



**Nombre de alumno: Selene Mancilla  
Avelar**

**Nombre del profesor: Rosario Gómez**

**Nombre del trabajo: introducción a la  
estadística**

**Materia: Estadística**

**Grado: 1er semestre**

**Grupo: semiescolarizado**

# ENSAYO

- **DEFINICIÓN DE ESTADÍSTICA:**

La estadística es una ciencia que estudia los números y los valores de estos, que con el paso de los tiempo ha ido cambiando tanto su valor y los diversos resultados, diferentes físicos, matemáticos y científicos como Galileo Galilei, Jacob Bernoulli, Junto con Arquímedes y Newton, fueron unos de los más grandes matemáticos quienes fueron descifrando cada uno de los obstáculos de las matemáticas si se decide llamarse así, las matemáticas se tiene conocimiento que se deriva de china lo cual es el lugar donde se cree que fueron los indicios de las estadísticas, la cual unos años más adelante se reflejo conocimiento tanto en Grecia como en roma aunque en Egipto también se tomaban datos estadísticos para construir las pirámides.

Las estadísticas son un sinfín de números y letras que nos ayudan a reflejar el resultado correcto de lo que buscamos, aunque también son graficas de de diferentes tamaños y formas con los cuales nos podemos guiar para un mejor entendimiento con base a la estadística.

Un mercader llamado John Graunt fue el primer fundador de la estadística en un pequeño libro, esto fue el primer intento para interpretar fenómenos biológicos de masa y de la conducta social.

Newsman llevó a cabo la primera demostración del Teorema Minimax, base fundamental de la Teoría de juegos.

La estadística es una de las herramientas mas utilizadas en el área empresariales administrativa El término estadística se refiere a datos numéricos, tales como promedios, medianas, porcentajes y números índices que ayudan a entender una gran variedad de negocios y situaciones económicas. Sin embargo, también nos comenta que la estadística no solo se basa en números y es mucho mas que solo datos.

La estadística se define como “el arte y la ciencia de reunir datos, analizarlos, presentarlos e interpretarlos”.

En los negocios y en la economía, la información obtenida al reunir datos, analizarlos, presentarlos e interpretarlos proporciona a directivos, administradores y personas que deben tomar decisiones una mejor comprensión del negocio o entorno económico, permitiéndoles así tomar mejores decisiones con base en mejor información.

La Estadística es de gran importancia en las diferentes empresas, enfocadas desde cualquier área profesional ya que ayudan a lograr una adecuada planeación y control apoyados en los estudios de pronósticos, presupuestos etc.

- **CLASIFICACIÓN DE LA ESTADÍSTICA:**

La estadística se clasifica en dos:

- La Estadística Descriptiva.
- La Estadística Inferencial.

- **Estadística Descriptiva:**

En esta estadística podemos encontrar diferentes tipos de gráficos, con diferente forma tamaños y colores los cuales nos pueden ayudar en diferentes casos y problemas estadísticos, con la finalidad de tener un mayor entendimiento con respecto a este.

La estadística emplea métodos descriptivos y de inferencia estadística. Los primeros se ocupan de la recolección, organización, tabulación, presentación y reducción de la información.

En el caso de la estadística descriptiva se sustituye o reduce el conjunto de datos obtenidos por un pequeño número de valores descriptivos, como pueden ser: el promedio, la mediana, la media geométrica, la varianza, la desviación típica, etc. Estas medidas descriptivas pueden ayudar a brindar las principales propiedades de los datos observados, así como las características clave de los fenómenos bajo investigación. Por lo general, la información proporcionada por la estadística descriptiva puede ser transmitida con facilidad y eficacia mediante una variedad de herramientas gráficas, como pueden ser: Gráficos de tendencia: es un trazo de una característica de interés sobre un periodo, para observar su comportamiento en el tiempo. Gráfico de dispersión: ayuda al análisis de la relación entre dos variables, representado gráficamente sobre el eje x y el correspondiente valor de la otra sobre el eje y. Histograma: describe la distribución de los valores de una característica de interés. Estos métodos gráficos son de mucha utilidad para entender con claridad un fenómeno.

- **La Estadística Inferencial:**

La estadística inferencial estudia la probabilidad de éxito de las diferentes soluciones posibles a un problema en las diferentes ciencias en las que se aplica y para ello utiliza los datos observados en una o varias muestras de la población. Mediante la creación de un modelo matemático infiere el comportamiento de la población total partiendo de los resultados obtenidos en las observaciones de las muestras.

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:** Es el conjunto de todos los posibles elementos que intervienen en un Experimento o en un estudio. La hay de dos tipos:
  - **Población finita:** Es aquella que indica que es posible alcanzarse o sobrepasarse al contar. Es aquella que posee o incluye un número limitado de medidas y observaciones.
  - **Población infinita:** Es infinita si se incluye un gran conjunto de medidas y observaciones que no pueden alcanzarse en el conteo. Son poblaciones infinitas porque hipotéticamente no existe límite en cuanto al número de observaciones que cada uno de ellos puede generar.}
  
- **Muestra:** Un conjunto de medidas u observaciones tomadas a partir de una población dada. Es un subconjunto de la población.
  - **Muestra representativa:** Un subconjunto representativo seleccionado de una población de la cual se obtuvo.
  
- **Muestreo:** Al estudio de la muestra representativa.
  
- **Muestreo Probabilístico:** Se basa en el principio de equiprobabilidad, esto quiere decir que todos los individuos de la muestra seleccionada, tendrán las mismas probabilidades de ser elegidos.
  
- **Muestreo No Probabilístico:** No sirven para hacer generalizaciones, pero sí para estudios exploratorios. En este tipo de muestras, se eligen a los individuos utilizando diferentes criterios relacionadas con las características de la investigación, no tienen la misma probabilidad de ser seleccionados ya que el investigador suele determinar la población objetivo
  
- **Censo:** Al estudio completo de la población.
  
- **Parámetro:** Lo constituyen las características medibles en una población completa. Se le asigna un símbolo representado por una letra griega.
  
- **Estadístico o estadígrafo:** Es la medida de una característica relativa a una muestra. La mayoría de los estadísticos muestrales se encuentran por medio de una fórmula y suelen asignárseles nombres simbólicos que son letras latinas.
  
- **Datos estadísticos (Variables):** Los datos son agrupaciones de cualquier número de observaciones relacionadas. Para que se considere un dato estadístico debe tener 2 características: a) Que sean comparables entre sí. b) Que tengan alguna relación.

- **Variable:** Una característica que asume valores.
- **Censo:** Al estudio completo de la población.
- **Parámetro:** Lo constituyen las características medibles en una población completa. Se le asigna un símbolo representado por una letra griega.
- **Estadístico o estadígrafo:** Es la medida de una característica relativa a una muestra. La mayoría de los estadísticos muestrales se encuentran por medio de una fórmula y suelen asignárseles nombres simbólicos que son letras latinas.
- **Datos estadísticos (Variables):** Los datos son agrupaciones de cualquier número de observaciones relacionadas. Para que se considere un dato estadístico debe tener 2 características: a) Que sean comparables entre sí. b) Que tengan alguna relación.
- **Variable:** Una característica que asume valores.
- **LA MEDIDA DE TENDENCIA CENTRAL:** Parámetro de tendencia central o medida de centralización es un número situado hacia el centro de la distribución de los valores de una serie de observaciones (medidas), en la que se encuentra ubicado el conjunto de los datos.

Las medidas de tendencia central más utilizadas son: media, mediana y moda. Cuando se hace referencia únicamente a la posición de estos parámetros dentro de la distribución, independientemente de que esté más o menos centrada, se habla de estas medidas como medidas de posición.

- **Promedio o media:** La medida de tendencia central más conocida y utilizada es la media aritmética o promedio aritmético. Se representa por la letra griega  $\mu$  cuando se trata del promedio del universo o población y por  $\bar{Y}$  (léase Y barra) cuando se trata del promedio de la muestra. Es importante destacar que  $\mu$  es una cantidad fija mientras que el promedio de la muestra es variable puesto que diferentes muestras extraídas de la misma población tienden a tener diferentes medias.
- **Mediana:** Otra medida de tendencia central es la mediana. La mediana es el valor de la variable que ocupa la posición central, cuando los datos se disponen en orden de magnitud.
- **Moda:** La moda de una distribución se define como el valor de la variable que más se repite. En un polígono de frecuencia la moda corresponde al valor de la variable que está bajo el punto más alto del gráfico. Una muestra puede tener más de una moda.

- **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.
  - **Rango de variación:** Se define como la diferencia entre el mayor valor de la variable y el menor valor de la variable.
  - **Desviación media:** Es la media aritmética de los valores absolutos de las diferencias de cada dato respecto a la media.
  - **Desviación estándar:** La desviación estándar mide el grado de dispersión de los datos con respecto a la media, se denota como  $s$  para una muestra o como  $\sigma$  para la población.
  - **Varianza:** Es otro parámetro utilizado para medir la dispersión de los valores de una variable respecto a la media. Corresponde a la media aritmética de los cuadrados de las desviaciones respecto a la media.
  - **Coefficiente de Variación:** Permite determinar la razón existente entre la desviación estándar ( $s$ ) y la media. Se denota como CV. El coeficiente de variación permite decidir con mayor claridad sobre la dispersión de los datos.

## Ejemplos

### 1.- Calcular la Media Aritmética del Conjunto de Datos { 1, 3, 5, 7, 9, 11, 13 }

$$\bar{x} = \Sigma x / n$$

$$\bar{x} = (1+3+5+7+9+11+13) / 7$$

$$\bar{x} = 49 / 7$$

$$\bar{x} = 7$$

### 2.- Detectar la Moda del Conjunto de Datos { 1, 3, 3, 4, 4, 4, 5, 5, 5, 5, 6, 6, 6, 7, 9, 9, 11, 13, 13 }

Hay que ver cuántas veces se encuentra enlistado cada término del conjunto

1: 1 vez, 3: 2 veces, 4: 3 veces, 5: 4 veces, 6: 3 veces, 7: 1 vez, 9: 2 veces, 11: 1 vez, 13: 2 veces

$M_o = 5$ , con 4 apariciones

La primera solución puede ser calcular la media de todas las desviaciones, es decir, si consideramos como muestra la de todas las desviaciones y calculamos su media. Pero esta solución es mala pues como veremos siempre va a ser 0.

$$\bar{D} = \sum_{i=1}^n \frac{d_i \cdot n_i}{N} = \sum_{i=1}^n \frac{(x_i - \bar{x}) \cdot n_i}{N} = \sum_{i=1}^n \frac{x_i \cdot n_i}{N} - \sum_{i=1}^n \frac{n_i}{N} \cdot \bar{x} = \bar{x} - \bar{x} = 0$$

Luego por lo tanto esta primera idea no es valida, pues las desviaciones positivas se contrarrestan con las negativas.

Para resolver este problema, tenemos dos caminos:

Tomar el valor absoluto de las desviaciones. Desviación media

Elevar al cuadrado las desviaciones. Varianza.

Desviación media:

Es la media de los valores absolutos de las desviaciones, y la denotaremos por  $d_m$ .

$$d_m = \sum_{i=1}^n \frac{|d_i| \cdot n_i}{N} = \sum_{i=1}^n \frac{|x_i - \bar{x}| \cdot n_i}{N}$$

Varianza:

Es la media de los cuadrados de las desviaciones, y la denotaremos por  $S_x^2$  o también por  $\sigma_x^2$ .

$$S_x^2 = \sigma_x^2 = \sum_{i=1}^n \frac{d_i^2 \cdot n_i}{N} = \sum_{i=1}^n \frac{(x_i - \bar{x})^2 \cdot n_i}{N}$$

Aunque también es posible calcularlo como:

$$S_x^2 = \sigma_x^2 = \sum_{i=1}^n \frac{x_i^2 \cdot n_i}{N} - \bar{x}^2$$

- **Tablas de frecuencias con datos no agrupados**

**Ejemplo 1:**

Se le pidió a un grupo de personas que indiquen su color favorito, y se obtuvo los siguientes resultados:

negro azul amarillo rojo azul  
 azul rojo negro amarillo rojo  
 rojo amarillo amarillo azul rojo  
 negro azul rojo negro amarillo

Con los resultados obtenidos, elaborar una tabla de frecuencias.

**Solución:**

En la primera columna, colocamos los valores de nuestra variable, en la segunda la frecuencia absoluta, luego la frecuencia acumulada, seguida por la frecuencia relativa, y finalmente la frecuencia relativa acumulada. Por ser el primer problema, no haremos uso de las frecuencias porcentuales.

Color	Frecuencia absoluta	Frecuencia acumulada	Frecuencia relativa	Frecuencia relativa acumulada
Negro	4	4	0,20	0,20
Azul	5	9	0,25	0,45
Amarillo	5	14	0,25	0,70
Rojo	6	20	0,30	1
Total	20			

**Ejemplo 2:**

En una tienda de autos, se registra la cantidad de autos Toyota vendidos en cada día del mes de Setiembre.

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

Con los datos obtenidos, elaborar una tabla de frecuencias.

**Solución:**

En la primera columna, colocamos los valores de nuestra variable, en la segunda la frecuencia absoluta, luego la frecuencia acumulada, seguida por la frecuencia relativa, y finalmente la frecuencia relativa acumulada. Ahora vamos a agregar la columna de frecuencia porcentual, y frecuencia porcentual acumulada.

Autos vendidos		Frecuencia absoluta		Frecuencia acumulada		Frecuencia	
relativa		Frec. relativa	relativa acumulada	Frecuencia porcentual		Frec. porcentual	porcentual acumulada
0	8	8	0,267	0,267	26,7%	26,7%	
1	7	15	0,233	0,500	23,3%	50,0%	
2	7	22	0,233	0,733	23,3%	73,3%	
3	5	27	0,167	0,900	16,7%	90,0%	
4	3	30	0,100	1	10,0%	100%	
Total	30		1		100%		

- **Tablas de frecuencias con datos agrupados**

**Ejemplo 1:**

Las notas de 35 alumnos en el examen final de estadística, calificado del 0 al 10, son las siguientes:

0; 0; 0; 0; 1; 1; 1; 1; 2; 2; 2; 3; 3; 3; 3; 4; 4; 4; 4; 5; 5; 5; 5; 6; 6; 6; 7; 7; 7; 8; 8; 8; 9; 10; 10.

Con los datos obtenidos, elaborar una tabla de frecuencias con 5 intervalos o clases.

**Solución:**

Hallamos el rango:  $R = X_{\max} - X_{\min} = 10 - 0 = 10$ .

El número de intervalos (k), me lo da el enunciado del problema:  $k = 5$ .

Calculamos la amplitud de clase:  $A = R/k = 10/5 = 2$ .

Ahora hallamos los límites inferiores y superiores de cada clase, y elaboramos la tabla de frecuencias.

Intervalo	Marca de clase	Frecuencia absoluta	Frecuencia acumulada	
	Frecuencia relativa	Frec. relativa acumulada		
[0 – 2)	1	8	0,229	0,229
[2 – 4)	3	7	0,200	0,429
[4 – 6)	5	8	0,229	0,658
[6 – 8)	6	6	0,171	0,829
[8 – 10]	9	6	0,171	1
Total	35	1		

**Ejemplo 2:**

Un grupo de atletas se está preparando para una maratón siguiendo una dieta muy estricta. A continuación, viene el peso en kilogramos que ha logrado bajar cada atleta gracias a la dieta y ejercicios.

0,2	8,4	14,3	6,5	3,4
4,6	9,1	4,3	3,5	1,5
6,4	15,2	16,1	19,8	5,4
12,1	9,6	8,7	12,1	3,2

Elaborar una tabla de frecuencias con dichos valores.

**Solución:**

Hallamos el rango:  $R = X_{\max} - X_{\min} = 19,8 - 0,2 = 19,6$ .

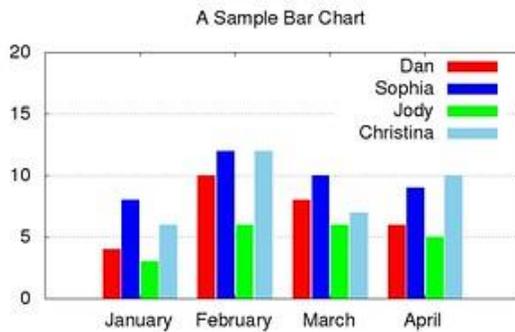
El número de intervalos (k), lo calculamos usando la regla de Sturges:  $k = 1 + 3,322 \log(n)$   
 $= 1 + 3,322 \cdot \log(20) = 5,32$ . Podemos redondear el valor de k a 5

Calculamos la amplitud de clase:  $A = R/k = 19,6/5 = 3,92$ . Redondeamos a 4.

Ahora hallamos los límites inferiores y superiores de cada clase, y elaboramos la tabla de frecuencias.

Intervalo	Marca de clase	Frecuencia absoluta	Frecuencia acumulada
	Frecuencia relativa	Frec. relativa acumulada	
[0 – 4)	2	5	0,25
[4 – 8)	6	10	0,50
[8 – 12)	10	4	0,20
[12 – 16)	14	4	0,20
[16 – 20]	18	2	0,10
Total	20	1	

- GRÁFICA DE BARRA:** Los gráficos de barra son aquellos que revelan cotejos entre elementos individuales. En este tipo de gráficas, las categorías se muestran organizadas de manera vertical; mientras que los valores se ordenan horizontalmente.

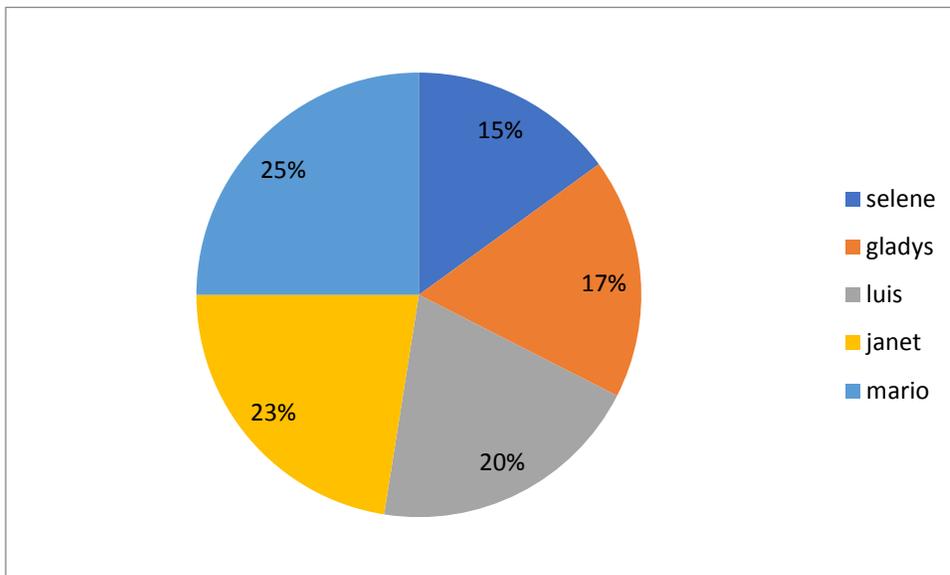


## LA GRAFICA DE PASTEL

La grafica de pastel o diagrama de pastel se utiliza para representar variables cuantitativas o categorías de preferencias nominales. Se deben representar las proporciones para cada categoría de la variable. La suma de las proporciones no debe exceder al 100%, Se debe utilizar para representar máximo 5 categorías.

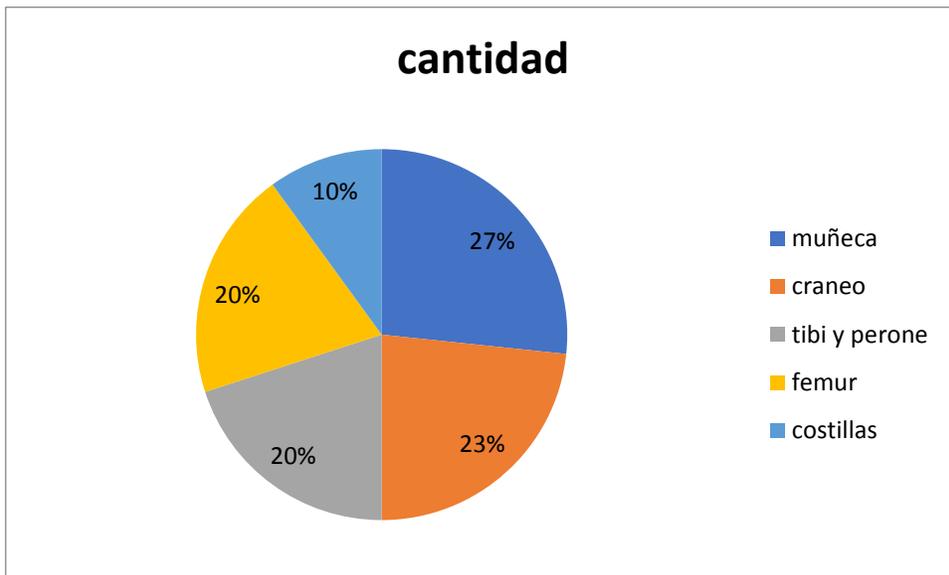
### Ejemplo de gráfica:

		fi	fa	(x-x) <sup>2</sup>
Selene	6	2	2	3.13
Gladys	7	2	4	.59
Luis	8	2	6	.05
Janet	9	2	8	1.51
Mario	10	1	9	4.97



**Ejemplo dos de grafica de pastel:**

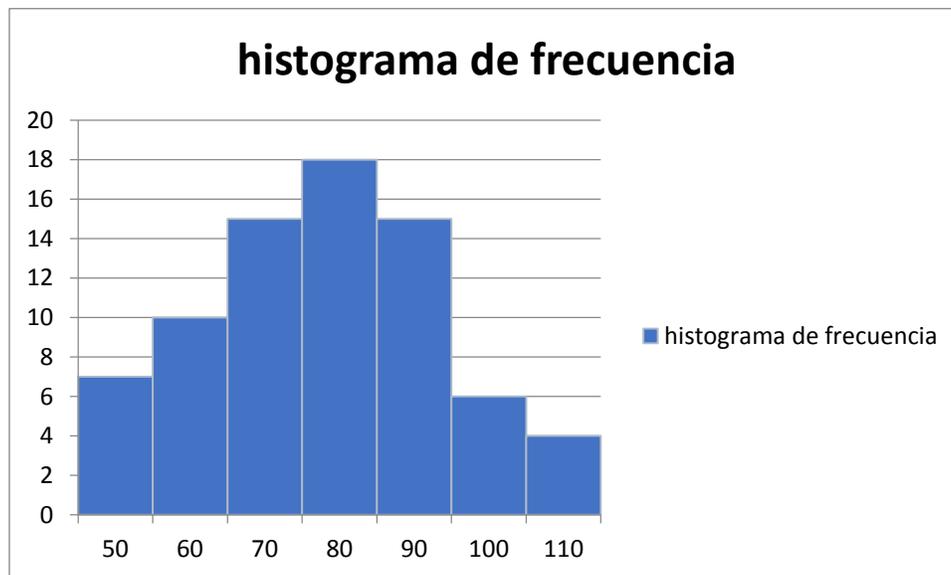
huesos rotos	cantidad
muñeca	8
cráneo	7
tibi y peroné	6
fémur	6
costillas	3



## Histograma

Es una representación gráfica variable en forma de barras donde la superficie de Cada barra es proporcional a la frecuencia de valores representados.

	fa	marca	fc
45-55	7	50	7
55-65	10	60	17
65-75	15	70	32
75-85	18	80	50
85-95	15	90	65
95- 105	6	100	71
105-115	4	110	75

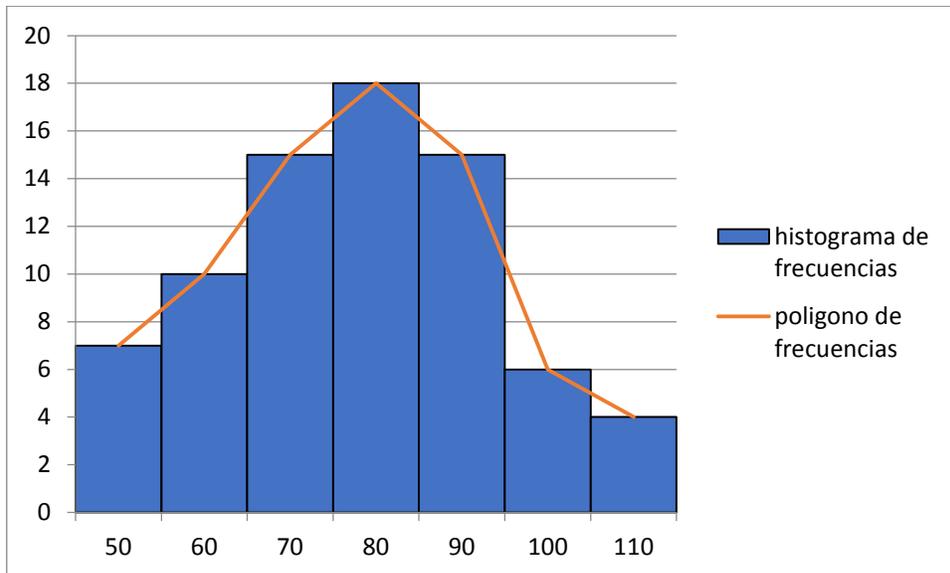


## Polígono de frecuencia

Los polígonos de frecuencia se forman a ytravez de un histograma de frecuencia. Es realizado uniendo los puntos de mayor altura de estas columnas

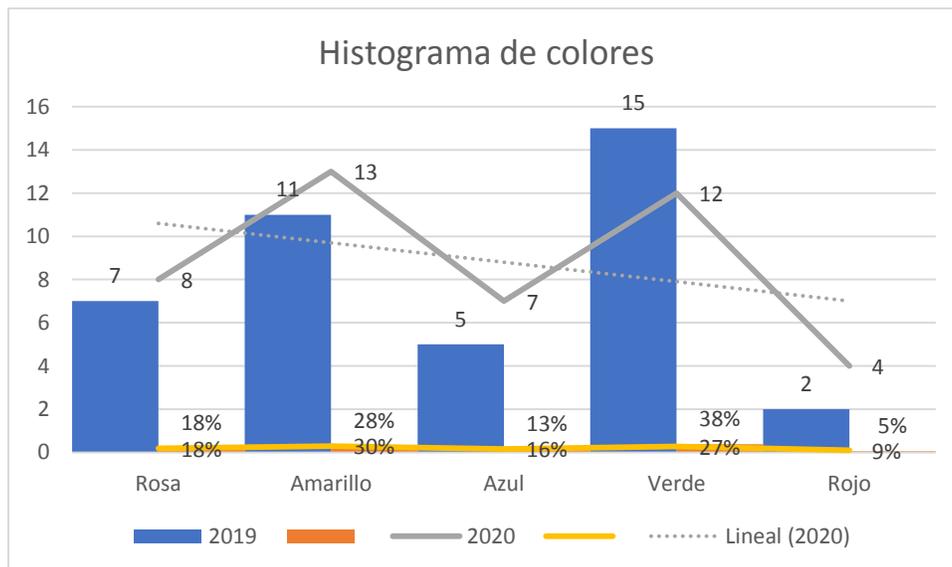
Ejemplo:

	fa	marca	fc
45-55	7	50	7
55-65	10	60	17
65-75	15	70	32
75-85	18	80	50
85-95	15	90	65
95- 105	6	100	71
105-115	4	110	75



Ejemplo 2 de histograma y polígono de frecuencia.

	2019		2020		fr
Rosa	7	18%	8	18%	
Amarillo	11	28%	13	30%	
Azul	5	13%	7	16%	
Verde	15	38%	12	27%	
Rojo	2	5%	4	9%	
	40		44		



## 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.

calificaciones	fi	fa	x-x)2
6	2	2	3.13
7	2	4	0.59
8	2	6	0.5
9	2	8	1.51
10	1	9	4.97

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

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{70}{9} = 7.77$$

$$\bar{X} = \frac{\sum X}{n} = 7.77$$

• Mediana

6 6 7 7 8 8 9 9 10      Datos impares

Mediana

6 6 7 7 8 8 9 9 10 6      Datos pares

8 mediana

• Moda

6 6 7 7 8 8 9 9 10

Tenemos cuatrimodal ó multimodal

	fi	f	x-x)²
Selene	6	2	3.13
Glady	7	2	0.59
Luis	8	2	0.05
Janet	9	2	1.51
Francis	10	1	4.97

- Moda

6 6 7 7 8 8 9 9 10

Tenemos cuatrimodal ó multimodal

		$f_i$	$f_n$	$(x - \bar{x})^2$
Selene	6	2	2	3.13
Glady	7	2	4	.59
Luis	8	2	6	.05
Janet	9	2	8	1.51
Ignacio	10	1	9	4.93

$$\bar{x} = \frac{\sum X}{n} = 7.77$$

- Varianza

$$S^2 = \frac{\sum (X - \bar{x})^2}{n - 1} = \frac{10.25}{8} = \boxed{1.28}$$

- Desviación estandar  $\sqrt{1.28} = 1.13$