. El histograma es una versión gráfica de dicha tabla. Luego dibujamos un cuadrante donde marcamos el eje horizontal con las fronteras de clases y el eje vertical lo marcamos con una escala apropiada para la frecuencia o frecuencia relativa. Finalmente, dibujamos una barra por cada clase partiendo de su límite inferior hasta su límite superior con altura correspondiente a su frecuencia o frecuencia relativa.

Ejemplo 14: La siguiente tabla resume los resultados al ver los registros de ausencias de 100 estudiantes de la Escuela Intermedia Eugenio María de Hostos durante el año académico 2009-2010. Construya un histograma de frecuencia para estos datos.

NÚMERO DE AUSENCIAS PARA EL AÑ0 2009-2010

Número de Ausencias	Frecuencia
1 – 3	10
4 – 6	25
7 – 9	35
10 – 12	20
13 – 15	10

Respuesta: Primero determinamos las fronteras de clases, las cuales son: 0.5, 3.5, 6.5, 9.5, 12.5 y 15.5. Luego al construir el histograma de frecuencias obtenemos una gráfica similar a la siguiente.



Al interpretar un histograma como éste podemos hacer varias observaciones, como por ejemplo:

1. El valor central de los datos está cercano a 8 ausencias pues la distribución es aproximadamente simétrica.

2. Los datos varían aproximadamente desde 1 hasta 15 ausencias en ese año.

3. La distribución no presenta valores extremos o aislados (valores que se encuentran lejos de los demás).

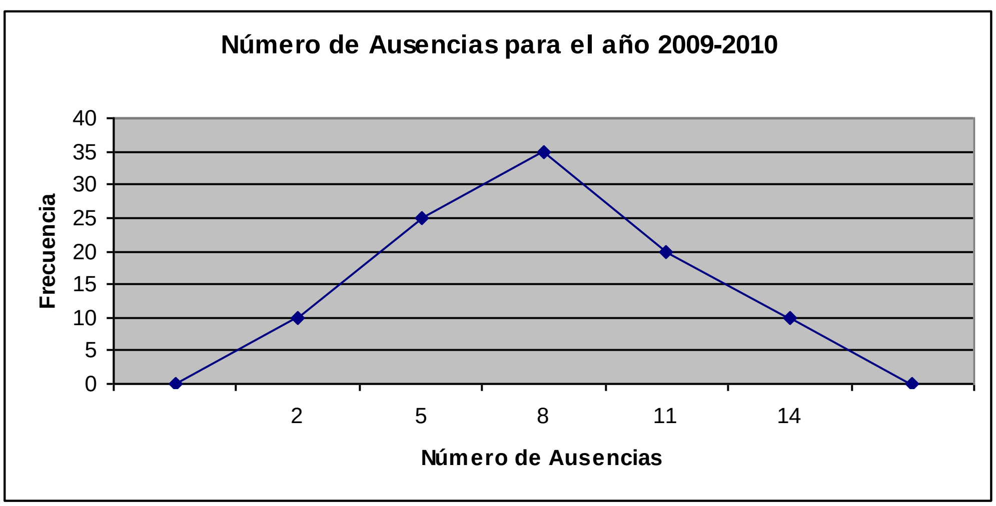
POLÍGONOS DE FRECUENCIAS

Los polígonos de frecuencias son gráficas que utilizan segmentos lineales conectando puntos localizados sobre los valores de las marcas de clase. Las alturas de estos puntos corresponden a las frecuencias o frecuencias relativas de clase. Para completar el polígono se conectan el primer punto y el último al eje horizontal en una marca anterior y posterior respectivamente con otros dos segmentos lineales.

Este tipo de gráfica se utiliza para representar datos cuantitativos de una distribución de frecuencias agrupada. Además el uso de frecuencias relativas en los polígonos de frecuencias permite comparar varios grupos de datos cuando los graficamos sobre los mismos ejes.

Ejemplo 15: Construya un polígono de frecuencias para los datos del ejemplo anterior.

Respuesta: Primero determinamos todas las marcas de clase las cuales son 2, 5, 8, 11 y 14. Luego construimos la gráfica quedando de la siguiente forma:



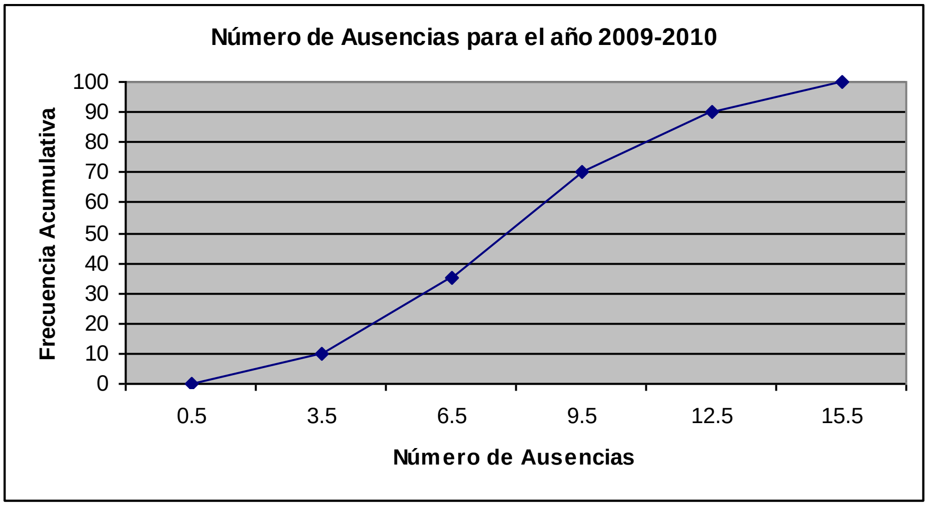
OJIVAS

Las ojivas son gráficas que utilizan segmentos lineales conectando puntos localizados sobre los valores de los límites superiores de clase. Las alturas de estos puntos corresponden a las frecuencias acumulativas de clase. Para completar la ojiva, se conecta el punto de la primera clase al eje horizontal en el límite inferior de la primera clase con un segmento lineal.

Este tipo de gráfica se utiliza para representar datos cuantitativos de una distribución de frecuencias acumuladas agrupada.

Ejemplo 16: Construya una ojiva para los datos del ejemplo anterior.

Respuesta: Primero determinamos todos los límites superiores de clase los cuales son 3.5, 6.5, 9.5, 12.5 y 15.5. Luego construimos la gráfica quedando de la siguiente forma:



GRÁFICAS DE TALLO Y HOJAS

Las gráficas de tallo y hojas se utilizan para representar datos que se separan en dos partes: el tallo (el dígito correspondiente a las decenas) y la hoja (el dígito correspondiente a las unidades). La siguiente ilustración muestra un diagrama de tallo y hojas conteniendo los años de servicio de 30 maestros. Dichas edades, ordenadas de forma creciente, son 3, 4, 5, 8, 8, 9, 10, 12, 12, 13, …, 32.

AÑOS DE SERVICIO DE LOS MAESTROS

Tallo (decenas)	Hojas (unidades)										
0	3	4	5	8	8	9					
1	0	2	2	3	3	5	5	6	7	8	9
2	0	1	2	2	3	3	5	5	6	8	8
3	0	2									

Notemos que cada dato está representado en este diagrama por sus respectivos tallo y hoja. Por ejemplo, el dato 17 tiene tallo 1 y hoja 7. Además las hojas se ubican de forma creciente de izquierda a derecha, todas con igual espacio entre ellas.

Una ventaja que nos ofrece la gráfica de tallo y hojas radica en que nos permite apreciar la distribución de los datos sin perder los datos, cosa que ocurre en las anteriores gráficas. Otra ventaja es que nos da los datos ordenados lo que en algunos procedimientos estadísticos requieren que se haga (como calcular mediana y percentiles).

Otras variantes de las gráficas de tallo y hojas se expanden utilizando más renglones en los tallos y otras se condensan utilizando menos renglones en los tallos. Véanse los siguientes ejemplos:

Versión Expandida:

AÑOS DE SERVICIO DE LOS MAESTROS						
Tallo (decenas)	Hojas (unidades)					
0	3	4				
0	5	8	8	9		
1	0	2	2	3	3	
1	5	5	6	6	8	9
2	0	1	2	2	3	3
2	5	5	6	8	8	
3	0	2				

Notemos que este diagrama tiene doble tallo para cada decena posible en los datos (excepto la decena 3 ya que no hay dato mayor que 34). En cada primer tallo anotamos las hojas correspondientes a los datos con unidades de 0 a 4 y en el segundo tallo las hojas con unidades de 5 a 9. Por ejemplo, en el tercer renglón tenemos los datos 10, 12, 12, 13 y 13, mientras que en el cuarto renglón tenemos los datos 15, 15, 16, 16, 18 y 19 todos con 1 en las decenas.

Versión Condensada:

PESO EN LIBRAS DE LOS ESTUDIANTES											
Tallo (decenas)	Hojas (unidades)										
2, 3	*	8	9								
4, 5	3	4	5	*	1	2	8	8	9		
6, 7	0	2	2	3	3						
8, 9	0	1	2	*	2	3	3	5	6	8	8

Notemos que el asterisco separa las hojas del primer tallo de las del segundo tallo. En el primer renglón tenemos que no hay datos en las 20's (libras) y en el tercer renglón no hay datos en las 70's (libras). En el segundo renglón observamos que los datos son 43, 44, 45, 51, 52, 58, 58 y 59 libras.

Ejercicios 4:

1. Una compañía automotriz realiza un sondeo a 100 compradores de autos nuevos, seleccionados al azar, para tener una idea de la preferencia en las marcas de autos en el mercado de Puerto Rico para vehículos de fabricación japonesa con valor menor a los $30,000. La siguiente tabla resume los hallazgos.

VENTA DE VEHÍCULOS EN PUERTO RICO

Marca	Frecuencia
Suzuki	15
Mitsubishi	20
Honda	13
Toyota	27
Subaru	2
Nissan	12
Mazda	10
Isuzu	1

Construya una gráfica de barras para estos datos.

2. A un grupo de 20 personas se les preguntó cuál es el color de su auto. Las respuestas fueron: blanco, rojo, verde, azul, blanco, verde, negro, rojo, rojo, blanco, azul, blanco, rojo, azul, negro, amarillo, azul, rojo, blanco y blanco. Construya una gráfica circular para estos datos.

3. La siguiente distribución de frecuencias resume las puntuaciones obtenidas por 60 estudiantes en el examen final de matemáticas.

PUNTUACIONES EN EL EXAMEN FINAL DE MATEMÁTICAS

Puntuación	Frecuencias Acumulativas
26 – 40	7
41 – 55	5
56 – 70	23
71 – 85	17
86 – 100	8

Construya para estos datos:

a. un histograma de frecuencias

b. un polígono de frecuencias

c. una ojiva

MEDIDAS DE TENDENCIA CENTRAL

Uno de los estadísticos básicos que se utiliza al describir, explorar y comparar las características fundamentales de un conjunto de datos es el valor central (o centro). Una ***medida de tendencia central*** es un valor que se encuentra en el centro de un conjunto de datos. Hay varias formas distintas de determinar el valor central de un conjunto de datos. A continuación discutiremos las más importantes.

Media Aritmética - La ***media aritmética*** de un conjunto finito de datos cuantitativos es la medida de tendencia central que se calcula al sumar todos los datos y dividir el total de la suma entre el número total de datos. La media aritmética es la medida más importante y es la más que se utiliza para describir datos pues es la que menos varía con respecto a las muestras. Por lo tanto, la utilizaremos con frecuencia y en adelante nos referiremos a ella simplemente como la ***media***.

La definición anterior de la media puede expresarse con las fórmulas que presentaremos a continuación. Utilizamos la letra griega Σ (sigma) para indicar una sumatoria de datos, $\boldsymbol{n}$ representa el número total de datos de una muestra y $\boldsymbol{N}$ el número total de datos de una población. Aunque la fórmula para determinar la media es básicamente la misma en muestras o poblaciones, el símbolo que se usa para denotarlas es distinto. Para la media muestral utilizamos $\bar{x}$ (y se lee "x barra") y para la media poblacional utilizamos la letra griega μ (y se lee "micro").

Media Muestral: $\bar{x} = \dfrac{\sum x}{n}$ (estadístico)

Media Poblacional: $\mu = \dfrac{\sum x}{N}$ (parámetro)

Ejemplo 17: A todos los 10 maestros de una pequeña escuela se les preguntó cuántos hermanos tienen y éstas fueron sus respuestas: 2, 1, 0, 3, 2, 2, 1, 1, 0 y 2. Halle la media para este conjunto de datos.

Respuesta: Notemos que es una población, por lo tanto

$$\mu = \frac{\sum x}{N} = \frac{2+1+0+3+2+2+1+1+0+2}{10} = \frac{14}{10} = 1.4$$

La media de los hermanos de estos maestros es 1.4 hermanos. Esto no quiere decir que es posible tener 1.4 hermanos sino que el "promedio" de hermanos que tienen estos maestros es 1.4.

Ejemplo 18: Un investigador desea saber la media de los salarios anuales de los empleados de una pequeña compañía. Para conseguirla toma una muestra aleatoria de 5 empleados y les pregunta su salario anual. Estos fueron sus datos: $25,000 ; $50,000 ; $75,000 ; $225,000 ; $25,000. ¿Cuál fue el resultado del investigador?

Respuesta: Notemos que es una muestra, por lo tanto

$$\bar{x} = \frac{\sum x}{n} = \frac{\$25,000 + \$50,000 + \$75,000 + \$225,000 + \$25,000}{5}$$

$$= \frac{\$400,000}{5} = \$80,000$$

La media de los salarios anuales de los empleados de esta compañía es $80,000.

Notemos que la media en este caso está por encima de la mayoría de los datos. Esto ocurre porque está siendo afectada por el dato aislado $225,000. De hecho, como la media toma en consideración todos los datos, es afectada por los datos extremos o aislados.

Cuando trabajamos con datos resumidos en una distribución de frecuencias, la media se denomina ***media ponderada*** y hay dos casos posibles: distribuciones de frecuencias no agrupadas (un solo dato posible por clase) y agrupadas (más de un dato posible por clase), entonces las fórmulas cambian. Veamos cada uno de estos casos.

Caso 1: Distribuciones de frecuencias no agrupadas

$$\bar{x} = \frac{\sum (x \cdot f)}{\sum f} \quad \text{o} \quad \mu = \frac{\sum (x \cdot f)}{\sum f}$$

donde $x \cdot f$ representa el producto de cada dato con su respectiva frecuencia, $\sum (x \cdot f)$

representa la suma de todos los productos $x \cdot f$ y $\sum f$ representa el número total de datos.

Ejemplo 19: Halle la media aritmética para la siguiente distribución de frecuencias que resume los datos obtenidos al tomar una muestra de 20 estudiantes y preguntarles el número de hermanos que tienen.

NÚMERO DE HERMANOS DE LOS ESTUDIANTES

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

Respuesta: Utilizando la tabla provista identificamos que en la primera columna (número de hermanos) se encuentran todos los posibles valores de los datos ($\boldsymbol{x}$), mientras que en la segunda columna tenemos sus respectivas frecuencias ($\boldsymbol{f}$). Entonces añadimos una tercera columna conteniendo los productos de cada dato $\boldsymbol{x}$ con su respectiva frecuencia $\boldsymbol{f}$, es decir $x \cdot f$. Finalmente sumamos todos los valores de $\boldsymbol{f}$ para hallar $\sum f$ y también sumamos todos los productos $x \cdot f$ para hallar $\sum (x \cdot f)$. Véase la siguiente tabla:

NÚMERO DE HERMANOS DE LOS ESTUDIANTES

Número de Hermanos (x)	Frecuencia (f)	$x \cdot f$
0	5	0
1	7	7
2	4	8
3	2	6
4	1	4
5	0	0
6	1	6
	$\sum f = 20$	$\sum (x \cdot f) = 31$

Luego la media será: $$\bar{x} = \frac{\Sigma(x \cdot f)}{\Sigma f} = \frac{31}{20} = 1.55 .$$

Caso 2: Distribuciones de frecuencias agrupadas

Cuando trabajamos con datos agrupados en una distribución de frecuencias, no sabemos con exactitud los datos que tenemos. Para poder hallar la media suponemos que todos los datos de cada clase son las marcas de clases. Entonces,

$$\bar{x} = \frac{\Sigma(x_i \cdot f)}{\Sigma f} \quad \text{o} \quad \mu = \frac{\Sigma(x_i \cdot f)}{\Sigma f}$$

donde $x_i \cdot f$ representa el producto de cada marca de clase con su respectiva frecuencia y Σf representa el número total de datos.

Ejemplo 20: Halle la media aritmética para la siguiente distribución de frecuencias que resume los datos obtenidos al tomar una población de 20 peces de una pecera y medir sus pesos en onzas.

PESO DE LOS PECES DE UNA PECERA

Peso en Onzas	Frecuencia
0 – 2	2
3 – 5	4
6 – 8	8
9 – 11	5
12 – 14	0
15 – 17	1

Respuesta: Utilizando la tabla provista identificamos que en la primera columna (peso en onzas) se encuentran todas las clases y añadimos una columna al lado de ésta conteniendo las marca de clase de cada una de las clases (estos son los valores de x_i). En la segunda columna de la tabla anterior tenemos sus respectivas frecuencias (f). Esta columna la estamos colocando como tercera en la siguiente tabla. Entonces añadimos una cuarta columna conteniendo los productos de cada marca de clase x_i con su respectiva frecuencia f, es decir $x_i \cdot f$. Finalmente sumamos todos los valores de f para hallar $\sum f$ y también sumamos todos los productos $x_i \cdot f$ para hallar $\sum (x_i \cdot f)$. Véase la siguiente tabla:

PESO DE LOS PECES DE UNA PECERA

Peso en Onzas	Marca de Clase (x_i)	Frecuencia (f)	$x_i \cdot f$
0 – 2	1	2	2
3 – 5	4	4	16
6 – 8	7	8	56
9 – 11	10	5	50
12 – 14	13	0	0
15 – 17	16	1	16
		$\sum f = 20$	$\sum (x_i \cdot f) = 140$

Luego la media será: $\mu = \dfrac{\sum (x_i \cdot f)}{\sum f} = \dfrac{140}{20} = 7$.

Recorrido Medio - El ***recorrido medio*** de un conjunto finito de datos cuantitativos es la medida de tendencia central que se calcula al sumar el dato mayor y el dato menor y dividimos el total de la suma entre dos.

$$\text{Recorrido Medio} = \frac{\text{Dato Mayor} + \text{Dato Menor}}{2}$$

Esta medida de tendencia central tiene la ventaja de ser sencilla y fácil de calcular pero tiene la desventaja de que es la medida que más se afecta con los datos extremos.

Ejemplo 21: El mismo investigador desea saber el recorrido medio de los salarios anuales de los empleados de la misma pequeña compañía anterior. Para conseguirlo utiliza los mismos datos de su muestra de 5 empleados:

$25,000 $50,000 $75,000 $225,000 $25,000.

¿Cuál fue el resultado del investigador?

Respuesta:

$$\text{El recorrido medio} = \frac{\$225{,}000 + \$25{,}000}{2} = \frac{\$250{,}000}{2} = \$125{,}000$$

Moda - La ***moda*** de un conjunto finito de datos es el dato con mayor frecuencia. Cuando dos datos tienen la misma frecuencia y ésta es la más alta, ambos datos son modas, por lo que el conjunto de datos es ***bimodal***. Si más de dos datos tienen la misma frecuencia y ésta es la más alta, todos estos datos son modas, por lo que el conjunto de datos es ***multimodal***. Y cuando ningún dato se repite, se dice que no hay moda.

Ejemplo 22: Halle la moda para cada uno de los siguientes conjuntos de datos:

a.	0.3	4.1	9.6	4.1	1.1	0.3	4.1		
b.	3	3	3	4	4	4	5	5	7
c.	1	2	3	4	6	8	9		
d.	rojo	azul	rojo	negro	azul	rojo			

Respuestas:

a. La moda es 4.1 porque es el dato con mayor frecuencia.

b. Los números 3 y 4 son modas, ya que ambos datos tienen igual frecuencia y ésta es la más alta.

c. No hay moda.

d. La moda es rojo porque es el dato con mayor frecuencia.

La moda tiene la ventaja de que es la única medida de tendencia central que puede ser utilizada para conjuntos de datos cualitativos. Pero la desventaja que tiene es que no siempre existe y no siempre es única. Por lo que, si necesitamos una medida del centro de los datos y usamos la moda podría no existir o tener múltiples centros, lo que en muchas ocasiones no es deseable (por ejemplo el salario promedio anual de $20,000 y $60,000).

Cuando los datos están resumidos en distribución de frecuencias agrupada, le llamamos la ***clase modal*** a aquella clase con mayor frecuencia.

Ejemplo 23: Halle la moda para cada uno de los siguientes conjuntos de datos resumidos en distribuciones de frecuencias.

a. **PESO DE LOS PECES DE UNA PECERA**

Peso en Onzas	Frecuencia
0 – 2	2
3 – 5	4
6 – 8	8
9 – 11	5
12 – 14	0
15 – 17	1

b. **NÚMERO DE HERMANOS DE LOS ESTUDIANTES**

Número de Hermanos	Frecuencia
0	5
1	7
2	4
3	2
4	1

Respuestas:

a. La clase modal es la segunda, 6 – 8, porque es la clase con mayor frecuencia.

b. La moda es 1 porque es el dato con mayor frecuencia.

Mediana - La ***mediana*** de un conjunto finito de datos cuantitativos ordenados es aquel valor que divide al conjunto en dos partes iguales, de forma que el número de datos mayor o igual a la mediana es igual al número de datos menores o iguales a ésta. Si un conjunto de datos ordenados tiene n elementos, entonces la mediana es el valor que estará en la posición $\frac{n+1}{2}$ cuando los datos originales se presentan en orden creciente o decreciente. Lo anterior implica que la mediana es un valor tal que el 50% de los datos es mayor que él y por ende el otro 50% es menor. La mediana suele denotarse con $\tilde{x}$.

La mediana tiene la ventaja de que es la medida de tendencia central que menos se afecta por datos extremos o aislados. Esto se debe a que los valores que afectan a la mediana son los centrales, cuando los datos están ordenados, y no los valores extremos que suelen estar por encima o por debajo de los restantes datos.

Para calcular la mediana de un conjunto de datos primero ordenamos los datos de forma creciente o decreciente y luego realizamos uno de los siguientes dos procedimientos:

1. Si el número total de datos es impar, la mediana es el dato que se encuentra exactamente en la mitad de la lista.
2. Si el número total de datos es par, la mediana se obtiene sumando los dos datos que se encuentran exactamente en la mitad de la lista y dividiendo la suma entre dos.

Ejemplo 24: Halle la mediana del siguiente conjunto de datos:

5 2 6 9 11 17 20 5 4

Respuesta: Notemos que los datos no están ordenados por lo tanto primero los ordenamos.

2 4 5 5 6 9 11 17 20

Luego como el número total de datos es impar (9), tenemos que la mediana es 6 (el dato en el centro de la lista ordenada).

Ejemplo 25: Halle la mediana del siguiente conjunto de datos:

12 3 24 19 8 15

Respuesta: Notemos que los datos no están ordenados por lo tanto primero los ordenamos.

3 8 12 15 19 24

Luego como el número total de datos es par (6), tenemos que la mediana es $\frac{12+15}{2} = 13.5$ (la suma de los dos datos centrales divida entre 2).

Notemos que en ambos ejemplos, el valor de la mediana divide la lista ordenada de datos en dos partes con igual número de datos (valor intermedio).

La extensión para el cálculo de la mediana en el caso de datos agrupados es realiza a continuación:

$$\tilde{x} = \mathrm{L}_{\inf} + \frac{\frac{n}{2} - f_{acum}}{f_{med}} \cdot A$$ donde

$L_{\inf}$ = es el límite inferior de la clase donde cae la mediana

n = número total de datos

f_{acum} = la frecuencia acumulativa de la clase anterior

f_{med} = la frecuencia de la clase donde cae la mediana

A = ancho de clase de la distribución

Ejemplo 26: Halle la mediana para las edades en la siguiente distribución de frecuencias que resume las edades de los pacientes atendidos en una clínica durante un fin de semana.

EDADES DE PACIENTES ATENDIDOS EN UNA CLÍNICA

Años cumplidos de los pacientes	Frecuencia
10 – 19	6
20 – 29	10
30 – 39	20
40 – 49	14
50 – 59	3
60 – 69	2

Respuesta: Retomemos la tabla del ejemplo mostrado para determinar la mediana de las edades de los atendidos por la clínica, añadiéndole la columna de la frecuencia acumulativa.

EDADES DE PACIENTES ATENDIDOS EN UNA CLÍNICA

Años cumplidos de los pacientes	Frecuencia	Frecuencia Acumulativa
10 – 19	6	6
20 – 29	10	16
30 – 39	20	36
40 – 49	14	50
50 – 59	3	53
60 – 69	2	55

Notemos que el número total de datos es 55, por lo tanto, la mediana es el dato de la posición $\frac{n+1}{2} = \frac{55+1}{2} = 28$. Esta posición cae en la tercera clase ya que hasta la segunda clase tenemos los primeros 16 datos y en la tercera están los siguientes 20 datos. Entonces tenemos que:

$L_{\inf} = 30$ (es el límite inferior de la tercera clase)

$n = 55$ (número total de datos)

$f_{acum} = 16$ (la frecuencia acumulativa de la segunda clase)

$f_{med} = 20$ (la frecuencia de la tercera clase)

$A = 10$ (ancho de clase de la distribución)

Por lo tanto, $\tilde{x} = L_{\inf} + \dfrac{\frac{n}{2} - f_{acum}}{f_{med}} \cdot A = 30 + \dfrac{\frac{55}{2} - 16}{20} \cdot 10 = 35.75$

Entonces podemos concluir que el 50% de las personas atendidas en esta clínica ese fin de semana tienen una edad inferior a los 35.75 años.

Resumen de las Ventajas y Desventajas de las Medidas de Tendencia Central

Hasta ahora hemos calculado la media, recorrido medio, moda y mediana como medidas de tendencia central. Pero, ¿cuál de ellas es mejor? La verdad es que no hay una respuesta única a esta pregunta pues no hay criterios objetivos para determinar la medida de tendencia central más representativa para todos los conjuntos de datos. Las diferentes medidas de tendencia central ofrecen diversas ventajas y desventajas, algunas de las cuales son presentadas en la siguiente tabla.

VENTAJAS Y DESVENTAJAS DE LAS MEDIDAS DE TENDENCIA CENTRAL

Medida	Ventaja(s)	Desventaja(s)	Cuándo conviene usarla
Media	**1.** varía menos con las muestras **2.** es la más utilizada	Es afectada por datos extremos o aislados	cuando los datos son cuantitativos y su distribución es simétrica
Recorrido Medio	fácil de calcular	la que más se afecta con los datos extremos o aislados	cuando los datos son cuantitativos y no hay datos extremos
Moda	la única que puede ser utilizada con datos cualitativos	**1.** no es siempre única **2.** no siempre existe	cuando el dato de mayor frecuencia indique una buena medida para el centro de los datos o los datos sean cualitativos
Mediana	es la que menos se afecta con los datos extremos o aislados	hay que ordenar los datos para hallarla, lo cual podría tomar tiempo cuando son muchos datos	cuando los datos son cuantitativos y su distribución no es simétrica

Ejercicios 5:

1. Considere el siguiente conjunto de datos muestrales:

   {2, 4, 3, 5, 3, 6, 9, 7, 2, 3}

   Halle para este conjunto de datos:

   a. la media
   b. el recorrido medio
   c. la moda
   d. la mediana

2. Considere la siguiente distribución de frecuencias que resume los datos obtenidos al tomar una muestra de 25 personas adultas y preguntarles el número de autos nuevos que haya comprado en su vida.

NÚMERO DE AUTOS NUEVOS COMPRADOS

Número de Autos	Frecuencia
0	3
1	10
2	8
3	2
4	1
5	0
6	1

Halle para este conjunto de datos:

a. la media
b. el recorrido medio
c. la moda
d. la mediana

3. Considere la siguiente distribución de frecuencias que resume los pesos en libras de una población de 30 estudiantes de música.

PESO DE ESTUDIANTES DE MÚSICA

Peso en libras	Frecuencia
100 – 119	2
120 – 139	7
140 – 159	8
160 – 179	9
180 – 199	3
200 – 219	1

Halle para este conjunto de datos:

a. la media

b. la clase modal

c. la mediana

RESPUESTAS A LOS EJERCICIOS

Ejercicios 1:

1. a. Los 3,600 estudiantes de esa universidad

   b. i. variable de interés = área de proveniencia isla o metro

      ii. muestra = 400 estudiantes

      iii. parámetro = 37% de los estudiantes provenientes de área metro

      iv. estadístico = 34% de los estudiantes provenientes de área metro

      v. tipo de datos = cualitativos

2. a. peso en libras de una persona – cuantitativa continua
   b. color de auto - cualitativa
   c. cantidad de dinero en cuenta de ahorro – cuantitativa discreta
   d. deporte favorito - cualitativa
   e. créditos matriculados por semestre – cuantitativa discreta
   f. temperatura actual en grados Fahrenheit – cuantitativa continua

Ejercicios 2:

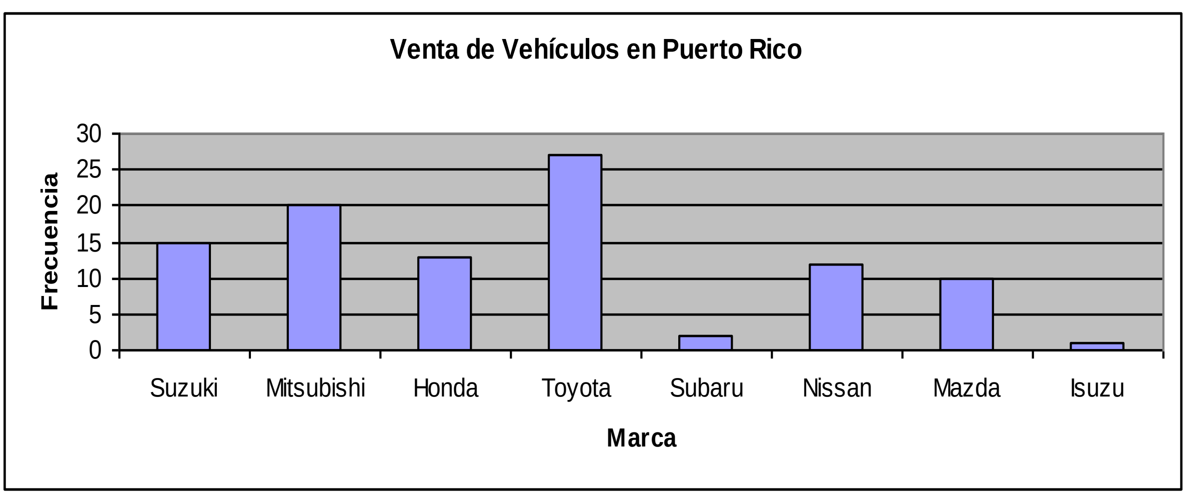
1. b 2. a 3. c 4. d

Ejercicio 3: **AÑOS DE SERVICIO DE MAESTROS**

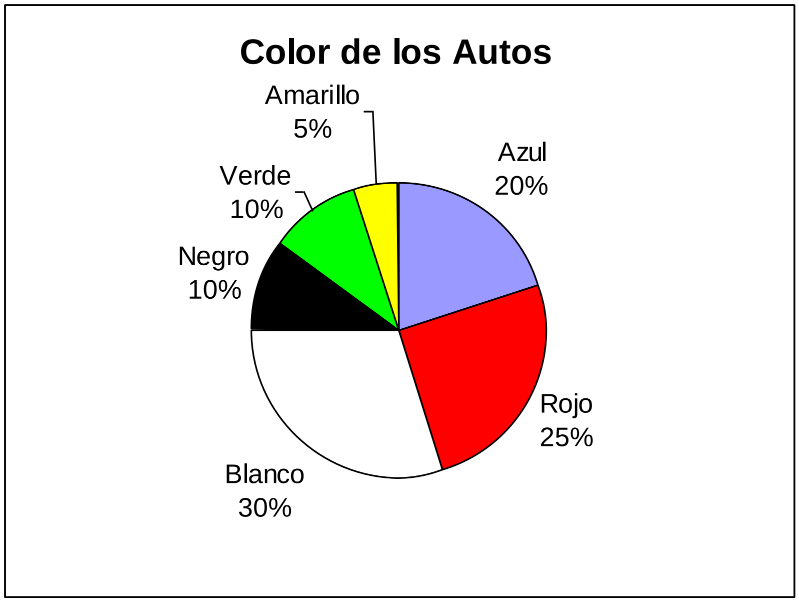
Años de Servico	Frecuencia	Frecuencia Relativa	Frecuencia Acumulativa
3 – 8	5	$\frac{5}{30}$	5
9 – 14	6	$\frac{6}{30}$	11
15 – 20	7	$\frac{7}{30}$	18
21 – 26	8	$\frac{8}{30}$	26
27 – 32	4	$\frac{4}{30}$	30

Ejercicios 4:

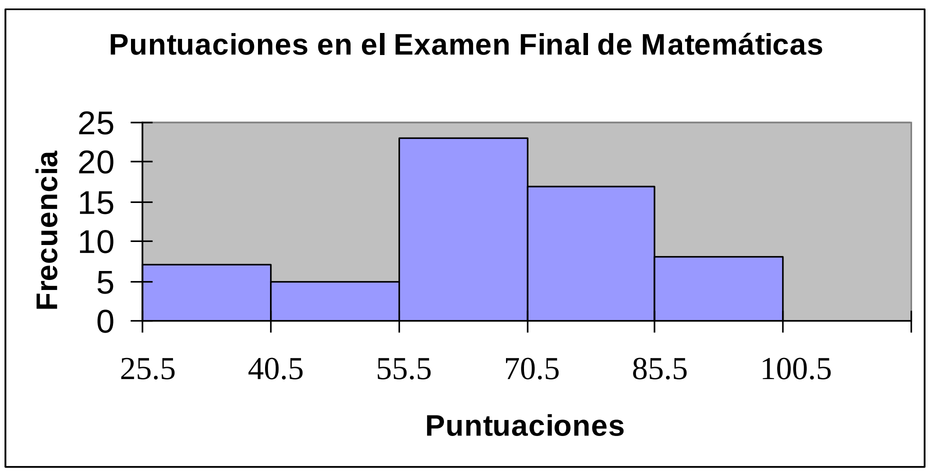
1.



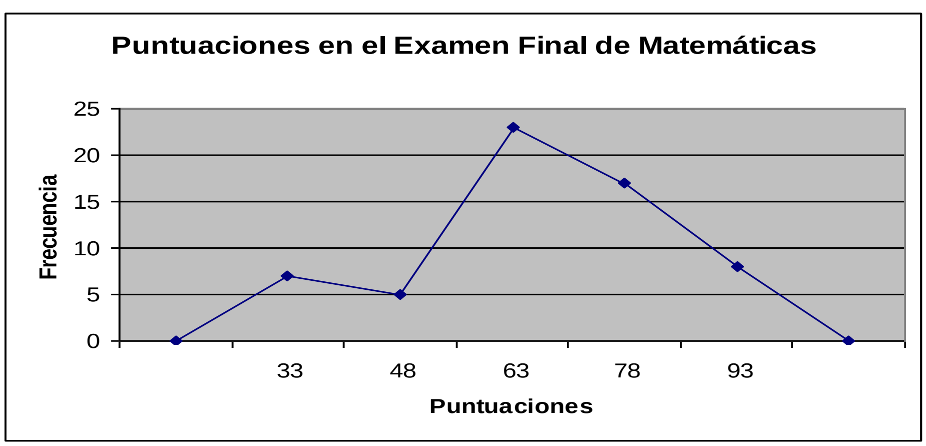
2.


Color de los Autos
Amarillo
5%
Azul
20%
Verde
10%
Negro
10%
Rojo
25%
Blanco
30%

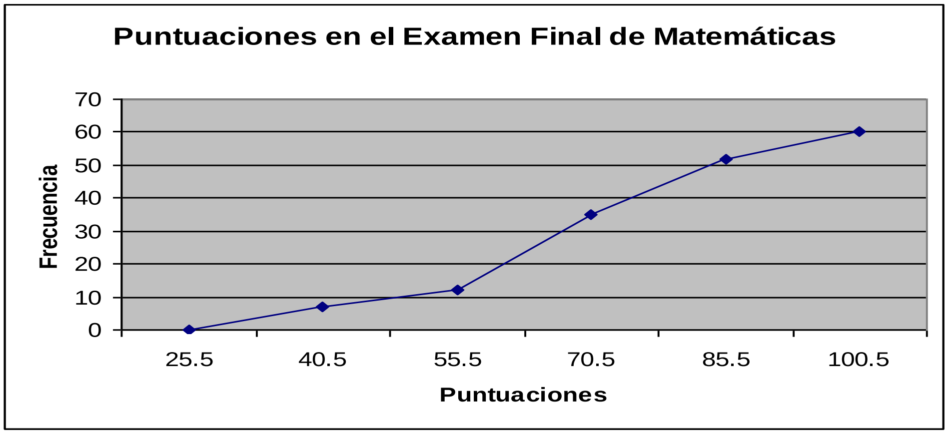
3. a.


Puntuaciones en el Examen Final de Matemáticas
Frecuencia
25
20
15
10
5
0
25.5
40.5
55.5
70.5
85.5
100.5
Puntuaciones

b.


Puntuaciones en el Examen Final de Matemáticas
Frecuencia
25
20
15
10
5
0
33
48
63
78
93
Puntuaciones

c.



Ejercicios 5:

1. a. $\bar{x} = 4.4$

   b. el recorrido medio = 5.5

   c. la moda = 3

   d. la mediana = 3.5

2. a. $\bar{x} = 1.68$

   b. el recorrido medio = 3

   c. la moda = 1

   d. la mediana = 1

3. a. $\mu = 154.167$

   b. la clase modal es la cuarta: 160 - 179

   c. la mediana = 155

POS-PRUEBA

Resuelva cada uno de los siguientes ejercicios y escoja la alternativa correcta:

1. La rama de la estadística que trata sobre las técnicas de colección, ordenación, clasificación y presentación de los datos es la estadística:

   a. descriptiva
   b. inferencial
   c. subjetiva
   d. probabilística

2. La variable, *cantidad de pupitres por salón*, es:

   a. Cualitativa
   b. Cuantitativa continua
   c. Cuantitativa discreta
   d. Ninguna de las anteriores

3. Una maestra quiere hacer un estudio sobre todos los estudiantes que ha tenido durante sus veinte años en dicha profesión. Ella decide seleccionar al azar 5 grupos completos de estudiantes de los 20 grupos que ha tenido. Este método de muestreo se le conoce como:

   a. Estratificado
   b. Por conglomerado
   c. Sistemático
   d. Por tómbola

Considere la siguiente distribución de frecuencias para contestar las preguntas 4 y 5.

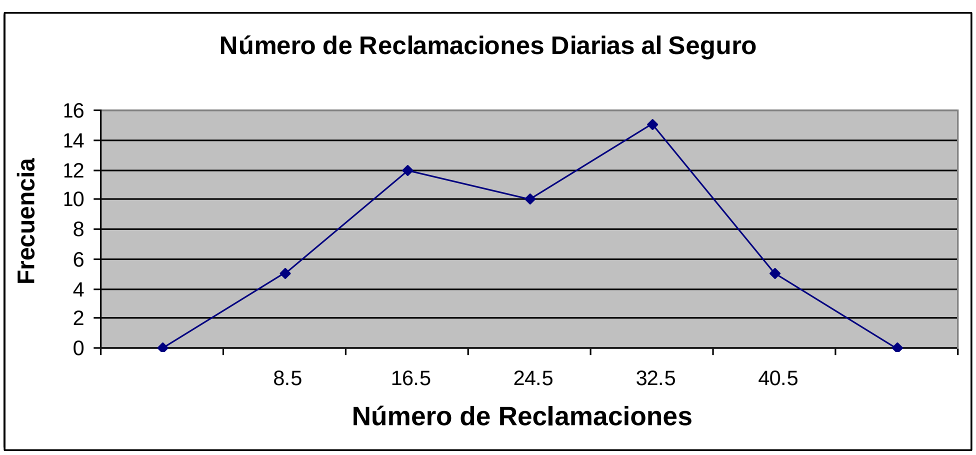
Edades de Ganadores del Oscar

Edad	Frecuencias
20 – 29	1
30 – 39	5
40 – 49	7
50 – 59	9
60 – 69	1

4. El ancho de clases de esta distribución es:
   a. 5
   b. 9
   c. 10
   d. 49

5. El límite superior de la segunda clase es:
   a. 30
   b. 35.5
   c. 39
   d. 39.5

Considere la siguiente gráfica para contestar las preguntas 6 y 7.



6. Esta gráfica se conoce como

a. ojiva

b. tallo y hojas

c. histograma

d. polígono de frecuencias

7. Los valores 8.5, 16.5, 24.5, 32.5 y 40.5 son:

a. límites de clases

b. marcas de clases

c. frecuencias de clases

d. fronteras de clases

8. La mediana en el siguiente conjunto de datos {3, 0, 4, 8, 4, 1, 2} es:

a. 3
b. 4
c. 5
d. 8

9. La moda de los datos resumidos en la siguiente distribución de frecuencias es:

a. 2
b. 4
c. 7
d. 11

Juegos Ganados por mes	5	6	7	8	9	10	11
Frecuencia	1	2	4	3	2	2	1

10. La medida de tendencia central que menos se afecta con los datos extremos o aislados es:

a. la media

b. el recorrido medio

c. la moda

d. la mediana

RESPUESTAS DE LA PRE-PRUEBA

1. b	6. a
2. b	7. d
3. a	8. b
4. c	9. c
5. b	10. b

RESPUESTAS DE LA POS-PRUEBA

1. a	6. d
2. c	7. b
3. b	8. a
4. c	9. c
5. c	10. d

REFERENCIAS

1. Estadística, Mario F. Triola, Décima Edición, 2009, Addison Wesley
2. Fundamentals of Statistics, Michael Sullivan, Third Edition, 2011, Prentice Hall