

PROYECTO DE ESTADISTICA

ESTADISTICA I
ROSARIO GOMEZ LUJANO

PRESENTA EL ALUMNO:

Mónica Alejandra Ruiz Solís

GRUPO, SEMESTRE y MODALIDAD:

Ier. Cuatrimestre Psicología Semi-escolarizado

Pichucalco, Chiapas

25 de septiembre de 2020

Estadística

La estadística es una ciencia, que nos sirve como herramienta, facilita la obtención de datos precisos de lo que estamos estudiando en el momento.

A través de su proceso hace más sencillo organizar y analizar los datos para dar un resultado, la estadística no son solo números o porcentajes, sino todo un proceso de investigación, observaciones, análisis y evaluaciones detalladas que son la base como tal de los resultados finales y exactos de la problemática o situación que se está estudiando.

La estadística nos sirve para estudiar todo tipo de cosa o situación, desde un grupo de personas, animales o plantas, hasta un grupo de objetos los cuales deseemos obtener cierta respuesta a lo que deseemos estudiar de ellos.

Clasificación de la estadística

La estadística se clasifica en dos: estadística descriptiva y estadística inferencial.

Estadística descriptiva. Como su nombre lo dice, describe las características de los datos que se tienen, se encarga también de recolectar, agrupar y presentar los resultados. Este proceso nos ayuda hacer más visible los datos, ya que los resultados se representan a través de graficas o tablas para tener una mejor organización y que se vean más estéticos los datos que se obtengan de toda la información que se recabo con anticipación.

Estadística inferencial. En ella se lleva a cabo una selección precisa de elementos claves, los cuales nos servirán para investigarlos y analizarlos, obteniendo de esa pequeña parte tomada el resultado deseado, todo esto partiendo de la muestra tomada, o sea cada uno de los elementos seleccionados.

Población

La población es el número total de personas o cosas a estudiar y analizar, la cual puede ser tomar toda completa o solo seleccionar una parte de ella, dependiendo lo grande o pequeña que esta sea. Y como se mencionó anteriormente a la pequeña

toma de toda la población se le conoce como muestra, lo que veremos a continuación de que se trata.

Muestra

La muestra, como ya se dijo, es una pequeña porción de toda la población que se desea o se está estudiando. La muestra facilita estudiar y analizar la población, porque es una selección específica de los elementos que puedan representar correctamente a toda la población, es decir que la muestra tenga características similares al de toda la población a la que pertenece, de la que fue tomada.

La muestra se puede obtener haciendo un censo de la población, con ese censo aplicado a la población se obtendrá ciertos datos, que son los que se tomaran como la muestra, lo cual representara a la población en sí.

Muestreo

El muestreo es una herramienta que se encarga de determinar cuál será la parte que se tomará de la población para poder estudiarla y sacar los resultados de ella.

Debe asegurarse de que la parte seleccionada sea realmente representativa de la población, debe dar a conocer las características de las partes tomada y compararlas con toda la población, para así tener la certeza de que son similares y no haya errores al momento de dar a conocer los datos obtenidos.

Las características pueden ser, sexo, edad, religión, etc., pero deben coincidir con la de la población completa para dar un resultado valido.

Censo

El censo son datos estadísticos de toda la población, para dar a conocer así las características específicas, que puedan dar respuesta de toda la población seleccionada.

El censo hace una recopilación de datos que servirán para saber que parte tomar para ser representación de toda la población que se tiene estudiando. Se toma

información cuantitativa y cualitativa para tener mejores resultados, y ya sea dependiendo lo que se esté estudiando se recolecta información como, edad, ingresos, egresos, sexo, religión entre otras cosas más, que puedan ser características y tengan similitud a la población y ayuden a que el resultado que se obtenga sea correcto.

Con todo el proceso que se realiza, si se hace de manera correcta será sencillo obtener cuales datos serán parte de la muestra y así también tener el resultado correcto que se desea saber.

Variables

Las variables son rasgos o características que pueden ser medidas, llegando a tener distintos valores según lo que se esté estudiando.

Existen tipos de variables: las cualitativas, que como su nombre lo dice, son las que describen las características; las cuantitativas, es obvio saber de qué se tratan, son las que se encargan de dar a conocer las cantidades, los números; las variables dependientes y las independientes, la primera es lo que es analizado y medido para ser posible la variable independiente y la segunda es toda aquella información recabada que es la causa o base de lo que se está estudiando y esta a su vez puede ser manipulada por el investigador.

medidas de tendencia central

las medidas de tendencia central son medidas estadísticas que resumen todos los valores obtenidos a solo uno. En las medidas de tendencia central se encuentran todos los datos recabados.

Las medidas de tendencia son: la media, la mediana y la moda, las cuales se encargan de promediar y de dar el resultado que buscamos.

Facilita en cierta forma obtener el resultado de nuestros estudios así se tengan como datos grandes cantidades, estas medidas nos ayudan a dar con el resultado de una manera más rápida y sencilla.

Medidas de dispersión

Las medidas de dispersión se encargan de medir que tanto de diferencia hay entre un valor a otro, que tan lejos o cerca están cada uno de esos datos.

Los dos tipos de medidas proporcionan información de las posiciones, la distancia y diferencia de los datos estudiados.

EJEMPLO DE DISTRIBUCION DE FRECUENCIA PARA DATOS NO AGRUPADOS.

Datos:

515, 622, 688, 701, 701, 749, 765, 800, 805, 872, 955, 961,993, 997, 1023, 1043, 1073, 1222, 1241, 1442

$$N^{\circ} \text{ clase} = \sqrt{\text{No datos}} =$$

$$= \sqrt{20} = 4.74 \cong 5$$

$$\text{Amplitud} = \text{valor} \geq - \text{valor} \leq / N^{\circ} \text{ clase} =$$

$$= 1444 - 515 / 5 = 185.4 \cong \mathbf{186}$$

$$LI = Li + \text{Amplitud} =$$

$$B = 515 + 186 = 701$$

$$C = 701 + 186 = 887$$

$$D = 887 + 186 = 1073$$

$$E = 1073 + 186 = 1259$$

$$LS = LI + \text{Amplitud} - U.V$$

$$A = 701 - 1 = 700$$

$$B = 887 - 1 = 886$$

$$C = 1073 - 1 = 1072$$

$$D = 1259 - 1 = 1258$$

$$E = 1445 - 1 = 1444$$

$$LI = Li - (LI - LS / 2) =$$

$$A = 515 - (701 - 700 / 2) = 515 - 0.5 = 514.5$$

$$B = 701 - (887 - 886 / 2) = 701 - 0.5 = 700.5$$

$$C = 887 - (1073 - 1072 / 2) = 887 - 0.5 = 886.5$$

$$D = 1073 - (1259 - 1258 / 2) = 1073 - 0.5 = 1072.5$$

$$E = 1259 - (1445 - 1444/2) = 1259 - 0.5 = 1258.5$$

$$LS = LS + (LI - LS/2) =$$

$$A = 700 + (701 - 700/2) = 700 + 0.5 = 700.5$$

$$B = 886 + (887 - 886/2) = 886 + 0.5 = 886.5$$

$$C = 1072 + (1073 - 1072/2) = 1072 + 0.5 = 1072.5$$

$$D = 1258 + (1259 - 1258/2) = 1258 + 0.5 = 1258.5$$

$$E = 1444 + (1445 - 1444/2) = 1444 + 0.5 = 1444.5$$

$$Y_i = LI + LS/2 =$$

$$A = 515 + 700/2 = 607.5$$

$$B = 701 + 886/2 = 793.5$$

$$C = 887 + 1072/2 = 979.5$$

$$D = 1073 + 1258/2 = 1165.5$$

$$E = 1259 + 1444/2 = 1351.5$$

CLASE	LIMITE INFERIOR	LIMITE SUPERIOR	LIMITE INFERIOR EXACTO	LIMITE SUPERIOR EXACTO	MARCA DE CLASE	FRECUENCIA ABSOLUTA	FRECUENCIA RELATIVA
A	515	700	514.5	700.5	607.5	3	0.15
B	701	886	700.5	886.5	793.5	7	0.35
C	887	1072	886.5	1072.5	979.5	6	0.3
D	1073	1258	1072.5	1258.5	1165.5	4	0.2
E	1259	1444	1258.5	1444.5	1351.5	0	0
						20	

Frecuencia absoluta

$F = n^\circ$ datos entre un valor y otro

Frecuencia relativa

$$f_i = f_i/n$$

$$A = 3/20 = 0.15$$

$$D = 4/20 = 0.2$$

$$B = 7/20 = 0.35$$

$$E = 0$$

$$C = 6/20 = 0.3$$

Datos:

35, 39, 46, 51, 58, 63, 67, 69, 72, 76, 85, 94, 97, 99, 101, 103, 106, 109, 112, 115

$$N^{\circ} \text{ clase} = \sqrt{\text{No datos}} =$$

$$= \sqrt{20} = 4.74 \cong 5$$

$$\text{Amplitud} = \text{valor} \geq - \text{valor} \leq / N^{\circ} \text{ clase} =$$

$$= 115 - 35 / 5 = 80 / 5 \cong 16$$

$$Ll = Li + \text{Amplitud} =$$

$$B = 35 + 16 = 51$$

$$C = 51 + 16 = 67$$

$$D = 67 + 16 = 83$$

$$E = 83 + 16 = 99$$

$$Ls = Ll + \text{Amplitud} - U.V$$

$$A = 51 - 1 = 50$$

$$B = 67 - 1 = 66$$

$$C = 83 - 1 = 82$$

$$D = 99 - 1 = 98$$

$$E = 115 - 1 = 114$$

$$Ll = Ll - (Ll - Ls / 2) =$$

$$A = 35 - (51 - 50 / 2) = 35 - 0.5 = 34.5$$

$$B = 51 - (67 - 66 / 2) = 51 - 0.5 = 50.5$$

$$C = 67 - (83 - 82 / 2) = 67 - 0.5 = 66.5$$

$$D = 83 - (99 - 98 / 2) = 83 - 0.5 = 82.5$$

$$E = 99 - (115 - 114 / 2) = 99 - 0.5 = 98.5$$

$$Ls = Ls + (Ll - Ls / 2) =$$

$$A = 50 + (51 - 50/2) = 50 + 0.5 = 50.5$$

$$B = 66 + (67 - 66/2) = 66 + 0.5 = 66.5$$

$$C = 82 + (83 - 82/2) = 82 + 0.5 = 82.5$$

$$D = 98 + (99 - 98/2) = 98 + 0.5 = 98.5$$

$$E = 114 + (115 - 114/2) = 114 + 0.5 = 114.5$$

$$Y_i = LI + LS/2 =$$

$$A = 35 + 50/2 = 42.5$$

$$B = 51 + 66/2 = 58.5$$

$$C = 67 + 88/2 = 74.5$$

$$D = 83 + 98/2 = 90.5$$

$$E = 99 + 114/2 = 106.5$$

CLASE	LIMITE INFERIOR	LIMITE SUPERIOR	LIMITE INFERIOR EXACTO	LIMITE SUPERIOR EXACTO	MARCA DE CLASE	FRECUENCIA ABSOLUTA	FRECUENCIA RELATIVA
A	35	50	34.5	50.5	42.5	3	0.15
B	51	66	50.5	65.5	58.5	3	0.15
C	67	82	66.5	81.5	74.5	4	0.2
D	83	98	82.5	97.5	90.5	3	1.15
E	99	114	98.5	113.5	106.5	7	0.35
						20	

Frecuencia absoluta

$F = n^\circ$ datos entre un valor y otro

Frecuencia relativa

$$fr_i = f_i/n$$

$$A = 3/20 = 0.15$$

$$D = 3/20 = 0.15$$

$$B = 3/20 = 0.15$$

$$E = 7/20 = 0.35$$

$$C = 4/20 = 0.2$$

Ejemplo de distribución de frecuencia para datos agrupados

CLASE	LI	LS	fi	Fi	Yi
A	1.43	1.52	5	5	1.475
B	1.53	1.62	7	12	1.575
C	1.56	1.72	4	16	1.675
D	1.73	1.82	5	21	1.775
E	1.83	1.92	1	22	1.885
			22		

Media:

$$\bar{x} = \frac{\sum(Y_i * f_i)}{n}$$

$$\bar{x} = (1.475*5) + (1.575*7) + (1.675*4) + (1.775*5) + (1.875*1) / 22 =$$

$$\bar{x} = 7.375 + 11.025 + 6.7 + 8.875 + 1.875 / 22 =$$

$$\bar{x} = 35.85 / 22 =$$

$$\bar{x} = \mathbf{1.62}$$

Mediana:

$$Me = LI + \left[\frac{\frac{n}{2} - Fi-1}{Fi - Fi-1} \right] * A$$

$$n/2 = 22/2 = 11$$

$$Fi - 1 = 5$$

$$Fi = 12$$

$$LI = 1.53$$

$$A = 0.09$$

$$Me = 1.53 + \left[\frac{11-5}{12-5} \right] * 0.09$$

$$Me = 1.53 + 0.07$$

$$Me = 1.53 + \left[\frac{6}{7} \right] * 0.09$$

$$Me = \mathbf{1.6}$$

$$Me = 1.53 + [0.85] * 0.09$$

Moda:

$$Mo = LI + \left[\frac{\Delta 1}{\Delta 2 + \Delta 1} \right] * A$$

$$\Delta 1 = FI - FI - 1 = 7 - 5 = 2$$

$$\Delta 2 = FI - FI + 1 = 7 - 4 = 3$$

$$A = 0.09$$

$$Mo = 1.53 + \left[\frac{2}{2+3} \right] * 0.09$$

$$Mo = 1.53 + [0.4] * 0.09$$

$$Mo = 1.53 + 0.036$$

$$Mo = \mathbf{1.566}$$

CLASE	LI	LS	fi	Fi	Yi
A	450	650	3	11	550
B	700	900	8	17	800
C	950	1150	6	23	1050
D	1200	1400	3	26	1300
E	1450	1650	2	28	1550
			22		

Media:

$$\bar{x} = \frac{\sum(Yi * fi)}{n}$$

$$\bar{x} = (550*3) + (800*8) + (1050*6) + (1300*3) + (1550*2) / 22 =$$

$$\bar{x} = 1650 + 6400 + 6300 + 3900 + 3100 / 22 =$$

$$\bar{x} = 21350 / 22 =$$

$$\bar{x} = \mathbf{970.45}$$

Mediana:

$$Me = LI + \left[\frac{\frac{n}{2} - Fi - 1}{Fi - Fi - 1} \right] * A$$

$$n/2 = 22/2 = 11$$

$$Fi - 1 = 11$$

$$Fi = 17$$

$$LI = 700$$

$$A = 200$$

$$Me = 700 + \left[\frac{11 - 11}{17 - 11} \right] * 200$$

$$Me = 700 + 1200$$

$$Me = 1900$$

Moda:

$$Mo = LI + \left[\frac{\Delta_1}{\Delta_2 + \Delta_1} \right] * A$$

$$\Delta_1 = FI - FI - 1 = 8 - 3 = 5$$

$$\Delta_2 = FI - FI + 1 = 8 - 6 = 2$$

$$A = 200$$

$$Mo = 700 + \left[\frac{5}{2+5} \right] * 200$$

$$Mo = 700 + [0.71] * 200$$

$$Mo = 700 + 142$$

$$Mo = \mathbf{842}$$

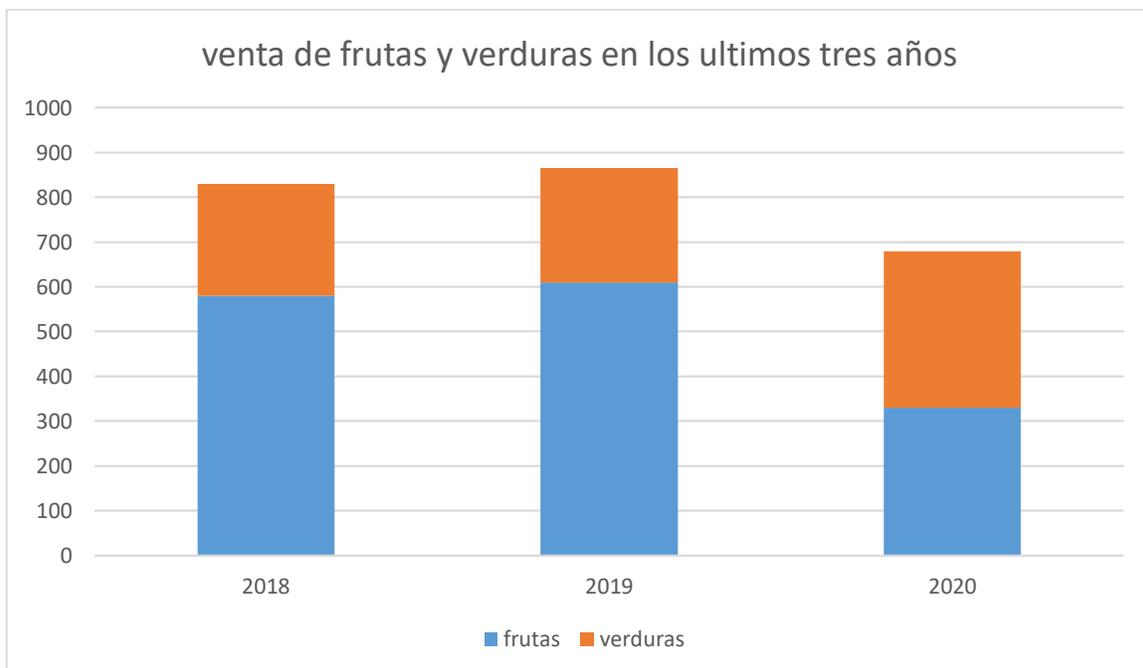
Definición de grafica de barras, pastel (circular), histogramas y polígonos de frecuencia.

Grafica de barras.

Un gráfico de barras es una forma de resumir un conjunto de datos por categorías. Muestra los datos usando varias barras de la misma anchura, cada una de las cuales representa una categoría concreta. La altura de cada barra es proporcional a una agregación específica (por ejemplo, la suma de los valores de la categoría que representa). Las categorías podrían ser desde grupos de edad a ubicaciones geográficas.

Si se aplica al crear el análisis, el gráfico de barras puede mostrar información adicional en líneas de referencia o varios tipos distintos de curvas. Estas líneas o curvas podrían, por ejemplo, mostrar si los puntos de los datos se adaptan bien a un ajuste de curva polinómica determinado, o resumir un conjunto de puntos de datos de muestra ajustándolos a un modelo que describirá los datos y mostrará una curva o una línea recta sobre la visualización. La curva normalmente cambia su aspecto en función de los valores que se hayan filtrado del análisis.

Ejemplo de grafica de barras:



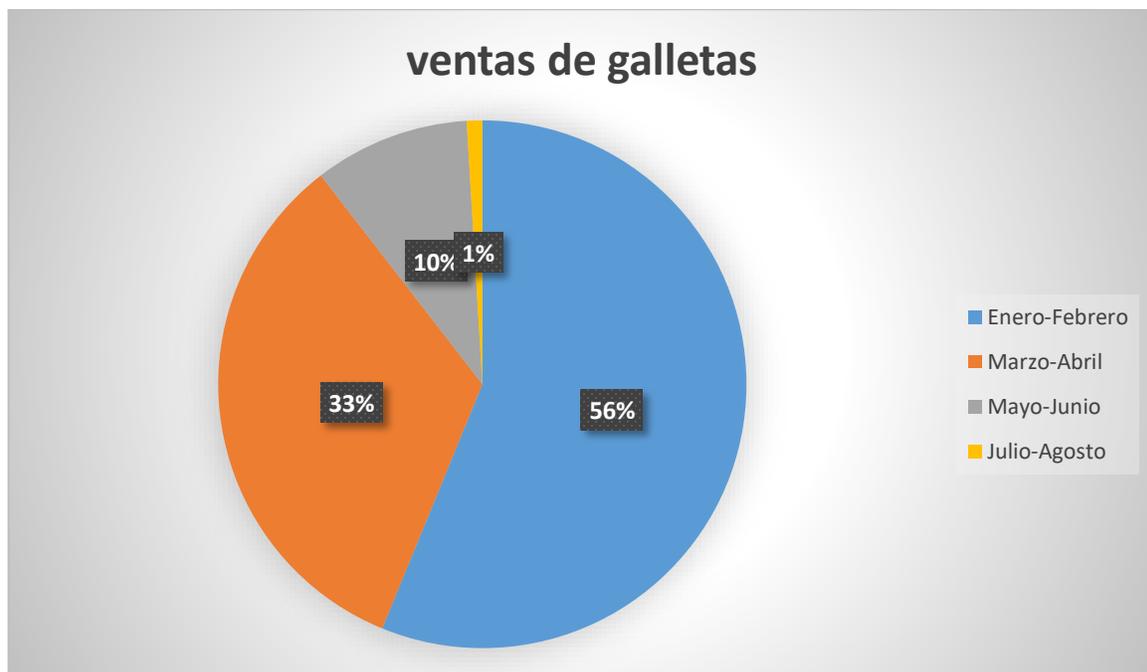
Grafica de pastel

Un diagrama de pastel es un círculo dividido en partes, donde el área de cada parte es proporcional al número de datos de cada categoría.

La gráfica de pastel se usa para representar variables cualitativas o categóricas, de preferencia nominales.

Se utiliza para mostrar la proporción le corresponde a cada categoría.

Ejemplo de grafica de pastel:



Histograma

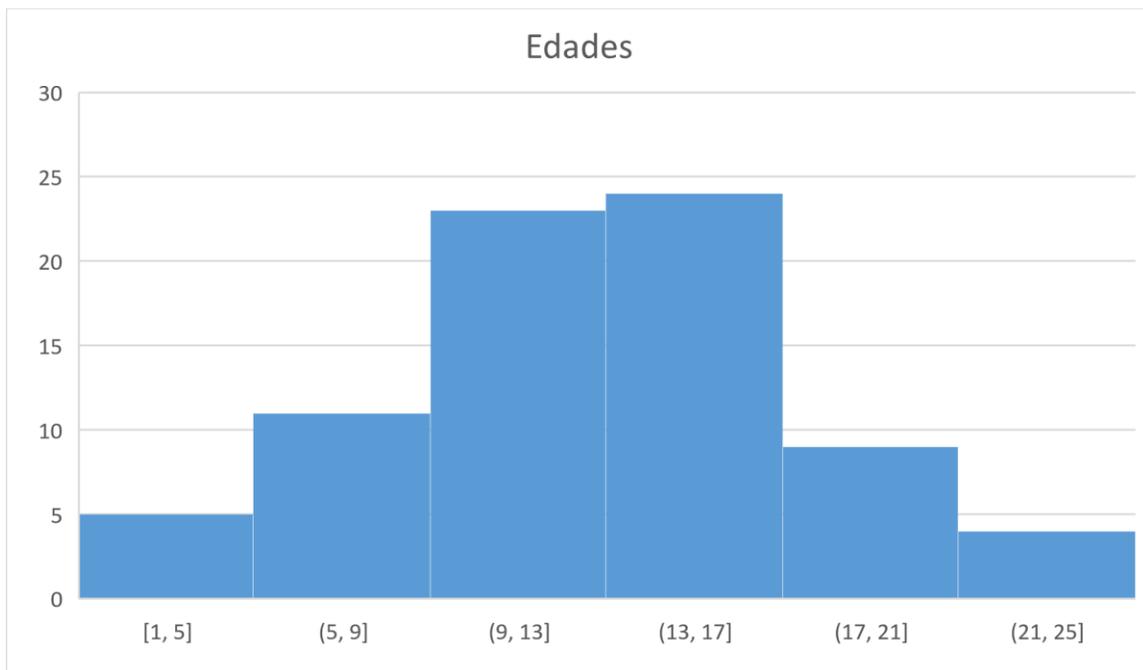
En estadística, un histograma es una representación gráfica de una variable en forma de barras, donde la superficie de cada barra es proporcional a la frecuencia de los valores representados. En el eje vertical se representan las frecuencias, y en el eje horizontal los valores de las variables, normalmente señalando las marcas de clase, es decir, la mitad del intervalo en el que están agrupados los datos.

En términos matemáticos, puede ser definida como una función inyectiva (o mapeo) que acumula (cuenta) las observaciones que pertenecen a cada subintervalo de una

partición. El histograma, como es tradicionalmente entendido, no es más que la representación gráfica de dicha función.

Se utiliza cuando se estudia una variable continua, como franjas de edades o altura de la muestra, y, por comodidad, sus valores se agrupan en clases, es decir, valores continuos. En los casos en los que los datos son cualitativos (no numéricos), como sexto grado de acuerdo o nivel de estudios, es preferible un diagrama de sectores.

Ejemplo de histograma:

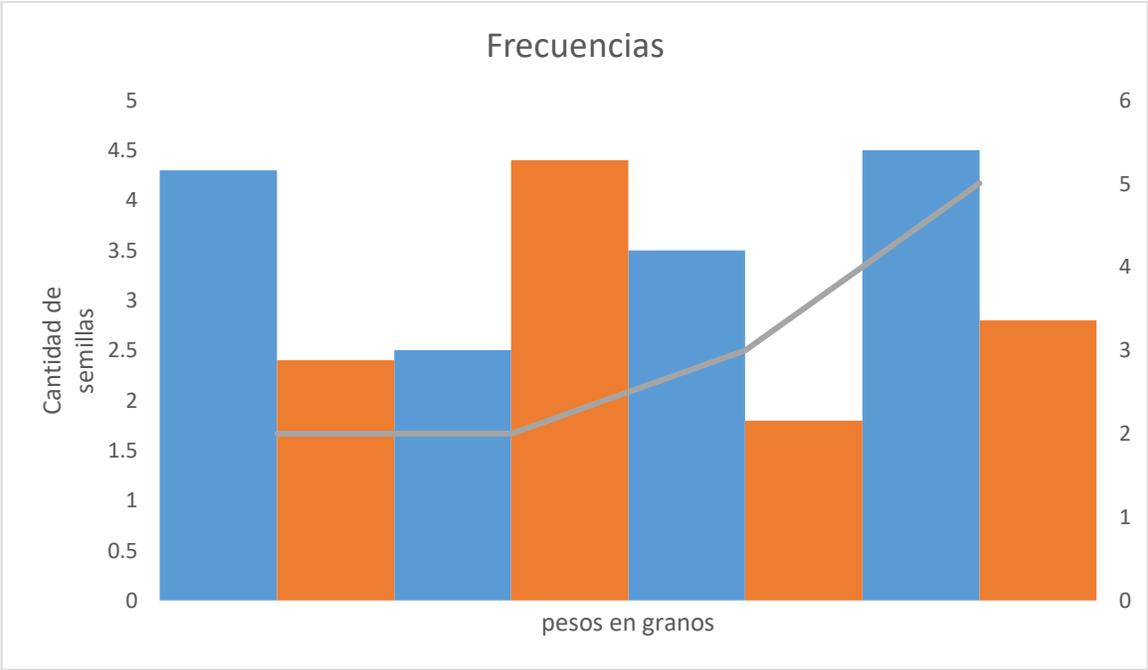


Polígonos de frecuencia

Un polígono de frecuencia es el nombre que tiene el tipo de gráfico que se realiza a partir de un histograma de frecuencia, en los cuales se utilizan columnas verticales para expresar las frecuencias, el polígono se crea al unir los puntos de máxima altura dentro de estas columnas.

Este gráfico es aquel que se puede establecer a partir de la unión de diferentes puntos medios de las columnas que forman la configuración de lo que conocemos

como histograma de frecuencia. El polígono se caracteriza por usar todo el tiempo las columnas de tipo vertical y por nunca dejar espacio entre una columna y otra.



Datos: 6, 6, 7, 7, 8, 8, 9, 9, 10

Media:

$$\bar{x} = \sum x_i / n = \bar{x} = 70/9 = \bar{x} = 7.7$$

Mediana:

$$(n+1) / 2 = (9+1) / 2 = 10/2 = 5$$

6, 6, 7, 7, **8**, 8, 9, 9, 10

Mediana: **8**

Moda: (multimodal) **6, 7, 8, 9**

Rango:

$$R = \text{Max } x_i - \text{Min } x_i = 10 - 6 = 4$$

Varianza:

$$\sigma^2 = (x_1)^2 + (x_2)^2 + (x_3)^2 \dots + (x_n)^2 / n - \bar{x}^2$$

$$\sigma^2 = 6^2 + 6^2 + 7^2 + 7^2 + 8^2 + 8^2 + 9^2 + 9^2 + 10^2 / 9 - 7.777^2$$

$$\sigma^2 = 560/9 - 60.481729 = 62.222222 - 60.481729$$

$$\sigma^2 = \mathbf{1.740493}$$

Desviación

$$S = \frac{\sqrt{\sum (x_i - \bar{x})^2}}{n-1} =$$

$$S = \sqrt{1.740493} = \mathbf{1.319227745376}$$