

Estadística descriptiva.

La estadística descriptiva es la parte de la estadística encargada de organizar y resumir conjuntos de datos numéricos en tablas o gráficos para poner en evidencia características y comportamientos que se puedan presentar como frecuencias, porcentajes, promedios, etc. sin extraer conclusiones (inferencias) sobre la población a la que pertenecen.

Se definirán algunas técnicas descriptivas básicas, con el fin de comprender la elaboración de tablas de frecuencias, la construcción de gráficos, las principales medidas de tendencia central y la forma que estas permitirán realizar la descripción de determinado conjunto de datos.

Por lo general, en cualquier análisis estadístico se pretende sacar conclusiones de un conjunto de interés llamado población; aunque a veces conocerla y analizarla totalmente es difícil por razones de costo y tiempo.

Población N: es el conjunto de todos los elementos que tienen una característica común para ser estudiados. La población puede estar definida por personas o cosas y también ser finita o infinita.

Muestra: es un subconjunto (representativo) de la población que se toma con el fin de extraer información.

Ejemplo:

Halle la población y una muestra de la siguiente información: Se desea realizar un estudio estadístico con algunas personas del municipio de Bello, acerca de la necesidad o no del pico y cédula, para las personas que deseen salir a comprar productos básicos.

Solución:

Dado que la población es el conjunto de estudio más general y grande posible, en este caso será representado por las personas del municipio de Bello que tienen cédula de ciudadanía y la muestra en este caso es un subconjunto de la población, es decir, personas del municipio de Bello (que tienen cédula de ciudadanía) con las que se realizó la encuesta sobre el pico y cédula.

Para poder aplicar alguna técnica estadística es necesario definir el tipo de variable con que se está trabajando.

Variable estadística: son las características consideradas con el propósito de describir cada uno de los individuos de la muestra.

Tipos de variables: se distinguen dos tipos de variables. Las cuantitativas y las cualitativas, las primeras hacen referencia aquellas que se pueden representar por medio de cantidades numéricas, por ejemplo, el peso, la estatura, el número asignaturas en una Universidad, entre otras. El segundo tipo de variable hace referencia a las categorías que no se pueden representar por medidas numéricas, por ejemplo, el sexo la raza, color de ojos, una calificación entre otras.

Distribuciones de frecuencia: son utilizadas para resumir la información a partir de una muestra de datos, su construcción es muy sencilla, pero en conjuntos de datos de gran tamaño el cálculo puede resultar tedioso.

Representaciones gráficas

En la estadística descriptiva existen diferentes graficas para representar la información de forma resumida e informativa en datos agrupados. El siguiente es un paso a paso, para elaborar una tabla de frecuencia, pero antes es importante tener en cuenta la terminología que se utilizará:

- N : tamaño de la población
- n : tamaño de la muestra
- x_i : característica cuantitativa observada en cada unidad investigada.
- f_i frecuencia absoluta: número de veces que se repite el valor de la variable.
- fr_i frecuencia relativa: se obtiene dividiendo cada frecuencia absoluta por el tamaño de la muestra o el tamaño de la población.
- F_i frecuencia absoluta acumulada: se hace por medio de sumas sucesivas de manera que la última frecuencia absoluta acumulada será igual a n .
- Fr_i : frecuencia relativa acumulada, para calcularla se procede igual que en frecuencia relativa.
- m : número de valores que toma la variable.



Ejemplo:

se tienen 300 cajas de madera en la bodega de una tienda y cada una de ellas contiene figuras en vidrio. El dueño del almacén le interesa examinar las cajas con el fin de saber el número de figuras que han sufrido alguna avería en el transporte desde la fábrica hasta la bodega. Pero por motivos de personal disponible, espacio físico y tiempo, se tomó la decisión de examinar una muestra equivalente al 10% de las cajas, por tal razón se tendrá 30 cajas de un total de 300. El siguiente recuadro contiene la información sobre dichas cajas, donde cada número indica la cantidad de figuras que han sufrido o no alguna avería:

1	2	1	2	2	2
2	3	2	3	2	3
1	1	3	0	1	2
0	2	0	3	4	1
3	4	2	2	2	3

Solución:

En este caso se puede definir x_i como el número de figuras que han sufrido o no alguna avería en cada caja, y dado que la cantidad de figuras que han sufrido o no alguna avería es 0, 1, 2, 3, o 4, entonces, $m=5$. Ahora, en función de esto armamos la tabla de frecuencia.

x_i	f_i	fr_i	F_i	Fr_i
0	3	$3/30 = 0,10$	3	0,10
1	6	$6/30 = 0,20$	9	0,30
2	12	$12/30 = 0,40$	21	0,70
3	7	$7/30 = 0,23$	28	0,93
4	2	$2/30 = 0,07$	30	1,00
Total	30	1,00		

A partir de la tabla de frecuencias construida se puede generalizar el resultado analizando las frecuencias relativas, donde la información indica que el 10% de las figuras no sufrieron averías.

Diagrama de barras o histograma: este tipo de gráfica es utilizado tanto para variables cuantitativas como cualitativas. Para construir este diagrama se utilizan dos ejes, en el eje horizontal se representan las categorías de la variable que se quiere representar y en el eje vertical se levantan las barras cuyas alturas son proporcionales a la frecuencia (absoluta o relativa) de cada dato.

Ejemplo:

Construir el diagrama de barras de acuerdo a la siguiente información: A un grupo de 20 personas se les preguntó acerca de su color favorito. Los resultados se muestran en la siguiente tabla de frecuencias:

Color	f_i
Negro	4
Azul	5
Amarillo	5
Rojo	6

Solución:

En este caso las variables del problema son de tipo cualitativo, para la elaboración del gráfico de barras se coloca en el eje horizontal los colores y en el eje vertical el valor de cada frecuencia absoluta, como se ilustra a continuación:

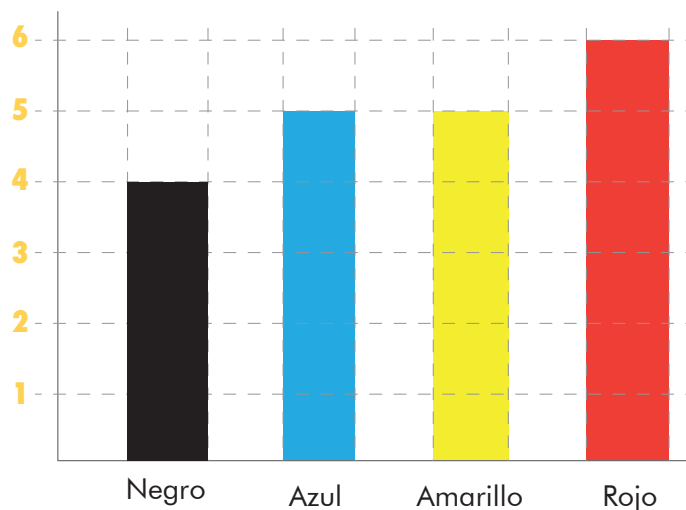


Diagrama de sectores: este diagrama también se utiliza para representar variables cuantitativas. En este se representan los datos y su frecuencia de manera que el círculo sea dividido de forma proporcional a la frecuencia de cada dato.

Ejemplo:

Representar en un diagrama circular la siguiente información:

En la siguiente tabla, se ilustra porcentualmente los fondos que destina la gobernación de cierto departamento anualmente en deporte, cultura, sociedad, tecnología, entre otros.

Sector	$fr_i(\%)$
Deporte	22.1
Cultura	20.3
Sociedad	19.9
Tecnología	17.9
Otros	19.8

Solución:

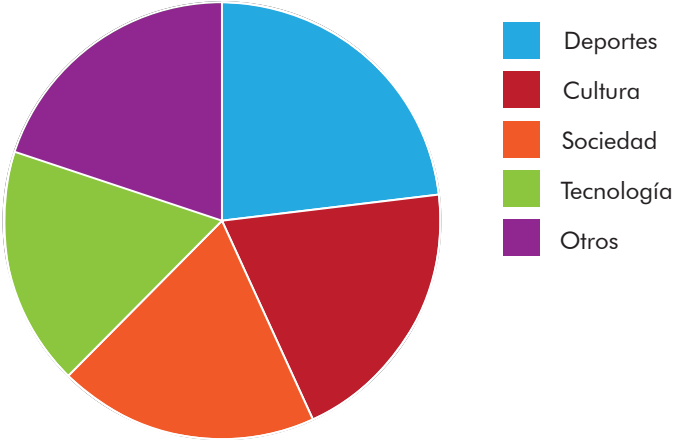
Para construir el diagrama circular, se debe tomar el valor de cada frecuencia relativa, multiplicarla por 360 y dividirla por 100, para conocer el ángulo de cada sector circular

$$\frac{Fr_i \times 360}{100}$$

Sector	Angulo
Deporte	79,56°
Cultura	73,1°
Sociedad	71,6°
Tecnología	64,4°
Otros	71,2°

Así, el diagrama circular es de la forma:

Visitas a contenidos



Ojiva: también conocido como diagrama acumulativo de frecuencias, se construye tomando en el plano los puntos (c_i, N_i) y uniéndolos con segmentos horizontales y verticales, de forma que se obtiene una función escalonada.

Comentario: si se utilizan las frecuencias relativas acumuladas, el valor máximo del diagrama es 1 y cuando se construye con las frecuencias absolutas acumuladas, el máximo será el total de datos de la muestra.

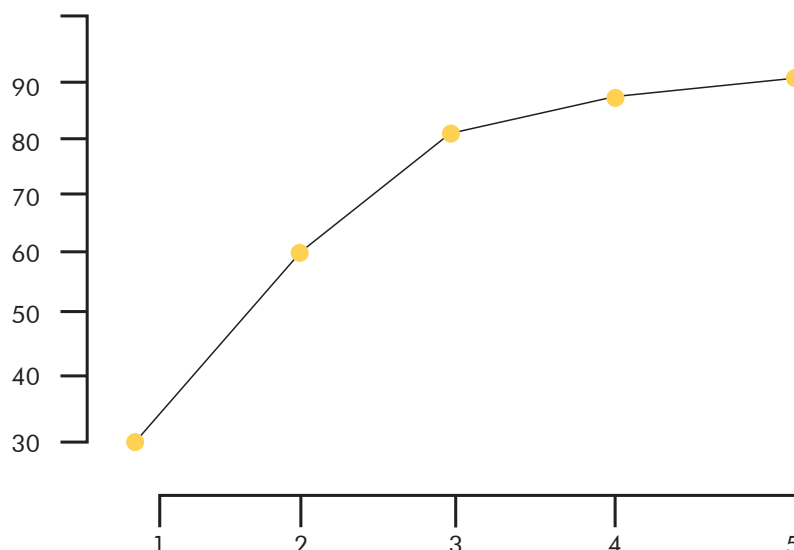
Ejemplo:

hallar el polígono de frecuencia de la siguiente información: El porcentaje de número de hijos por cada 100 familias en cierta ciudad se representa en la siguiente tabla:

Nº de hijos	$fr_i(\%)$	$FR_i(\%)$
1	30	30
2	35	65
3	15	80
4	10	90
5	10	100

Solución:

Para realizar la gráfica, se coloca en el eje horizontal el número de hijos y en el eje vertical el porcentaje de la frecuencia acumulada relativa. La representación gráfica queda de la siguiente manera:



Medidas de tendencia central: son las medidas que se ubican alrededor de los datos observados de la muestra, se distinguen la media aritmética, la mediana y la moda.

Media aritmética: es el promedio de todos los datos de la muestra

$$\bar{X} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}$$

donde x_i es valor de cada dato y n es el tamaño de la muestra. En el caso de datos agrupados en tablas la media se puede calcular por medio de la expresión:

$$\bar{X} = \frac{\sum_{i=1}^n f_i \cdot x_i}{n}$$

donde x_i es valor de cada dato, f_i es la frecuencia absoluta de cada dato y n es el tamaño de la muestra.

Ejemplo:

Calcular la media de 8, 16, 4, 12 y 10.

Solución:

Se tienen 5 datos, así la media de estos datos es:

$$\bar{X} = \frac{8+16+4+12+10}{5} = 10$$

Mediana: si los datos están ordenados de mayor a menor, la mediana es el valor hasta el cual se

encuentran el 50 % de los casos, así la mediana dejará la mitad de las observaciones por debajo de su valor y la otra mitad por encima.

Ejemplo:

hallar la mediana del siguiente conjunto de datos:

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

Solución:

Organizando los datos de menor a mayor se tiene
1 1 1 1 1 1 1 2 2 2 2 2 2 3 3 3 3 4 4 4 5 5 5 5 5
el elemento intermedio es 2, así que la mediana es de este conjunto de datos es 2

Moda: En un conjunto de datos, la moda es el dato de mayor frecuencia.

Ejemplo:

Hallar la moda en el siguiente conjunto de datos: 5, 3, 6, 5, 4, 5, 2, 4, 6, 4, 5

Solución:

Se organizan los datos en una tabla resumen de la siguiente manera:

N°	2	3	4		
frecuencia	1	1	3		

así el dato de mayor frecuencia en este conjunto de datos es el número 5, el cual se registra cuatro veces en la lista de números.



Actividad 7

1. En el siguiente conjunto de datos,
1 1 1 1 1 1 1 2 2 2 2 2 2 3 3 3 3 4 4 4 5 5 5 5
la media es:

- a. 2,5
- b. 2,72
- c. 3
- d. 4

2. En el siguiente conjunto de datos,
18, 18, 19, 17, 23, 20, 21, 18 la mediana es:

- a. 17
- b. 18
- c. 18.5
- d. 23

3. En el siguiente conjunto de datos,
20, 21, 18, 19, 18, 17, 18 la moda es:

- a. 19
- b. 18
- c. 18.71
- d. 18.5

Ejercicios 4 – 5

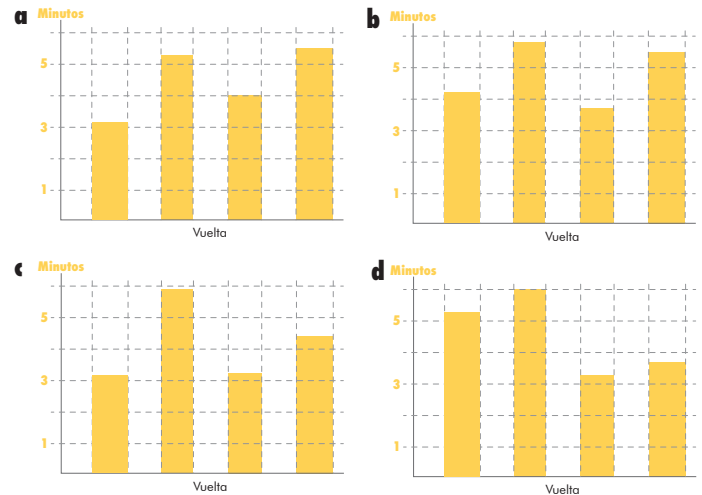
Felipe dio cuatro vueltas alrededor de una pista de automovilismo en su nuevo carro eléctrico. El tiempo que él empleó en dar cada vuelta se muestra a continuación:

Vuelta	Tiempo empleado
Vuelta #1	4 minutos y 15 segundos
Vuelta #2	5 minutos y 45 segundos
Vuelta #3	3 minutos y 30 segundos
Vuelta #4	4 minutos y 45 segundos

4. El tiempo total empleado por Felipe en dar las cuatro vueltas alrededor de la pista fue de:

- a. 16 minutos
- b. 16 minutos y 30 segundos
- c. 17 minutos
- d. 18 minutos y 15 segundos

5.Cuál de las siguientes graficas representa mejor el tiempo que Felipe tardó en cada vuelta:



Ejercicios 6 – 8

Una distribución estadística de datos viene dada por la siguiente tabla:

x_i	61	64	67	70	73
f_i	5	18	42	27	8

6. La moda de la distribución es:

- a. 67
- b. 64
- c. 42
- d. 73

7. La media de la distribución es:

- a. 68
- b. 67.5
- c. 69
- d. 68.5

8. La mediana de la distribución es:

- a. 67
- b. 68
- c. 69
- d. 65

Ejercicios 9 – 15

En una clínica dental observan el número de caries en 100 niños analizados. La información obtenida aparece resumida en la siguiente tabla



Nº de caries	f_i	f_i
0	25	0,25
1	20	0,2
2	x	z
3	15	0,15
4	15	0,15

9. El valor de x es:

- a. 35
- b. 30
- c. 40
- d. 25

10. El valor de z es:

- a. 0,25
- b. 0,30
- c. 0,35
- d. 0,4

11. El valor de y es:

- a. 5
- b. 10
- c. 0
- d. 15

12. El valor promedio de la información es:

- a. 2
- b. 1
- c. 1,55
- d. 0,5

13. La mediana de la información es:

- a. 15
- b. 35
- c. 40
- d. 30

14. La mediana de la información es:

- a. 15
- b. 35
- c. 40
- d. 30

15. La moda de la información es:

- a. 15
- b. 20
- c. 40
- d. 35

