ESTADISTICA EN LAS ORGANIZACIONES

ESTADISTICA

ROSARIO GOMEZ LUJANO

PRESENTA EL ALUMNO:

Rosa Linney Melo Castellanos

GRUPO, SEMESTRE y MODALIDAD:

1er. Semestre Semiescolarizado

Pichucalco, Chiapas.

25 de septiembre de 2020.

Estadística

Definición:

La estadística consiste en métodos, procedimientos y fórmulas que permiten recolectar información para luego analizarla y extraer de ella conclusiones relevantes. Se puede decir que es la Ciencia de los Datos y que su principal objetivo es mejorar la comprensión de los hechos a partir de la información disponible.

El origen de la palabra estadística se suele atribuir al economista Gottfried Achenwall (prusiano, 1719-1772) que entendía la estadística como "ciencia de las cosas que pertenecen al Estado".

Clasificación de la estadística:

La estadística para su mejor estudio se ha dividido en dos ramas las cuales son: estadística descriptiva y estadística inferencial.

Estadística descriptiva: Consiste en la presentación de datos en forma de tablas y gráficas. Esta comprende cualquier actividad para resumir o describir los mismos factores pertinentes adicionales, esto se refiere a no intentar nada que vaya más allá de los datos.

Estadística inferencial: Se deriva de las observaciones hechas solo a una parte de un conjunto numeroso de elementos; implicando así que su análisis requiera de generalizaciones que van más allá de los datos, como consecuencia la característica más importante del crecimiento de la estadística ha sido un cambio en el énfasis de los métodos que sirven para generalizarlas. En otras palabras, la estadística inferencial investiga y analiza una población partiendo de una muestra tomada.

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.

Algunas definiciones:

- "Una población es un conjunto de todos los elementos que estamos estudiando, acerca de los cuales intentamos sacar conclusiones". Levin & Rubin (1996).
- "Una población es un conjunto de elementos que presentan una característica común". Cadenas (1974).

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

Muestra:

La muestra es una representación significativa de las características de una población, que bajo, la asunción de un error (generalmente no superior al 5%) estudiamos las características de un conjunto poblacional mucho menor que la población global.

- "Se llama muestra a una parte de la población a estudiar que sirve para representarla". Murria R. Spiegel (1991).
- "Una muestra es una colección de algunos elementos de la población, pero no de todos". Levin & Rubín (1996).
- "Una muestra debe ser definida en base de la población determinada, y las conclusiones que se obtengan de dicha muestra solo podrán referirse a la población en referencia", Cadenas (1974).

Muestreo:

En la referencia estadística se conoce como muestreo a la técnica para la selección de una muestra a partir de una población estadística.

Al elegir una muestra aleatoria se espera conseguir que sus propiedades sean extrapolables a la población. Este proceso permite ahorrar recursos, y a la vez obtener resultados parecidos a los que se alcanzarían si se realizase un estudio de toda la población. En las investigaciones llevadas por empresarios y de la medicina se usa muestreo extensivamente en recoger información sobre poblaciones.

Cabe mencionar que para que el muestreo sea válido y se pueda realizar un estudio adecuado (que consienta no solo hacer estimaciones de la población sino estimar también los márgenes de error correspondientes a dichas estimaciones), debe cumplir ciertos requisitos. Nunca podremos estar enteramente seguros de que el resultado sea una muestra representativa, pero sí podemos actuar de manera que esta condición se alcance con una probabilidad alta.

Censo:

En estadística descriptiva, se denomina censo al recuento de individuos que conforman una población estadística, definida como un conjunto de elementos de referencia sobre el que se realizan las observaciones. El censo de una población estadística consiste básicamente en obtener mediciones del número total de individuos mediante diversas técnicas de recuento y se realiza cada determinado período.

Variable:

Una variable es una propiedad que puede fluctuar y cuya variación es susceptible de adoptar diferentes valores, los cuales pueden medirse u observarse. Las variables adquieren valor cuando se relacionan con otras variables, es decir, si forman parte de una hipótesis o de una teoría. En este caso se las denomina construcciones hipotéticas.

Medidas de tendencia central:

Las medidas de tendencia central son parámetros estadísticos que informan sobre el centro de la distribución de la muestra o población estadística.

A veces, tratamos con una gran cantidad información. Variables que presentan muchos datos y muy dispares. Datos con muchos decimales, de diferente signo o longitud. En estos casos, siempre es preferible calcular medidas que nos ofrezcan información resumida sobre dicha variable. Por ejemplo, medidas que nos indiquen cuál es el valor que más se repite.

Media: La media es el valor promedio de un conjunto de datos numéricos, calculada como la suma del conjunto de valores dividida entre el número total de valores. A continuación, se muestra la fórmula de la media aritmética:

$$Media\ aritmética = \frac{\sum_1^N x_i}{N} = \frac{x_1 + \ x_2 + \ x_3 \ .... + x_n}{N}$$

Como se explica en el artículo enlazado anteriormente, existen muchos tipos de media. La elección de cada tipo de media tiene que ver, principalmente con el tipo de dato sobre el que se calcula.

Mediana: La mediana es un estadístico de posición central que parte la distribución en dos, es decir, deja la misma cantidad de valores a un lado que a otro. Las fórmulas propuestas no nos darán el valor de la mediana, lo que nos darán será la posición en la que está dentro del conjunto de datos. Las fórmulas que indica la posición de la mediana en la serie son las siguientes:

- Cuando el número de observaciones es par:

Mediana = (n+1) / 2 → Media de las posiciones observaciones

- Cuando el número de observaciones es impar:

Mediana = (n+1) / 2 → Valor de la observación

Moda: La moda es el valor que más se repite en una muestra estadística o población. No tiene fórmula en sí mismo. Lo que habría que realizar es la suma de las repeticiones de cada valor. Por ejemplo, ¿cuál es la moda de la siguiente tabla de salarios?

Trabajador	Salario	
1	€	1.236
2	€	1.236
3	€	859
4	€	486
5	€	1.536
6	€	1.536
7	€	1.621
8	€	978
9	€	1.236
10	€	768

La moda sería 1.236€. Si vemos los salarios de los 10 trabajadores, veríamos que 1.236€ se repite en tres ocasiones.

Medidas de dispersión:

Parámetros estadísticos que indican como se alejan los datos respecto de la media aritmética. Sirven como indicador de la variabilidad de los datos. Las medidas de dispersión más utilizadas son el rango, la desviación estándar y la varianza.

Recorrido: La medida de dispersión más inmediata es el recorrido de la distribución estadística, también llamado rango o amplitud. Dada una serie de valores x1, x2, ..., xn, su recorrido es la diferencia aritmética entre el máximo y el mínimo de estos valores:

$$Re = x_i\ (\text{máx}) - x_i\ (\text{mín}), \text{ siendo } i = 1, 2, ..., n.$$

Desviación media: Como medida de dispersión más frecuentemente utilizada, la desviación media se define como la media aritmética de los valores absolutos de la desviación de cada valor de la variable con respecto a la media. Su formulación matemática es la siguiente:

$$DM = \frac{f_1|x_1 - \overline{x}| + f_2|x_2 - \overline{x}| + \ldots + f_n|x_n - \overline{x}|}{f_1 + f_2 + \ldots + f_n} = \\ = \frac{\sum f_i |x_i - \overline{x}|}{\sum f_i} \text{ con } i = 1, 2, 3, ..., n.$$

Varianza y desviación típica: La desviación media no siempre suministra una idea clara del grado de separación entre los valores de una variable estadística. Para estudios científicos, se prefiere utilizar una pareja de parámetros relacionados que se conocen como varianza y desviación típica.

La varianza se define como el cociente entre la suma de los cuadrados de las desviaciones de los valores de la variable y el número de datos del estudio. Matemáticamente, se expresa como:

$$V = \frac{f_1(x_1 - \overline{x})^2 + f_2(x_2 - \overline{x})^2 + \ldots + f_n(x_n - \overline{x})^2}{f_1 + f_2 + \ldots + f_n} = \\ = \frac{\sum f_i(x_i - \overline{x})^2}{\sum f_i} \text{ con } i = 1, 2, 3, ..., n.$$

Por su parte, la desviación típica, simbolizada por s, se define sencillamente como la raíz cuadrada de la varianza:

$$\sigma = \sqrt{V}$$

Por lo tanto, se tiene que:

$$\sigma = \sqrt{\frac{\sum f_i(x_i - \overline{x})^2}{\sum f_i}}$$

La varianza y la desviación típica, cada una con su respectiva valor, se usan indistintamente en los estudios estadísticos.

Distribución o tabla de frecuencias para datos no agrupados y agrupados.

Datos no agrupados

Datos diferentes: Consideraremos como un dato diferente, a cada uno de los distintos datos que se presentan en la muestra, los denotaremos por xi, y al número total de datos diferentes lo denotaremos por m.

Cuando el tamaño de la muestra (n) es finito y el número de datos diferentes es pequeño (consideraremos pequeño k ≤ 10), es fácil hacer un análisis de los datos tomando cada uno de los datos diferentes y ordenándolos cualitativa o cuantitativamente.

EJEMPLO 1: Cierta universidad realizó un experimento sobre el coeficiente intelectual (C.I.) de sus alumnos, para lo cual aplicó un examen de C.I. a un grupo de 20 alumnos escogidos al azar, obteniendo los siguientes resultados:

119, 109, 124, 119, 106, 112, 112, 112, 112, 109, 112, 124, 109, 109, 109, 106, 124, 112, 112,106.

Datos	Repeticiones
106	3
109	5
112	7
119	2
124	3

Toda vez que se tienen los datos, se ordenan de menor a mayor o viceversa.

106, 106, 106, 109, 109, 109, 109, 109, 112, 112, 112,

112, 112, 112, 112, 119, 119, 124, 124, 124

Resultado:

Coeficiente Intelectual				
x_i	f	f_a	f_r	f_{ra}
106	3	3	0.15	0.15
109	5	8	0.25	0.40
112	7	15	0.35	0.75
119	2	17	0.10	0.85
124	3	20	0.15	1.00
Total	20		1.00	

(Frecuencia de datos no agrupados)

EJEMPLO 2. Se preguntó a un grupo de alumnos de primer año del Plantel “X”, por la asignatura de su preferencia, arrojándose los siguientes resultados:

Asignaturas

Mate	Social	Taller	Quím.	Infor	Mate	Inglés
Mate	Quím.	Infor	Inglés	Ética	Inglés	Social
Inglés	Ética	Mate	Taller	Quím.	Mate	Taller
Social	Mate	Inglés	Infor	Inglés	Ética	Infor
Mate	Inglés	Infor	Ética	Quím.	Taller	Inglés
Social	Inglés	Ética	Taller	Infor	Quím.	Taller
Taller	Infor	Mate	Quím.	Infor	Mate	Infor
Inglés						

Asignatura	Repeticiones (frecuencias)
Ética y valores	5
Informática	9
Inglés	10
Matemáticas	9
Química	6
Sociales	4
Taller de lectura	7
Total	50

Resultado:

Asignatura de Preferencia				
x_i	f	f_a	f_r	f_{ra}
Ética y valores	5	5	0.1	0.1
Informática	9	14	0.18	0.28
Inglés	10	24	0.2	0.48
Matemáticas	9	33	0.18	0.66
Química	6	39	.12	0.78
Sociales	4	43	0.08	0.86
Taller de lectura	7	50	0.14	1.00
Total	50		1.00	

(*Frecuencia de datos no agrupados*)

Datos agrupados

Cuando el tamaño de la muestra es considerable o grande y los datos numéricos son muy diversos (n>15), conviene agrupar los datos de tal manera que permita establecer patrones, tendencias o regularidades de los valores observados. De esta manera podemos condensar y ordenar los datos tabulando las frecuencias asociadas a ciertos intervalos de los valores observados.

Intervalos de Clase: Son los intervalos en los que se agrupan y ordenan los valores observados. Cada uno de estos intervalos está delimitado (acotado) por dos valores extremos que les llamamos límites.

Pasos a seguir para construir intervalos de frecuencia.

1. Determinar la cantidad de intervalos apropiada

La selección del número adecuado de intervalos y los límites entre ellos dependen del criterio o experiencia de quien realiza el estudio. Sin embargo, existen reglas empíricas para ***1.- Calcular el número de intervalos:*** la más empleada es la Regla de Sturges, cuya expresión es: K= 1 + 3.3 $Log\ n$

Dónde: K=Número de intervalos el cual siempre debe ser un número entero. Razón por la cual se deberá redondear el resultado al entero más cercano.

n= Número de datos.

Log = logaritmo en base 10.

Otra regla utilizada es la de Velleman que establece que el número de Intervalos se obtiene de la raíz cuadrada del número de datos; es decir K= $\sqrt{n}$, recomendable para tamaños de muestra pequeños (n< 50)

El número de intervalos determinado mediante cualquier regla se aproxima al valor entero más cercano, pero deberá ser responsabilidad de quien realiza el estudio, pudiendo utilizar éste en ocasiones uno menor o mayor al obtenido por cualquier regla, si esto le permite tener intervalos con la misma amplitud. Sin embargo, la mayoría de las reglas subestiman el número de intervalos.

2.- Calcular el rango de los datos.

Llamamos rango al número de unidades de variación presente en los datos recopilados y se obtiene de la diferencia entre el dato mayor y el dato menor. Se representa con la letra R. R= Dato mayor menos dato menor

3.- Obtención de la amplitud o anchura que tendrá cada intervalo.

Se encuentra dividiendo el rango por el número de intervalos regularmente es de 5 a 6. Se representa con la letra *A* de tal manera que $\boldsymbol{Ac = \frac{R}{K}}$.

4.- Construcción de los intervalos.

Los intervalos de clase son conjuntos numéricos y deben ser excluyentes y exhaustivos; es decir, si un dato pertenece a un intervalo determinado, ya no podrá pertenecer a otro, esto quiere decir excluyentes y además todos y cada uno de los datos deberá estar contenido en alguno de los intervalos, esto les da el valor de exhaustivos.

Las dos características mencionadas anteriormente se logran construyendo intervalos cerrados por la izquierda y abiertos por la derecha; esto se simboliza a través del uso de corchetes y paréntesis respectivamente. Por razones naturales, el último intervalo será cerrado por ambos extremos.

El primer intervalo se construye de la siguiente manera: Habrá de iniciar con el dato menor, el cual será el extremo inferior del intervalo; el otro extremo se obtiene de la suma del dato menor y la amplitud, con este mismo valor iniciamos el segundo intervalo, del cual el segundo extremo se encuentra sumando al valor anterior la amplitud y este proceso se repite sistemáticamente hasta completar el total de intervalos indicado por la regla elegida, por ejemplo, la de Sturges.

Los valores extremos o límites de intervalo.

Los intervalos de clase deben estar definidos por límites que permitan identificar plenamente si un dato pertenece a uno u otro intervalo. Estos límites son los valores extremos de cada intervalo.

Límite inferior: Es el valor menor de cada intervalo, se denota por Li

Límite superior: Es el número mayor de cada intervalo, se denota por Ls.

También será muy útil conocer y calcular la Marca de Clase (MC) de cada intervalo: Se refiere al Punto Medio del intervalo y a través de él representaremos a todo el intervalo y una de las maneras de calcularla es promediando los valores límite de cada intervalo, su fórmula es:

$$MC = \frac{Li + Ls}{2}$$

Ejemplo 1.-

Un grupo de investigadores pertenecientes a la secretaría de seguridad pública, tomó una muestra aleatoria de las velocidades (km/h) registradas por 30 vehículos en el trayecto Hermosillo a Ures, con el fin de establecer nuevos límites máximos de velocidad para una carretera. La muestra arrojo los datos siguientes:

90, 99, 104, 99, 119, 98, 95, 112, 95, 120, 100, 90, 116, 96, 114, 108, 98, 118, 100, 106, 114, 100, 112, 106, 100, 115, 111, 105, 114, 97

Toda vez que se tienen los datos, se recomienda ordenarlos de menor a mayor o viceversa

90, 90, 95, 95, 96, 97, 98, 98, 99, 99, 100, 100, 100, 104, 105,106, 108, 111,

112, 112, 114, 114, 115, 116, 118, 119, 120

Ahora llevamos a la práctica los pasos descritos anteriormente para la construcción de los intervalos.

1º obtendremos el número de intervalos que vamos a utilizar, para lo cual empleamos la Regla de Sturges:

K = 1 + 3.3Log (30) = 1+ 3.3 (1.4771212547) =1+ 4.87 = 5.87 ≈ 6

2º calculamos el rango de variación, R = 120 – 90 = 30

3º obtenemos la amplitud de cada intervalo de clase como sigue:

$$Ac = \frac{30}{6} = 5$$

4º construimos los intervalos: el primero de ellos inicia con 90 que es el extremo inferior que, sumado a 5 obtenemos 95, que será el extremo superior; este extremo será el inferior del segundo intervalo; y al sumar nuevamente la amplitud tendremos 100 que será el extremo superior y así sucesivamente hasta completar los 6 intervalos., que se muestran enseguida:

[90 – 95), [95 – 100), [100 – 105), [105 – 110), [110 – 115) y [115 – 120]

Los corchetes expresan que el valor extremo se incluye en el intervalo y los paréntesis dan a entender que el valor extremo del intervalo no se incluye en el.

Para la construcción de distribuciones de frecuencias, contamos el número de datos que le corresponden a cada intervalo; es decir obtenemos las frecuencias absolutas y de estas podemos generar los demás tipos de frecuencias y presentarlas en una tabla de resumen como la que a continuación se muestra:

Distribuciones de frecuencias para las velocidades

x_i Intervalos de Clase	f	f_a	f_r	f_{ra}	m_c
[90 – 95)	2	2	0.07	0.07	92.5
[95 – 100)	8	10	0.27	0.34	97.5
[100 – 105)	5	15	0.17	0.51	102.5
[105 – 110)	4	19	0.13	0.64	107.5
[110 – 115)	6	25	0.20	0.84	112.5
[115 – 120]	5	30	0.16	1.00	117.5
Total	30		1.00		

Ejemplo 2.-

Los datos que se dan a continuación corresponden a los pesos en Kg. de ochenta personas:

a) Obténgase una distribución de datos en intervalos de amplitud 5, siendo el primer intervalo [50; 55].
b) Calcúlese el porcentaje de personas de peso menor que 65 Kg.
c) ¿Cuántas personas tienen peso mayor o igual que 70 Kg? pero menor que 85?}

60 ; 66 ; 7 7; 7 0 ; 6 6 ; 6 8 ; 57 ; 7 0; 6 6 ; 5 2 ; 7 5 ; 65 ; 6 9; 7 1 ; 5 8 ; 6 6 ; 67 ; 7 4; 6 1;
63 ; 69 ; 8 0; 5 9 ; 6 6 ; 7 0 ; 67 ; 7 8; 7 5 ; 6 4 ; 7 1 ; 81 ; 6 2; 6 4 ; 6 9 ; 6 8 ; 72 ; 8 3; 5 6;
65 ; 74 ; 6 7; 5 4 ; 6 5 ; 6 5 ; 69 ; 6 1; 6 7 ; 7 3 ; 5 7 ; 62 ; 6 7; 6 8 ; 6 3 ; 6 7 ; 71 ; 6 8; 7 6;
61 ; 62 ; 6 3; 7 6 ; 6 1 ; 6 7 ; 67 ; 6 4; 7 2 ; 6 4 ; 7 3 ; 79 ; 5 8; 6 7 ; 7 1 ; 6 8 ; 59 ; 6 9; 7 0; 6 6 ; 62 ; 6 3; 66

Solución: Como se trata de efectuar una distribución de datos agrupados, debemos obtener primero los intervalos correspondientes, situando los datos en sus lugares respectivos:

L_{i-1} - L_i	n_i	N_i
[50;55)	2	2
[55; 60)	7	9
[60; 65)	17	26
[65;70)	30	56
[70; 75)	14	70
[75; 80)	7	77
[80; 85]	3	80
	80	

Observando la columna de frecuencias acumuladas se deduce que existen N3 = 26 individuos cuyo peso es menor que 65 Kg., que en términos de porcentaje corresponden a:

$$\frac{26}{80}\cdot 100 = 32{,}5$$

El número de individuos con peso comprendido entre 70 y 85 Kg. es:

$n_5 + \mathrm{n}_6 + n_7 = 14 + 7 + 3 = 24$

lo que es equivalente a: $N_7 - N_4 = 80 - 56 = 24$

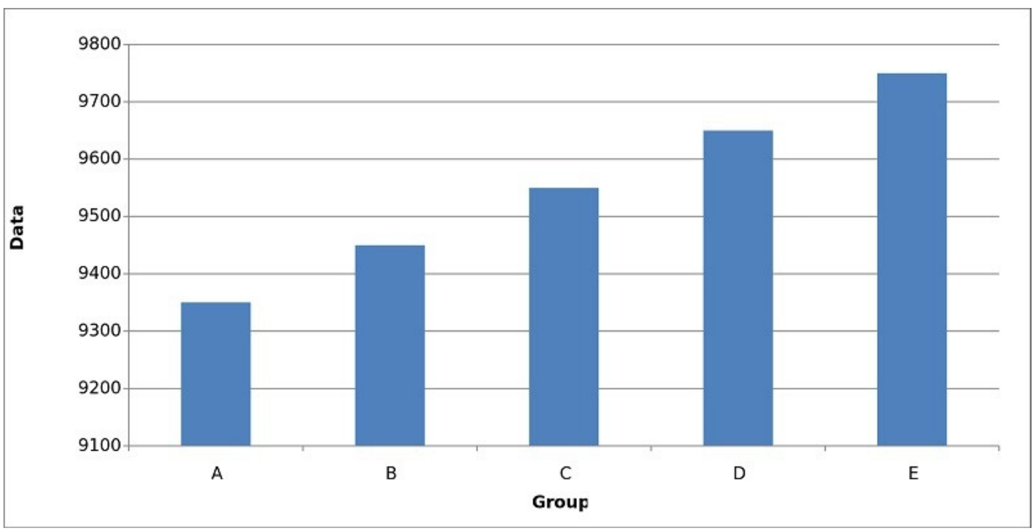
Tipos de gráficas

Existen muy diversos tipos de gráficas, generalmente aplicándose unas u otras en función de lo que se pretenda representar o simplemente de las preferencias del autor. A continuación, indicamos algunas de las más conocidas y comunes.

1. Gráfico de barras

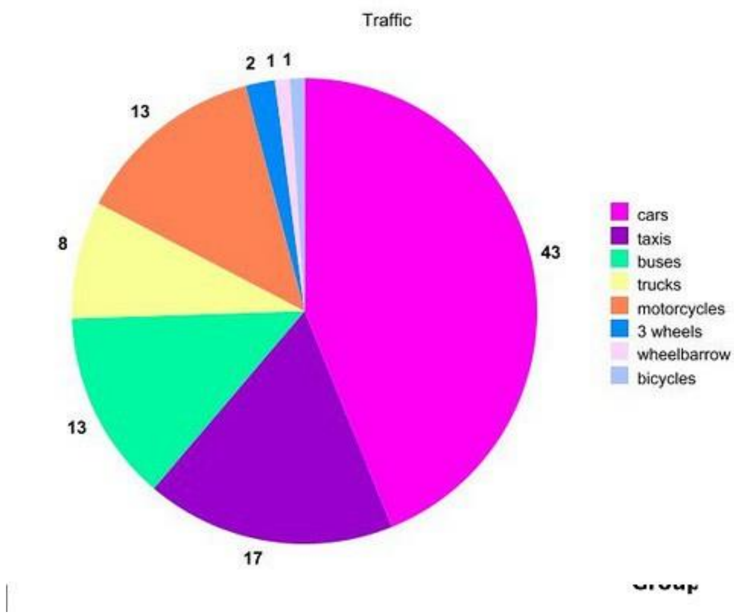
El más conocido y utilizado de todos los tipos de gráficos es el gráfico o diagrama de barras. En éste, se presentan los datos en forma de barras contenidas en dos ejes cartesianos (coordenada y abscisa) que indican los diferentes valores. El aspecto visual que nos indica los datos es la longitud de dichas barras, no siendo importante su grosor.

Generalmente se emplea para representar la frecuencia de diferentes condiciones o variables discretas (por ejemplo, la frecuencia de los diferentes colores del iris en una muestra determinada, que solo pueden ser unos valores concretos). Únicamente se observa una variable en las abscisas, y las frecuencias en las coordenadas.



2. Gráfico circular

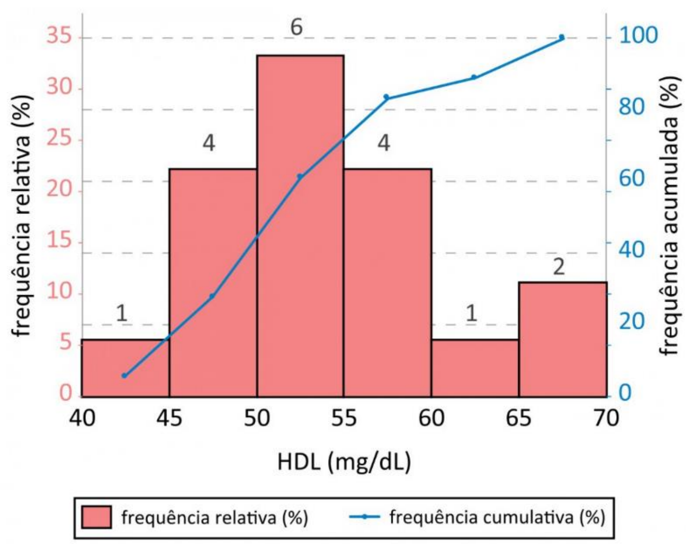
El también muy habitual gráfico en forma de "quesito", en este caso la representación de los datos se lleva a cabo mediante la división de un círculo en tantas partes como valores de la variable investigada y teniendo cada parte un tamaño proporcional a su frecuencia dentro del total de los datos. Cada sector va a representar un valor de la variable con la que se trabaja. Este tipo de gráfico o diagrama es habitual cuando se está mostrando la proporción de casos dentro del total, utilizando para representarlo valores percentuales (el porcentaje de cada valor).



3. Histograma

Aunque a simple vista muy semejante al gráfico de barras, el histograma es uno de los tipos de gráfica que a nivel estadístico resulta más importante y fiable. En esta ocasión, también se utilizan barras para indicar a través de ejes cartesianos la frecuencia de determinados valores, pero en vez de limitarse a establecer la frecuencia de un valor concreto de la variable evaluada refleja todo un intérvalo. Se observa pues un rango de valores, que además podrían llegar a reflejar intervalos de diferentes longitudes.

Ello permite observar no solo la frecuencia sino también la dispersión de un continuo de valores, lo que a su vez puede ayudar a inferir la probabilidad. Generalmente se utiliza ante variables continuas, como el tiempo.



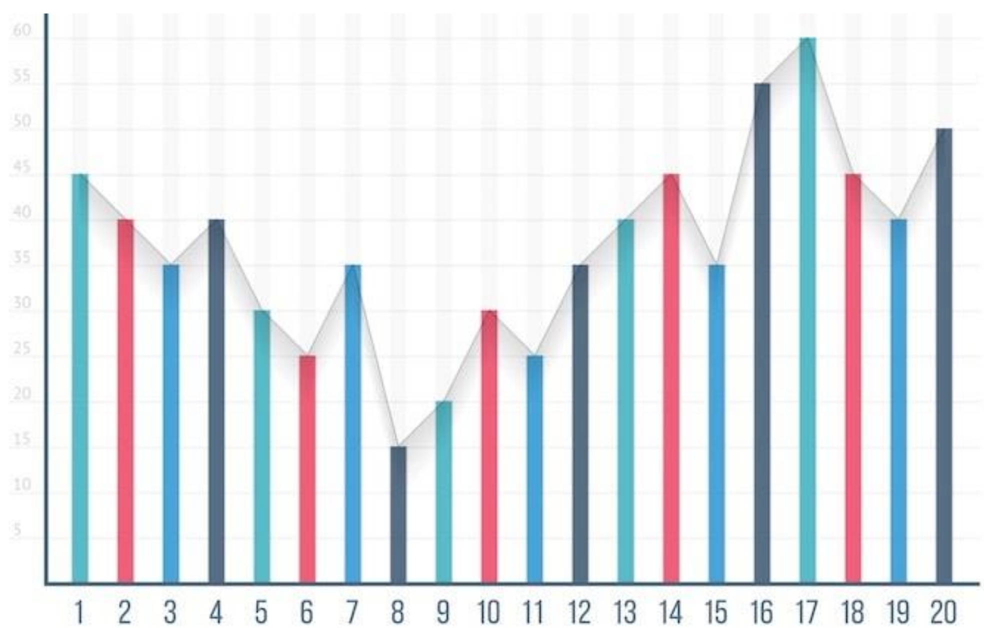
4. Polígono de frecuencias

Un polígono de frecuencia es una herramienta gráfica que se emplea a partir de un histograma de frecuencia (es decir, otro tipo de gráfico que expresa las frecuencias mediante columnas verticales). Para ello, se unen con una línea los distintos puntos medios de las columnas del histograma, sin dejar espacio entre una y otra, logrando así una forma geométrica o polígono.

Con esta herramienta gráfica pueden representarse variables cuantitativas o distribuciones diferentes, cosa que tradicionalmente no hace un histograma, de un modo rápido y sencillo. Además, cuenta con la virtud de ser apreciable a simple vista.

Por esta razón es sumamente empleado dentro de las ciencias sociales y ciencias económicas, permitiendo así establecer comparaciones útiles entre los distintos resultados de un mismo proceso.

Se emplean los polígonos de frecuencias cuando es necesario graficar o resaltar distintas distribuciones conjuntas o bien una clasificación cruzada de una variable cuantitativa continua, junto con otra variable cualitativa o cuantitativa discreta, todo dentro de un mismo gráfico.



Resuelve el siguiente ejercicio:

Calcula la media aritmética o promedio, mediana, moda, rango, desviación estándar, varianza para las siguientes calificaciones que obtuvo un alumno LPSI de la UDS: 6, 7, 8, 9, 9,8,7, 6,10

Promedio: 7.7

6 + 6 + 7 + 7 + 8 + 8 + 9 + 9 + 10 = 7.7

Mediana: 8

Moda: 6, 7, 8 y 9

Rango: 6 – 10 = 4

Desviación estándar: 1.39

Varianza: 1.93