

ESTADISTICA

Para qué nos sirve?

La estadística va más allá de la simple recolección y descripción de datos: analiza e interpreta la información con el fin de llegar a conclusiones válidas, de proporcionar instrumentos para la toma de decisiones con base en información actualizada.



Casi toda persona está familiarizada con este tipo de frases:

- ◆ *Las estadísticas muestran que el costo de vida está subiendo.*
- ◆ *Los salarios aumentaron un promedio del 4% este año*

Estas frases y muchas otras que se nos presentan a diario son situaciones estadísticas.

✂ *Recorta, del periódico o de revistas, frases y/o gráficos estadísticas y pégalas en una hoja de tu cuaderno o al respaldo de esta hoja si imprimiste el taller.*

Definición



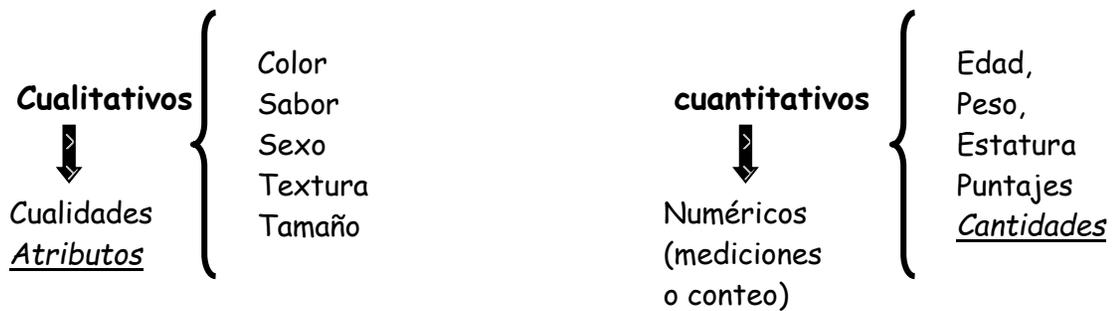
- ◆ La **Estadística** es la parte de la matemática que se encarga de recolectar, clasificar, describir e interpretar datos. Son muchas y muy amplias las áreas donde se encuentra terreno para su aplicación.

- ◆ La **Población** es el Conjunto de objetos bajo investigación
- ◆ La **muestra** es cualquier subconjunto de la población

- 🔗 **Ejemplo:** estamos interesados en un estudio sobre la contaminación ambiental de todas las ciudades colombianas de más de 100.000 habitantes. Determinemos la población y un ejemplo de muestra
- 🔗 La población es el conjunto de todas las ciudades de Colombia que tienen más de 100.000 habitantes, luego: $P = \{x/x \text{ es ciudad de Colombia con más de 100.000 habitantes}\}$
- 🔗 Una muestra puede ser: $M = \{\text{Cali, Medellín}\}$

Tipos de datos

Fundamentalmente, hay dos tipos de datos:



Una variable es un símbolo que representa indistintamente, a uno cualquiera de los elementos de un conjunto de datos

- ✎ Subraya las variables cuantitativas
Color de ojos, peso, número de hijos, estatura, programa de televisión favorito, código, nombre, pasatiempos, marca de automóvil, libros de la biblioteca

Las variables cuantitativas pueden ser continuas o discretas

Son continuas cuando toman cualquier valor en un intervalo real de valores, por ejemplo la estatura humana 1,3 m o 1,58 m

Una variable cuantitativa es discreta cuando solo puede tomar valores aislados por ejemplo el número de hijos en un matrimonio 0, 1, 2, 3 etc.

- ✎ Determina cual de las siguientes variables cuantitativas son **continuas** y cuales **discretas** subraya de color rojo las primeras y en azul las segundas

Peso	número de hermanos	estatura
edad	Tiempo de trabajo	Precio de un artículo
Tiempo de sueño	número de votantes	número de carros

🌀 Ejemplo:

Deseamos conocer algo sobre el color del cabello de cada uno de los estudiantes del colegio "José Joaquín Casas" Determinemos la población, la muestra, los datos y una variable

- ☞ La población: el conjunto de todos los estudiantes del colegio "José Joaquín Casas"
- ☞ Una muestra: los estudiantes de grado 9°
- ☞ Los datos: son los colores del cabello de todos los estudiantes del colegio, si en el colegio solo hay estudiantes de cabello negro y rubio entonces el conjunto de datos es: {negro, rubio} en este caso los datos son cualitativos y el atributo es el color del cabello
- ☞ Una variable aquí sería C , que representa el color negro o rubio

Ejercicio

✎ Resolver los siguientes ejercicios (al respaldo de la hoja o en tu cuaderno)

1. Una investigación está encaminada a determinar el porcentaje de fumadores en la ciudad de Chía ¿Cuál es la población? ¿Cuál es una muestra?
2. Haz una investigación para determinar la estatura de los estudiantes de tu curso ¿Cuál es la población? ¿Cuál es una muestra?
3. Estas investigando sobre el número de habitantes de todas las poblaciones colombianas ¿Cuál es la población? ¿Cuál es una muestra? ¿Qué tipo de datos se encuentran? ¿Cuál es la variable y que representa?
4. Se desea investigar el peso de cada uno de los niños que nacen en un determinado hospital en un día. ¿Cuál es la población? ¿Cuál es la muestra? ¿Cuál es el conjunto de datos? ¿Son los datos cuantitativos? ¿Cuál es la variable? ¿qué actividad planearías para obtener los datos de la investigación?
5. Juanito ha realizado una encuesta a sus compañeros y compañeras y observa lo que contestó una de ellas:

Sexo	Femenino <input checked="" type="checkbox"/>	Azul <input type="checkbox"/>	Rock <input type="checkbox"/>
	Masculino <input type="checkbox"/>	Color Rojo <input checked="" type="checkbox"/>	Pop <input type="checkbox"/>
		Favorito Negro <input type="checkbox"/>	Música Romántica <input checked="" type="checkbox"/>
		Verde <input type="checkbox"/>	favorita Vallenato <input type="checkbox"/>
Nombre: <u>Mercedes</u>		Color del cabello	Rubio <input checked="" type="checkbox"/> Negro <input type="checkbox"/>

✎ Quien contestó la encuesta? _____

✎ Podríamos decir lo que contexto la persona encuestada?

Disposición y clasificación de datos de frecuencia

Para presentar los datos recogidos de una muestra, se utiliza una tabla de distribución de frecuencias.

Consideremos una encuesta a 50 alumnos de una universidad, sobre el número de faltas de asistencia durante el primer semestre de este año

Las respuestas recogidas arroja el siguiente conjunto de datos numéricos:

3, 2, 4, 3, 1, 2, 3, 4, 3, 8, 3, 5, 6, 6, 5, 3, 4, 1, 2, 3, 2, 5, 1, 3, 3, 3, 2, 4, 1, 2, 2, 3, 3, 5, 5, 6, 3, 4, 4, 1, 2, 4, 3, 7, 7, 3, 7, 6, 5, 3

☞ Observa que 5 encuestados afirmaron tener una sola falla, ya que el número 1 se encuentra 5 veces en la lista de datos recogidos, al encontrar ocho veces el número dos (2) concluimos que 8 estudiantes fallaron 2 veces, y así sucesivamente, de esta forma organizamos la recolección de datos en una forma práctica y sencilla antes de construir la tabla de frecuencias

1 falta ⇒ 5 estudiante	5 faltas ⇒ 6 estudiantes
2 faltas ⇒ 8 estudiantes	6 faltas ⇒ 4 estudiantes
3 faltas ⇒ 16 estudiantes	7 faltas ⇒ 3 estudiantes
4 faltas ⇒ 7 estudiantes	8 faltas ⇒ 1 estudiante

☞ Una vez organizada la información iniciamos la tabla de distribución de frecuencias así:

Fallas de 50 estudiantes de una universidad	
Valores de la variable	Frecuencia absoluta
Número de fallas	Número de estudiantes
1	5
2	8
3	16
4	7
5	6
6	4
7	3
8	1
numero total de estudiantes = 50	

En la columna encabezada por *Valores de la variable* deben aparecer los distintos valores, bien sean cualitativos o cuantitativos como en este caso, se recomienda que estos datos recogidos aparezcan en orden ascendente, como se muestra.

La segunda columna muestra el número de veces que se da la variable por cada dato, en la muestra, llamada frecuencia absoluta (f)

Así en la columna de frecuencias aparece el número de estudiantes que falló el número de veces indicado a la izquierda

Algunas conclusiones:

- ☞ Bastantes alumnos fallaron 3 veces
- ☞ Solo un estudiante tiene 8 fallas
- ☞ La mitad de los estudiantes presentan menos de 4 fallas



A las conclusiones anteriores se les llama **Inferencias estadísticas**

Para obtener información mas detallada es necesario saber, que porcentaje de la muestra corresponde a cada valor de la variable, el cual llamaremos **frecuencia relativa (% f)** así:

x	f	%f
1	5	10
2	8	16
3	16	32
4	7	14
5	6	12
6	4	8
7	3	6
8	1	2
Totales	n = 50	100%

- X valores de la variable
- n es el tamaño de la muestra
- f la frecuencia absoluta correspondiente al valor x
- % f frecuencia relativa

tenemos que $\%f = \frac{f}{n} * 100\%$

la frecuencia relativa correspondiente a 3 fallas es: x = 3 f = 16 n = 50

$\%f = \frac{16}{50} * 100 = 32$ significa que 32% de los

alumnos fallaron 3 veces

La tabla de distribución de frecuencias queda completa si le agregamos dos columnas más de **frecuencias acumuladas (F)** **frecuencias relativas acumuladas (%F)**

En la columna de frecuencias acumuladas(F) se consignan las sumas de las frecuencias correspondientes a todos los valores de las variables menores o iguales que el valor indicado por la misma

x _i	f _i	%f _i	F _i
1	5	10	5
2	8	16	13
3	16	32	29
4	7	14	36
5	6	12	42
6	4	8	46
7	3	6	49
8	1	2	50
Totales	n = 50	100%	

- F frecuencia acumulada

$F_i = f_1 + f_2 + f_3 + \dots + f_i$

Por ejemplo para

$x_i = 5,$

$f_1 = 5 \quad f_2 = 8 \quad f_3 = 16 \quad f_4 = 7 \quad f_5 = 6$

entonces $F_5 = 5 + 8 + 16 + 7 + 6 = 42$

este 42 significa que 42 estudiantes tuvieron hasta 5 fallas durante el trimestre o que los 8 restantes presentaron más de 5 fallas

Y por ultimo agregamos la columna de frecuencia relativa acumulada (%F), donde se consigna el porcentaje de frecuencias acumuladas que corresponden al valor de la variable x_i así: ↓

Faltas de 100 estudiantes del colegio x				
x_i	f_i	$\%f_i$	F_i	$\%F_i$
1	5	10	5	10
2	8	16	13	26
3	16	32	29	58
4	7	14	36	72
5	6	12	42	84
6	4	8	46	92
7	3	6	49	98
8	1	2	50	100
	n = 50	100%		

▪ %F frecuencia relativa acumulada

$$\%F_i = \frac{F_i}{n} * 100\%$$

para $x_4=4$, $F_4 = 36$, $n= 50$

$$\%F_4 = \frac{36}{50} * 100 = 72$$

el 72% de los estudiantes presentaron entre 1 y 4 fallas

☞ Interpretemos algunos resultados que se muestran en la tabla final de frecuencias

- el 46 que aparece en la tabla , en la columna encabezada por F_i significa que 46 estudiantes de la muestra tienen fallas menores o iguales a 6
- el 58 que aparece en la tabla , en la columna encabezada por $\%F_i$ significa que el 58% de los estudiantes fallaron 3 o menos veces
- Los alumnos que presentaron 5 fallas, representan el 12% de la muestra

✎ Que representa el número 32 en la tabla de frecuencias? _____

✎ Que porcentaje de estudiantes obtuvo el mayor número de fallas? _____

✎ Cuantos estudiantes presentaron menos de 4 fallas? _____

✎ Que porcentaje de alumnos falló una sola vez? _____

Que otras inferencias estadísticas puedes anotar?

✎ _____

✎ _____

✎ _____

✎ El 32% de los alumnos registró 3 fallas

✎ _____

✎ Con la información registrada en la página 4, "encuesta a 20 personas" elabora cada una de las tablas completas de frecuencia para las variables: Edad, Estatura y Peso (trabaja en tu cuaderno o al respaldo de esta hoja)

RECUERDA:



Una tabla de distribución de frecuencias contiene las siguientes columnas:

- La primera muestra todos los **valores de la variable** (x_i)
- La segunda conocida como **frecuencia absoluta** (f_i) aparece el número de veces que se da la variable por cada dato,
- La tercera llamada **frecuencia relativa** ($\%f_i$), $\%f = \frac{f}{n} * 100\%$ es el porcentaje del total de la muestra que corresponde a cada valor de la variable donde n es el tamaño de la muestra
- La cuarta columna la **frecuencia acumulada** dada por $(F_i) = f_1 + f_2 + \dots + f_i$
- La quinta, **frecuencia relativa acumulada** correspondiente a: $(\%F_i) = \frac{F_i}{n} * 100\%$ indica el porcentaje de frecuencias acumuladas que corresponden al valor de la variable x_i

🌀 **Ejercicio:**

1. En cada uno de los siguientes ejercicios organiza los datos y preséntalos mediante un cuadro completo de frecuencias:

a. Al tomar las estaturas de 100 estudiantes de un colegio, encontramos los siguientes datos

1.50m	⇒ 1 alumno	1.61m	⇒ 40 alumnos
1.52m	⇒ 8 alumnos	1.70m	⇒ 8 alumnos
1.60m	⇒ 13 alumnos	1.45m	⇒ 1 alumnos
1.59m	⇒ 28 alumnos	1.71m	⇒ 1 alumnos

Organicemos la información en una tabla de frecuencias. Complétala

Estatura de 100 estudiantes del colegio x				
x_i	f_i	$\%f_i$	F_i	$\%F_i$
1.45	1		1	1
1.50	1		2	2
1.52		8	10	10
1.59		28	38	
1.60	13			
1.61		40		
1.70	8			
1.71		1	100	
	n =	100%		

- d. para investigar la preferencia que se tiene por cierta marca de automóviles se utilizo el teléfono y se obtuvieron los resultados que presenta la tabla

Marca de Autos	No de respuestas afirmativas
A	////////
B	//////////
C	////////
D	//////////
E	//////////

Preferencia De Carros				
x_i	f_i	$\%f_i$	F_i	$\%F_i$
A				
B				
	n=			

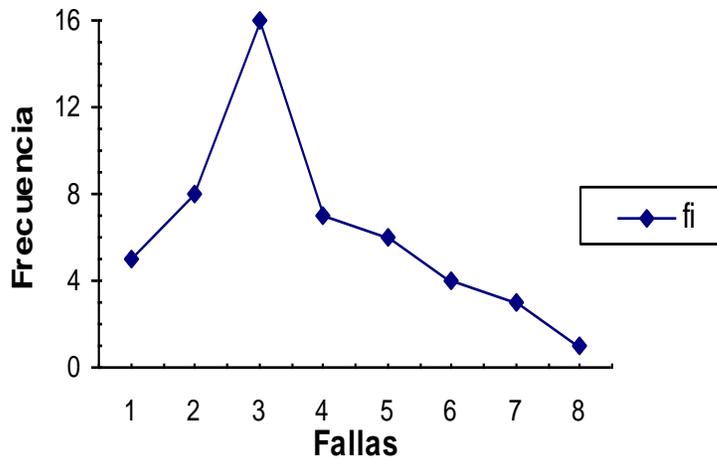
2. Analiza las tablas de frecuencia de los ejercicios anteriores (págs. 8, 9 y 10) y completa las siguientes inferencias estadísticas
- Cuántos estudiante miden 1.61m de estatura? _____
 - El porcentaje de estudiantes que mide menos de 1.60m de estatura es _____
 - Cuantos estudiantes miden menos de 1.59? _____
 - Cuantos paquetes de café pesaron menos de 500 gramos? _____
 - Que porcentaje de paquetes de café presentan el máximo peso? _____
 - Que porcentaje de estudiantes obtuvieron mas de 75 puntos en las olimpiadas? _____
 - Cuantos estudiantes perdieron el puntaje de las olimpiadas? _____
 - Si el mínimo puntaje para pasar las olimpiadas matemáticas es 70, entonces cuantos alumnos perdieron la prueba? _____
 - Que porcentaje de alumnos obtuvieron puntajes mayores a 80? _____
 - Según la información anterior cuantos pasaron la prueba? _____
 - Cual fue el puntaje mínimo de la muestra de estudiantes que participaron en las olimpiadas? _____
 - Que porcentaje de encuestados prefieren la marca de auto E? _____
 - Que marca de auto es la menos preferida? _____
 - Que marca de carro, según la información prefiere usted? _____

Representación Gráfica De Datos Estadísticos

Una vez elaboradas las tablas de frecuencia correspondientes a los datos de una investigación debemos hacer una representación de los mismos en un diagrama que nos permita visualizar rápidamente lo que ha ocurrido.

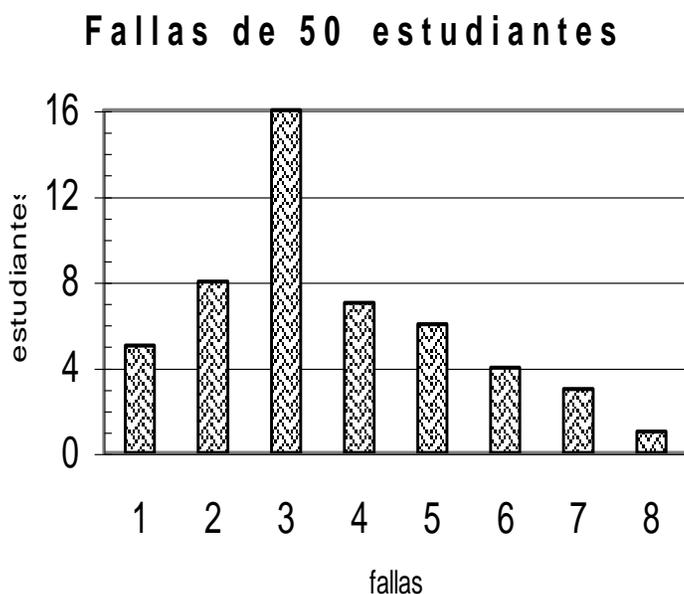
Existen varias gráficas estadísticas como:

➤ Polígonos de frecuencias



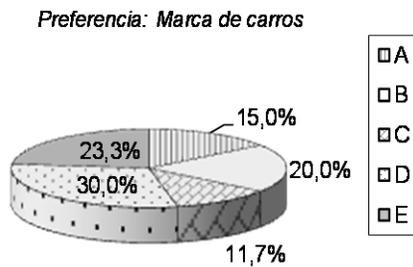
El gráfico elaborado corresponde a la tabla de frecuencias de la página 5. (Fallas de 50 estudiantes)
 Sobre el eje horizontal (X) marcamos el distinto número de fallas que cometieron los estudiantes de la muestra
 Sobre el eje vertical (Y) marcamos las distintas frecuencias, como la frecuencia mayor es 16, debemos dividir el eje Y de tal forma que el punto mas alto corresponda a dicha frecuencia

➤ Gráfico de barras



Los gráficos de barras consisten en trazar para cada valor muestral una barra vertical cuya altura sea igual a la frecuencia medida sobre una escala vertical. Observando el gráfico, cual fue el mayor número de fallas presentado? _____ y por cuantos alumnos? _____
 Cuantos alumnos fallaron 4 días? _____

➤ **Diagrama circular**



Llamada por su forma pastel o torta,

Analicemos el ejercicio d de la página 9 y así entenderemos la gráfica

Que Columna de la tabla de frecuencias está representada en esta torta? _____

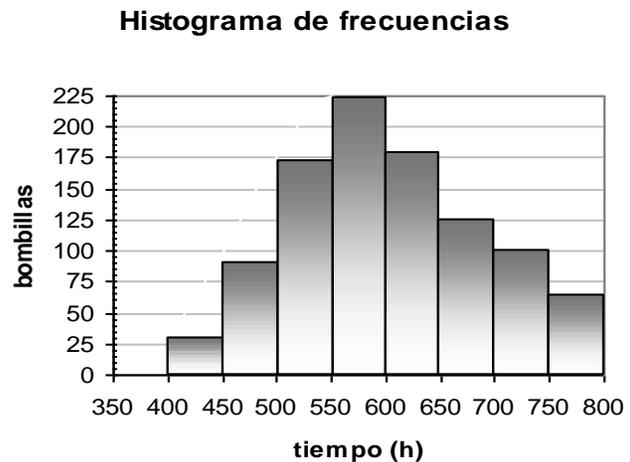
Que significa 23,3%? _____

Que carro prefieren los encuestados? _____

➤ **Histograma**

Para ilustrar esta gráfica tengamos en cuenta la siguiente tabla de frecuencias que indica la duración en horas, de 1000 bombillas

Duración de las bombillas	
Duración (horas)	Frecuencia
400-450	30
450-500	100
500-550	175
550-600	225
600-650	180
650-700	125
700-750	100
750-800	65



El histograma es utilizado para representar a la variable continua, requiere que la amplitud del intervalo sea constante y se compone de barras o rectángulos unidos, su ancho está dado por el intervalo de clase y su altura por la frecuencia.

Que conclusiones o inferencias puedes establecer de acuerdo al histograma?

Teniendo en cuenta el ejercicio de la página 4 "encuesta a 20 personas" elabora:

1. Polígono de frecuencias de la edad
2. Grafico de barras según la estatura
3. Diagrama circular mostrando color preferido.

Medidas de tendencia central

➤ **La Media \bar{X}** → Es la suma de todos los valores que toma la variable dividida por el número de observaciones.

Analicemos el siguiente ejemplo (ejercicio b de la página 8)

Peso de 50 paquetes de café		$\bar{X} = \frac{\sum X_i}{n} = \frac{\text{suma de todos los valores observados}}{\text{número de observaciones}}$ $\bar{X} = \frac{\sum x_i * f_i}{n} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{\text{número de observaciones}}$
x_i	f_i	
482	2	\sum significa suma $\bar{X} = \frac{x_i * f_i}{n} = \frac{482 * 2 + 487 * 10 + 499 * 13 + 500 * 5 + 501 * 17 + 508 * 3}{50}$
487	10	
499	13	$\bar{X} = \frac{964 + 4870 + 6487 + 2500 + 8517 + 1524}{50} = \frac{24862}{50} = 497.24$
500	5	
501	17	Entonces el peso promedio de los 50 paquetes es: $\bar{X} = \underline{497.24}$
508	3	
	n=50	

✎ Calcule la Media del ejercicio a de la página 8 (trabaje al respaldo de la hoja)

➤ **La Moda (Md) ó (Mo)** → Es el valor muestral de mayor frecuencia absoluta

estaturas		Observemos que la estatura que presenta mayor frecuencia es 1.55 m por lo tanto $Md = \underline{1.55}$ Cuál es la Moda de la tabla anterior (peso paquetes de café)? _____ por qué? _____
x_i (m)	f_i	
1.20	5	
1.45	3	
1.55	8	
1.60	6	

➤ **La Mediana (Me)** → es el dato tal que entre el y los menores que él, totalizan el 50% de los datos de la muestra, en otras palabras es el valor que se encuentra en el punto medio de todos los datos

Peso de 40 alumnos			Se halla el cociente $\frac{n}{2}$ Como son 40 datos entonces: $\frac{40}{2} = 20$ Y en la columna de frecuencias acumuladas se busca el primer valor mayor o igual a 20, la frecuencia acumulada con esta característica es 23, que corresponde al valor 56 kg por lo tanto $Me = 56$
x_i (kg)	f_i	F_i	
44	6	6	
50	7	13	
56	10	23	
57	12	35	
63	5	40	

✎ Averigua cual es la mediana del ejercicio c de la página 9 (Puntaje olimpiadas matemáticas del grado 9º)