

Señores estudiantes grados Octavos, a continuación, encontrarán las definiciones y ejercicios de estadística, bajados de internet está dividido en varias actividades para desarrollar y entregar en la fecha y día que se les indique en cada curso.

ESTADÍSTICA: TABLAS Y GRÁFICOS

1.- ORIGEN Y SIGNIFICADO DE LA ESTADÍSTICA

En sus orígenes históricos, la Estadística estuvo ligada a cuestiones de Estado (recuentos, censos, etc.) y de ahí proviene su nombre. Hoy en día está presente en todos los ámbitos humanos, tanto individuales como colectivos.

La Estadística surge ante la necesidad de poder tratar y comprender conjuntos numerosos de datos. Los estudios estadísticos, en la actualidad, impregnan numerosos ámbitos: sanidad, mundo empresarial, medios de comunicación, etc.

Definición.- La Estadística es la ciencia que se ocupa de la recogida de datos, su organización y análisis; así como de las predicciones que, a partir de estos datos, pueden hacerse.

Fases de un estudio estadístico:

- Recogida de datos
- Recuento de datos: tablas y gráficos estadísticos.
- Análisis de los datos: parámetros estadísticos.
- Extracción de conclusiones de los datos.
- Toma de decisiones.

Podemos distinguir entre dos clases de Estadística:

Definición.- La Estadística descriptiva se ocupa de tomar los datos de un conjunto, organizarlos en tablas o en representaciones gráficas y del cálculo de unos números que nos informen de manera global del conjunto estudiado.

Definición.- La Estadística inferencial trata sobre la elaboración de conclusiones para la población, partiendo de los resultados de una muestra y del grado de fiabilidad de estas conclusiones.

2.- CONCEPTOS BÁSICOS EN UN ESTUDIO ESTADÍSTICO.

- **Población.-** Es el conjunto formado por todos los elementos que existen para el estudio de un determinado fenómeno.
- **Individuo u objeto.-** Es cada elemento de la población.
- **Muestra.-** Es el subconjunto que tomamos de la población para determinar el estudio del fenómeno.
- **Tamaño de la muestra.-** Es el número de individuos que componen la muestra.
- **Variable o carácter estadístico.-** Es la cualidad o propiedad del elemento de la población que se analiza en el estudio estadístico.

Tipos de variables:

- **Variables cualitativas:** no se pueden medir y se describen con palabras.
Ejemplos: razas de perro, estado civil de una persona, color de ojos...
- **Variables cuantitativas:** se pueden medir y expresar con números.

Pueden ser de dos tipos:

- **Discretas:** solo pueden tomar un número finito de valores numéricos:

Ejemplos: número de hermanos, número de parados de una ciudad, número de habitantes menores de edad...

- **Continuas:** pueden tomar cualquier valor en un intervalo dado.

Ejemplos: estatura de una persona, peso de una persona...

Ejemplo completo:

Se desea realizar un estudio sobre el número de hijos por familia en el Municipio de Chía

- Población: Familias de Chía
- Individuo: cada una de las familias.
- Muestra: elegimos una muestra aleatoria (al azar) de 1000 familias distribuidas por todo el municipio
- Tamaño de la muestra: 1000
- Variable: número de hijos. Es una variable cuantitativa y discreta.

ACTIVIDAD 1- REALIZARLA EN EL CUADERNO

1.- Clasifica las siguientes variables estadísticas:

- a) Número de músculos de los animales vertebrados.
- b) Intención de voto.
- c) Velocidad que, en un instante dado, llevan las motocicletas que circulan por las carreteras y calles de una gran ciudad española.
- d) Talla de pantalones de los estudiantes de tu curso
- e) Tipos de comida que prefieren los adolescentes.
- f) Temperatura mínima en tu ciudad cada día del año.
- g) Las marcas de los carros que venden en Chía
- h) Deporte practicado por los chicos y chicas de tu colegio
- i) La duración de cada pila eléctrica producida por una empresa durante un semestre.

3.- CONTEO. TABLAS ESTADÍSTICAS.

El primer paso de cualquier estudio estadístico es **recoger los datos**. Normalmente se suele llevar a cabo a través de encuestas o entrevistas, según la población a estudiar, su tamaño, el tiempo de que dispongamos.

Una vez que tenemos los datos recogidos, pasamos a hacer el **conteo**: contando el número de veces que aparece cada valor de la variable a estudiar.

Ejemplo 1.- Preguntamos a 20 estudiantes el número de miembros de su familia, y sus respuestas fueron:

3, 5, 4, 3, 5, 6, 8, 3, 3, 5, 7, 5, 6, 5, 4, 4, 7, 4, 5, 3

Miembros por familia x_i	Frecuencia
3	5
4	4
5	6
6	2
7	2
8	1

Los valores de las variables estadísticas continuas se agrupan por **intervalos o clases**. Además, si la variable es discreta y toma muchos valores, también se suele agrupar por **intervalos o clases**.

El valor medio de cada clase o intervalo se llama **marca de clase** y se calcula como la semisuma de los extremos del intervalo.

Para construir los intervalos tenemos que tener en cuenta:

- Es conveniente que el número de intervalos que debemos considerar en cualquier estudio esté entre 5 y 10.
- Usualmente tomamos los intervalos con igual amplitud o longitud.
- El **recorrido de la variable** es la diferencia entre el valor más grande y el más pequeño. La amplitud de cada intervalo se calcula dividiendo el recorrido de la variable entre el número total de intervalos.

Ejemplo 2.- A los 100 empleados de una empresa de piezas de precisión, se les ha realizado una prueba de habilidad manual. En una escala de 0 a 100 se han obtenido las siguientes puntuaciones:

27, 66, 32, 55, 46, 37, 75, 81, 18, 33, 47, 74, 37, 52, 47, 66, 80, 87, 37, 29,
 46, 15, 29, 90, 76, 67, 23, 35, 94, 23, 25, 56, 73, 78, 17, 28, 76, 58, 45, 36,
 55, 60, 17, 56, 23, 82, 64, 50, 51, 45, 37, 65, 62, 26, 69, 36, 54, 42, 40, 54,
 27, 62, 28, 65, 46, 92, 36, 33, 23, 66, 18, 82, 47, 49, 59, 45, 73, 43, 47, 83,
 78, 65, 39, 36, 53, 91, 38, 35, 68, 78, 91, 23, 34, 43, 55, 56, 74, 56, 62, 38.

Observamos que los valores extremos son 15 y 94. La amplitud total entre los datos es de 80 puntos, ya que ambas puntuaciones están incluidas.

Agruparemos los datos en 8 intervalos de amplitud 10:

$(14,24]$, $(24,34]$, ..., $(84,94]$. Realizando el conteo con atención, se obtiene la tabla que sigue:

Habilidad manual	Marca de clase	Frecuencias
	x_i	f_i
$(14,24]$	19	10
$(24,34]$	29	12
$(34,44]$	39	17
$(44,54]$	49	18
$(54,64]$	59	13
$(64,74]$	69	13
$(74,84]$	79	11
$(84,94]$	89	6

Para construir una **tabla estadística completa** tenemos que calcular:

- **Frecuencia absoluta (f_i) de cada valor x_i :** es el número total de veces que aparece el dato x_i .
- **Frecuencia absoluta acumulada (F_i) de cada valor x_i :** es la suma de todas las frecuencias absolutas correspondientes a los valores anteriores a x_i y a la suya propia. No tiene sentido para variables cualitativas.
- **Frecuencia relativa (h_i) de cada valor x_i :** se calcula dividiendo la frecuencia absoluta correspondiente f_i entre el número total de datos N. $h_i = \frac{f_i}{N}$
- **Frecuencia relativa acumulada (H_i) de cada valor x_i :** es la suma de todas las frecuencias relativas correspondientes a los valores anteriores a x_i y a la suya propia. No tiene sentido para variables cualitativas.

Ejemplo 1.- Preguntamos a 20 alumnos el número de miembros de su familia, y sus respuestas fueron:
3, 5, 4, 3, 5, 6, 8, 3, 3, 5, 7, 5, 6, 5, 4, 4, 7, 4, 5, 3

Miembros por familia x_i	Frecuencia absoluta f_i	Frecuencia absoluta acumulada F_i	Frecuencia relativa h_i	Frecuencia relativa acumulada H_i
3	5	5	0,25	0,25
4	4	9	0,2	0,45
5	6	15	0,3	0,75
6	2	17	0,1	0,85
7	2	19	0,1	0,95
8	1	20	0,05	1

ACTIVIDAD

2.- Una determinada especie de mamíferos tiene en cada parto un número variable de hijos. Se observa que las camadas de 35 familias durante un año han sido las que se recogen en la tabla adjunta:

Número de hijos	0	1	2	3	4	5	6	7
Número de familias	2	3	10	10	5	0	5	0

Elabora una tabla estadística completa con todos los tipos de frecuencias existentes.

3.- La realización de una prueba de habilidad motora por parte de 60 niños han dado los resultados que siguen:

15, 35, 18, 23, 75, 81, 19, 27, 15, 18, 63, 45, 31, 32, 45, 18, 29, 17, 30, 77,
76, 75, 19, 15, 23, 35, 81, 15, 81, 41, 76, 24, 27, 69, 15, 18, 13, 18, 76, 14,
29, 31, 52, 46, 18, 17, 35, 62, 44, 31, 18, 27, 32, 74, 19, 31, 47, 19, 82, 50.

- Agrupar estos datos en intervalos de amplitud 5, realizando la correspondiente tabla estadística completa.
- Responder a las mismas cuestiones del apartado anterior tomando clases de amplitud 10.

4.- En 1797 el científico inglés Henry Cavendish midió la densidad de la Tierra a través de una balanza de torsión. Realizó 29 observaciones y obtuvo los siguientes valores (en g/cm^3).

5,50 5,61 4,88 5,07 5,26 5,55 5,36 5,29 5,58 5,65
5,57 5,53 5,63 5,29 5,44 5,34 5,79 5,10 5,27 5,39
5,42 5,47 5,63 5,34 5,46 5,30 5,75 5,68 5,85

Agrupar los datos en 5 clases de amplitud 0,25, considerando como límite inferior de la primera clase el valor 4,75 y construir la correspondiente tabla completa de frecuencias.

4.- GRÁFICOS PARA VARIABLES ESTADÍSTICAS CUALITATIVAS.

La recopilación de datos y la tabulación pueden traducirse gráficamente mediante representaciones convenientemente elegidas: barras, sectores circulares, mapas curvas, etc.

Los gráficos permiten visualizar e interpretar el fenómeno que se estudia, en forma más clara.

Las **barras** se utilizan generalmente para representar atributos cualitativos o cuantitativos discreto. La longitud es igual a la frecuencia de cada observación. Pueden ser barras simples o múltiples, según se trate de representar uno o más atributos.

Las barras pueden ser horizontales o verticales.

Gráf. de barras: Evaluación del gobierno X

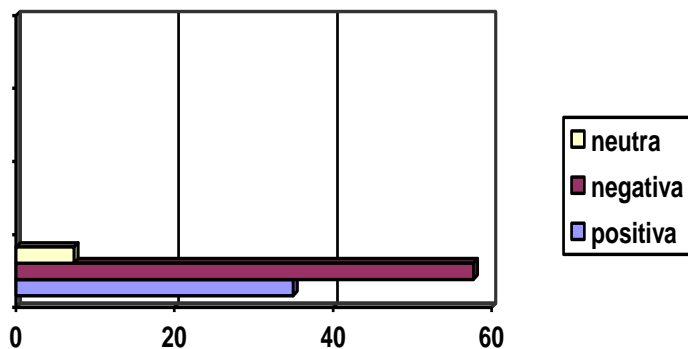
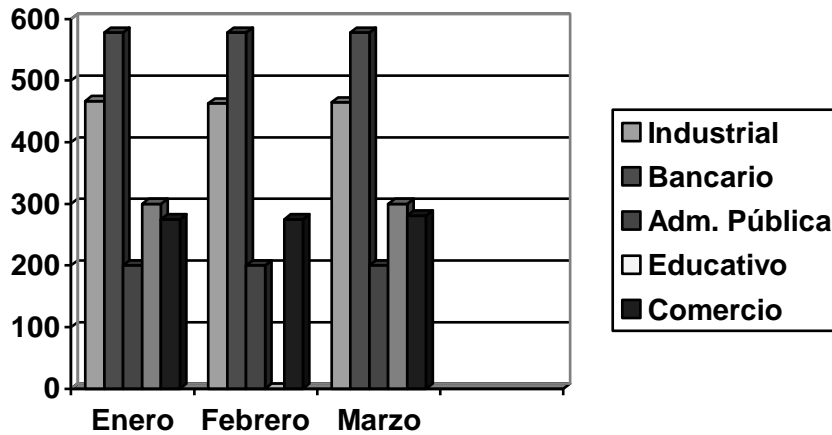
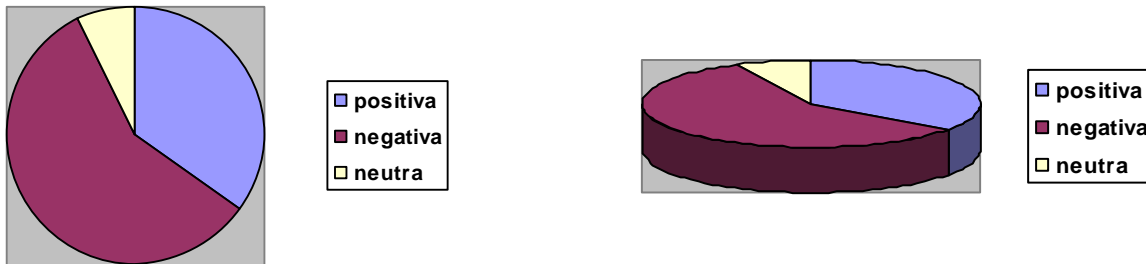


Gráfico de barras compuesto: Remuneraciones medias (año Z)



Los **gráficos circulares** o **gráficos de torta** son útiles para comparar datos pues, en general, trabajan con porcentuales. El área de cada sector representa el porcentaje que corresponde a la frecuencia de un cierto valor de la variable. Esta representación es conveniente cuando el número de sectores es pequeño y sus áreas están bien diferenciadas.



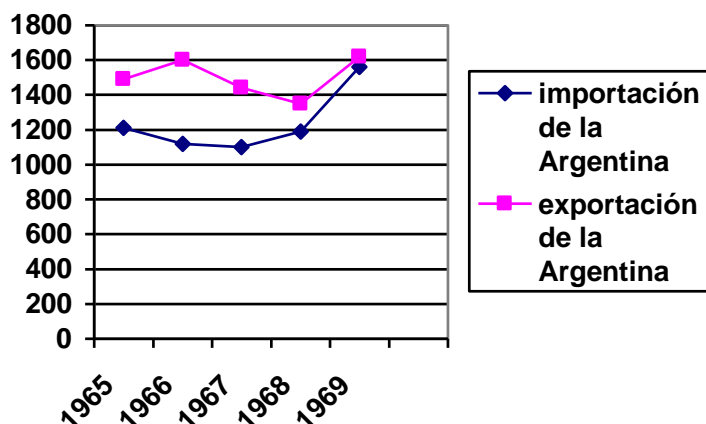
Evaluación del gobierno X

El **histograma** se utiliza para representar una tabla de frecuencias de intervalos de clase. Sobre el eje horizontal se representan los intervalos de clase y sobre el eje vertical, las frecuencias de los intervalos. El gráfico consiste en un conjunto de rectángulos adyacentes cuya base representa un intervalo de clase y cuya altura representa la frecuencia del intervalo.

El **polígono de frecuencias** se construye uniendo los puntos medios de los lados opuestos de las bases de cada rectángulo. Si se quiere cerrar el rectángulo, se agregan dos intervalos: uno anterior y otro posterior al último y se prolonga el polígono hasta los puntos medios de estos intervalos.

Las **curvas** se utilizan generalmente para representar la variación de una variable a través del tiempo (años, meses, horas, etc.). Sobre el eje horizontal figuran los períodos de tiempo.

Variación del valor de las importaciones y exportaciones de la Argentina en millones de dólares



Estas son sólo algunas de las formas posibles de graficación y las que encontrarás con más frecuencia.

c) Análisis y medición de datos

Para describir un conjunto de datos, se calculan algunas medidas que resumen la información y que permiten realizar comparaciones.

Medidas de posición: se utilizan para encontrar un valor que represente a todos los datos. Las más importantes son: la **media aritmética**, la **moda** y la **mediana**.

- La **media aritmética** o **promedio** (\bar{x}) de varios números se calcula como el *cociente entre la suma de todos esos números y la cantidad de números que sumamos*.
- La **moda (Mo)** es el *valor que más se repite*. Puede suceder que haya más de una moda o ninguna (si todos los valores tienen igual frecuencia).
- La **mediana (Me)** es el *valor que ocupa el lugar central al ordenar los datos de menor a mayor*. Si la cantidad de datos es par, la mediana es el promedio entre los dos valores centrales.

EJEMPLO: El entrenador de un equipo de natación debe elegir a uno de sus integrantes para la próxima competencia de estilo libre. Según los tiempos en segundos que obtuvieron los postulantes de las cinco últimas carreras de 100 m de estilo libre, ¿qué nadador le conviene elegir?

Diego	61,7	61,7	62,3	62,9	63,1
Tomás	61,5	62,9	62,9	63,7	63,7
Sergio	60,7	62,4	62,7	62,7	63,2

Para poder decidir, calcula las medidas de posición de cada uno.

	promedio	moda	mediana
Diego	62,34	61,7	62,3
Tomás			
Sergio			

En promedio, los nadadores más rápidos son y, pero esto no significa que hayan tenido el mismo rendimiento; por eso necesitamos las otras medidas de posición: de ellos dos, tanto la moda como la mediana indican que fue más veloz. Sin embargo, para elegir el nadador adecuado, no basta con considerar las medidas de posición, ya que también es necesario que su rendimiento *sea parejo*, es decir, que los tiempos de sus 100 m libres *no tengan mucha dispersión*.

Medidas de dispersión: nos informan cómo están distribuidos los datos. La más importante es el **desviación estándar (σ)**, que mide *la dispersión de los datos con respecto al promedio*. Cuanto menor es el desvío estándar, menos dispersos están los datos con respecto al promedio.

Para calcular el desvío estándar, seguimos los siguientes pasos:

- Calculamos la diferencia entre cada uno y el promedio.
- Elevamos al cuadrado cada una de las diferencias anteriores.
- Sumamos todos los valores hallados en el paso anterior y dividimos el resultado por la cantidad de datos. Así obtenemos la **varianza**.
- Calculamos el **desviación estándar (σ)** como la **raíz cuadrada de la varianza**.

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}} \quad n: \text{número de datos}$$

- Diego y Sergio, dos de los nadadores del ejercicio anterior, obtuvieron el mismo promedio y sin embargo sus tiempos están distribuidos de manera diferente.
Calcula los desvíos estándares de los tiempos de los nadadores

Tiempos de Diego

x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
61,7	-0,64	
61,7	-0,64	
62,3	-0,04	
62,9	0,56	
63,1	0,76	
total		

Tiempos de Sergio

x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
total		

$$\sigma_{Diego} = \sqrt{\frac{\quad}{5}} \cong \dots\dots\dots$$

$$\sigma_{Sergio} = \sqrt{\frac{\quad}{\quad}} \cong \dots\dots\dots$$

Entonces:

Podemos ver que el desvío estándar de es menor que el de, lo cual indica que el promedio representa mejor los datos de, porque sus tiempos fueron menos dispersos.

Entonces, aunque cinco datos son muy pocos para hacer estadística, si con esa información hay que elegir un nadador de ese equipo para la próxima competencia, conviene que sea

ACTIVIDAD

5.- Realizar el trabajo de todos los estudiantes del grado octavo en que se encuentre con la edad de cada estudiante, encontrar la tabla con los datos de frecuencias y graficar los resultados mediante diagrama de barras y circulas

6.- Las dianas logradas en un campeonato por 25 tiradores fueron:

8, 10, 12, 12, 10, 10, 11, 11, 10, 13, 9, 11, 10, 9, 9, 11, 12, 9, 10, 9, 10, 8, 10

Resume los datos anteriores en una tabla de frecuencias absolutas y relativas, y dibuja el correspondiente diagrama de barras.

7.- Se ha realizado un test de habilidad numérica a los alumnos de una clase. Los resultados obtenidos son:

Puntos	[10,15)	[15,20)	[20,25)	[25,30)	[30,35)	[35,40)	[40,45)	[45,50)
Nº de alumnos	4	6	6	10	8	10	3	3

Representa los datos mediante un histograma

8.- Se ha aplicado un test a los empleados de una fábrica, obteniéndose la siguiente tabla:

x	(38,44]	(44,50]	(50,56]	(56,62]	(62,68]	(68,74]	(74,80]
Nº trabajadores	7	8	15	25	18	9	6

Construye el histograma y el polígono de frecuencias absolutas.