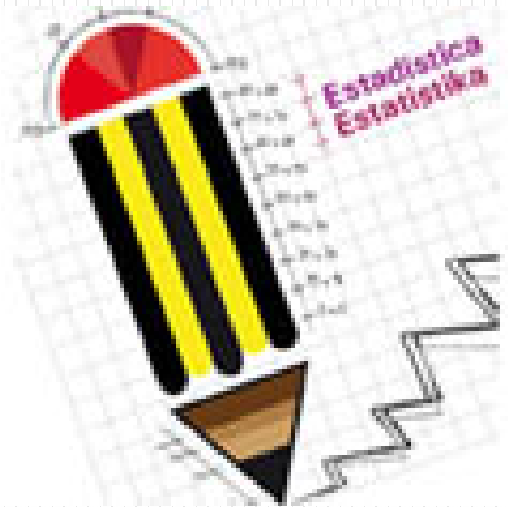


BREVE ANÁLISIS DE LA PRESENTACIÓN DE LA ESTADÍSTICA EN LIBROS DE TEXTO



III Jornadas de Enseñanza de las Matemáticas en Navarra

PLANTEAR SITUACIONES ESTADÍSTICAS CLARAS EN LAS QUE LA POBLACIÓN Y LA CARACTERÍSTICA EN ESTUDIO SE DEDUZCAN CON FACILIDAD

Tomás ha elaborado una tabla con las mascotas preferidas de sus 30 compañeros de clase. Calcula la frecuencia relativa de cada una.

**Característica de interés:
Mascota preferida**

	Frecuencia absoluta	Frecuencia relativa
Perro	12	$\frac{12}{30} =$
Gato	9	
Tortuga	3	
Hámster	6	
Total	30	

Población: compañeros de la clase de Tomás.

Característica de interés: Mascota preferida

1) Frecuencia absoluta y frecuencia relativa

Carlos, Irene y Andrés practican el lanzamiento de tiros libres.

¿Quién ha obtenido mejores resultados?



Para averiguarlo han colocado los resultados en una tabla de frecuencias:

	Frecuencia absoluta
Carlos	40
Irene	30
Andrés	70

La frecuencia absoluta indica el número de veces que se repite un dato.

Andrés es el que más canastas ha conseguido, pero también es el que más veces ha lanzado.

Para saber quién ha conseguido mejores resultados, hay que tener en cuenta el número total de intentos. Para ello calculamos la frecuencia relativa:

	Frecuencia absoluta	Número de intentos	Frecuencia relativa
Carlos	40	80	$\frac{40}{80} = 0,5$
Irene	30	40	$\frac{30}{40} = 0,75$
Andrés	70	100	$\frac{70}{100} = 0,70$

La frecuencia relativa se obtiene dividiendo la frecuencia absoluta entre el número total de observaciones.

La frecuencia relativa mayor es la de Irene: ella ha obtenido los mejores resultados.

Población:

¿los tres amigos?

Característica de interés:

¿canastas encestadas?

¿porcentaje de acierto?

Población:

¿Canastas encestadas?

Característica de interés:

¿amigos?

La frecuencia relativa es sobre el mismo total.

¡¡¡NO ES UN EJERCICIO DEL TEMA DE ESTADÍSTICA!!!

NOTACIÓN MUESTRAL FRENTE A NOTACIÓN POBLACIONAL

3. El número de hijos de los trabajadores de una fábrica es:

1 0 2 1 0	0 0 2 1 1	2 3 1 0 1	2 1 3 1 0	2 1 1 1 0
3 1 2 0 1	1 2 0 0 1	2 1 3 0 1	2 0 1 2 0	0 1 2 0 2

Población: TRABAJADORES DE LA FÁBRICA

Característica de interés: X: Número de hijos

Observamos a una parte

ESTUDIO MUESTRAL

Media

\bar{X}

$$\bar{X} = \frac{x_1 \cdot n_1 + x_2 \cdot n_2 + \dots + x_n \cdot n_n}{N}$$

Varianza (muestral): S^2

Observamos a toda la población

ESTUDIO CENSAL O POBLACIONAL

Media (poblacional)

μ

$$\sigma^2 = \frac{(x_1 - \bar{x})^2 \cdot n_1 + (x_2 - \bar{x})^2 \cdot n_2 + \dots + (x_n - \bar{x})^2 \cdot n_n}{N}$$

Varianza (poblacional) σ^2

CONSECUENCIAS: MALA COMPRENSIÓN DE ENUNCIAD

Cuando se explica inferencia en 2º Bachiller de Humanidades y C. Sociales NO se mezclar la notación

48 ●

PAU

Se quiere estimar la media de la nómina mensual que reciben los directivos de las compañías multinacionales que operan en Europa.

- a) Si la varianza de la nómina en la población es de 1000 €, ¿cuál es la varianza de la media muestral cuando el tamaño de la muestra es de 100? ¿Var[\bar{X}]?
- b) Si en las condiciones del apartado anterior la media muestral fuera de 4008 €, ¿se rechazaría con un nivel de confianza del 0,95 la hipótesis de que la nómina media es de 4000 €?

ESTUDIO MUESTRAL

Se toma una muestra de 100 directivos

\bar{X}

S^2

ESTUDIO CENSAL O POBLACIONAL

X: nómina directivos

$E[X] = \mu$

$Var[X] = \sigma^2$

CONSECUENCIAS: MALA DISTINCIÓN ENTRE VALORES FIJOS Y ALE

ESTUDIO MUESTRAL

Se toma una muestra de 100 directivos

\bar{X}

S^2

ESTUDIO CENSAL O POBLACIONAL

X: nómina directivos

$E[X] = \mu$

$\text{Var}[X] = \sigma^2$

- NO son valores fijos, cambian con cada muestra que se toma.

- Si la muestra se toma al azar: SON ESTADÍSTICOS

- Como la muestra se toma al azar, tienen

UNA DISTRIBUCIÓN DE PROBABILIDAD.

Son valores fijos (generalmente desconocidos) y se desean estimar

Así, el parámetro p que se desea estimar sigue una $N(0,28; 0,01)$.

Tipificando la variable, resulta que $Z = \frac{p - 0,28}{0,01}$ sigue una $N(0, 1)$.

USO DE LA NOTACIÓN ESTÁNDAR EN ESTADÍSTICA

Curso	Frecuencias absolutas	Frecuencias relativas	Frecuencias acumuladas absolutas	Frecuencias acumuladas relativas
1	10	33,33 %	10	33,3 %
2	8	26,67 %	18	60,0 %
3	6	20,00 %	24	80,0 %
4	6	20,00 %	30	100,0 %
Todos los cursos	30			

$$\bar{X} = \frac{x_1 \cdot n_1 + x_2 \cdot n_2 + \dots + x_n \cdot n_n}{N}$$

MÁS HABITUAL

$$\bar{X} = \frac{\sum x_i f_i}{N}$$

Valores	n_i	f_i	p_i	N_i
0	2	2:20=0,1	0,1·100=10 %	2
1	5	5:20=0,25	0,25·100=25 %	5+2=7
2	8	8:20=0,4	0,4·100=40 %	7+8=15
3	3	3:20=0,15	0,15·100=15 %	15+3=18
4	2	2:20=0,1	0,1·100=10 %	18+2=20
Total	20	1	100 %	

N.º de hijos x_i	Recuento	F. absoluta f_i	F. relativa h_i	F. absoluta acumulada, F_i	F. relativa acumulada, H_i
0	//// //	15	$\frac{15}{50}$	15	$\frac{15}{50}$
1	//// //	19	$\frac{19}{50}$	34	$\frac{34}{50}$
2	//// //	12	$\frac{12}{50}$	46	$\frac{46}{50}$
3	///	4	$\frac{4}{50}$	50	$\frac{50}{50}$
		50	1		

USO COHERENTE DEL SOFTWARE ESTADÍSTICO: EXCEL

En una base de datos que recopila los datos de una estadística sigue una norma bien establecida:

- CADA FILA ES UN INDIVIDUO (una observación)
- CADA COLUMNA ES UNA VARIABLE ESTADÍSTICA (una característica observada en el individuo)

Más coherente

Excel

	A	B	C	D	E
1	APELLIDOS	NOMBRE	EDAD	SALARIO	
2	Aliaga	Jose A.	34	34000	
3	Belmonte	Andrés	32	28000	
4	Casado	Ricardo	45	30000	
5	Martínez	Luis A.	21	20000	
6	Pérez	Alfonso	33	27000	
7	Zaragoza	Antonio	38	40000	
8					
9					

	A	B	C	D	E
1	59,4	60,6	55,3	62,4	61,9
2	66,8	72,1	80	90,4	65,8
3	58,9	68,6	74,5	71,5	92,8
4	60,6	75,4	85,7	89,6	74,3
5	61	85,4	63,5	66,8	84,3
6	94,4	71,2	65,7	62	64,3
7	79,8	81,3	65,9	73,4	78,6
8	81,4	90,2	48,5	100,1	56,9
9	63,7	56,2	85,4	93,6	102,2
10	74,8	62,5	86,1	90	59,8
11	63,7	115	83,9	64,8	91,3
12	73,8	64,2	102,5	96	85,6

Los datos que aparecen en el margen corresponden a la superficie (en metros cuadrados) de las viviendas de 60 familias. Vamos a ver cómo se pueden distribuir en 5 intervalos.

1. Tecllea los datos en las filas y columnas que aparecen en la tabla.

EL SOFTWARE ESTADÍSTICO AUTOMATIZA OPERACIONES, PERO NO REPRODUCE EL PROCEDIMIENTO QUE SE HACE EN EL PAPEL

Excel

	A	B	C	D	E
1	59,4	60,6	55,3	62,4	61,9
2	66,8	72,1	80	90,4	65,8
3	58,9	68,6	74,5	71,5	92,8
4	60,6	75,4	85,7	89,6	74,3
5	61	85,4	63,5	66,8	84,3
6	94,4	71,2	65,7	62	64,3
7	79,8	81,3	65,9	73,4	78,6
8	81,4	90,2	48,5	100,1	56,9
9	63,7	56,2	85,4	93,6	102,2
10	74,8	62,5	86,1	90	59,8
11	63,7	115	83,9	64,8	91,3
12	73,8	64,2	102,5	96	85,6

Tabla de frecuencias de datos agrupados en intervalos

Los datos que aparecen en el margen corresponden a la superficie (en metros cuadrados) de las viviendas de 40 familias, vamos a ver cómo se pueden distribuir en 5 intervalos.

- Teclas los datos en las filas y columnas que aparecen en la tabla.
- Combina las celdas F1 y G1 y escribe DTD MENOR=. Repite el proceso en las filas siguientes y escribe los textos que aparecen en la tabla del margen.

3. Anota las siguientes fórmulas en las celdas que se indican:

H1 → =MIN(A1:B2) I7 → =H1

H2 → =MAX(A1:B2) I8 → =I7 + \$I\$5

H3 → =I2 - I1

I4 → =5

H5 → =H3/I4

Haz clic en F9 y lleva el cursor a la esquina inferior derecha de la celda. Cuando aparezca una cruz, arrastra el ratón hasta la celda F12. De este modo, habrá copiado la fórmula contenida en F9 en las celdas F9, F10, F11 y F12. Obsérvate que has obtenido los extremos de los 5 intervalos.

- Combina las celdas A14 y B14 y escribe INTERVALOS. En la misma fila y en columna sucesivas teclea sí, no, sí, no. Teclea los extremos de los intervalos en las celdas de los columnas A y B de las filas 15, 16, 17, 18 y 19.

5. En la celda C15 escribe la fórmula =A15-B15/2. Cópiala después en las celdas C16 a C19.

6. En la celda D15 teclea la siguiente fórmula: =CONTAR.SI(A\$1:E\$12;">=48,5")-CONTAR.SI(A\$1:E\$12;">=61,8")

Con ella se calcula cuántas viviendas son mayores o iguales que 48,5 y cuántas son mayores o iguales que 61,8; la diferencia de ambos resultados será el número de datos del intervalo [48,5; 61,8]; es decir, su frecuencia absoluta.

Copia la fórmula de D15 en las celdas D16, D17, D18 y D19. Después, haz clic sobre D16 y sustituye 48,5 y 61,8 por 61,8 y 75,1, respectivamente. Cambia también los valores de las fórmulas de D17, D18 y D19. Selecciona el rango de celdas D15-D20 y haz clic en el botón de Autosuma (Σ).

7. Teclea estas fórmulas en las celdas indicadas:

F19 → =D15/D20

F18 → =REDONDEAR(F19*100;0)

Copia estas fórmulas hasta la E19 y la F19, respectivamente y calcula Autosuma.

8. Escribe las siguientes fórmulas en las celdas correspondientes:

G15 → =D15

G16 → =G15+D16

Copia la fórmula de G16 en G17, G18 y G19.

14	INTERVALOS	ci	ni	f	h	hi
15	48,5; 61,8	55,15	10	0,166666667	16,7	10
16	61,8; 75,1	68,45	24	0,4	40	34
17	75,1; 88,4	81,75	13	0,216666667	21,7	47
18	88,4; 101,7	95,05	10	0,166666667	16,7	57
19	101,7; 115	108,35	3	0,05	5	60
20	TOTAL		60	1	100,1	

6. En la celda D15 teclea la siguiente fórmula:

$$=CONTAR.SI(A\$1:E\$12;">=48,5")-CONTAR.SI(A\$1:E\$12;">=61,8")$$

Con ella se calcula cuántas viviendas son mayores o iguales que 48,5 y cuántas son mayores o iguales que 61,8; la diferencia de ambos resultados será el número de datos del intervalo [48,5; 61,8]; es decir, su frecuencia absoluta.

Copia la fórmula de D15 en las celdas D16, D17, D18 y D19. Después, haz clic sobre D16 y sustituye 48,5 y 61,8 por 61,8 y 75,1, respectivamente. Cambia también los valores de las fórmulas de D17, D18 y D19. Selecciona el rango de celdas D15-D20 y haz clic en el botón de Autosuma Σ .

EL SOFTWARE ESTADÍSTICO AUTOMATIZA OPERACIONES, PERO NO REPRODUCE EL PROCEDIMIENTO QUE SE HACE EN EL PAPEL

Herramienta: análisis de datos (instalada en cualquier Excel)

Excel

	A	B	C	D	E
1	59,4	60,6	55,3	62,4	61,9
2	66,8	72,1	80	90,4	65,8
3	58,9	68,6	74,5	71,5	92,8
4	60,6	75,4	85,7	89,6	74,3
5	61	85,4	63,5	66,8	84,3
6	94,4	71,2	65,7	62	64,3
7	79,8	81,3	65,9	73,4	78,6
8	81,4	90,2	48,5	100,1	56,9
9	63,7	56,2	85,4	93,6	102,2
10	74,8	62,5	86,1	90	59,8
11	63,7	115	83,9	64,8	91,3
12	73,8	64,2	102,5	96	85,6



	A
1	59,4
2	66,8
3	58,9
4	60,6
5	61
6	94,4
7	79,8
8	81,4
9	63,7
10	74,8
11	63,7
12	73,8
13	60,6
14	72,1
15	68,6
16	75,4
17	85,4



61,8
75,1
88,4
101,7
115



Clase	Frecuencia	% acumulado
61,8	8	22,22%
75,1	14	61,11%
88,4	10	88,89%
101,7	2	94,44%
115	2	100,00%
y mayor...	0	100,00%

Histograma

Entrada

Rango de entrada:

Rango de clases:

Rótulos

Opciones de salida

Rango de salida:

En una hoja nueva:

En un libro nuevo

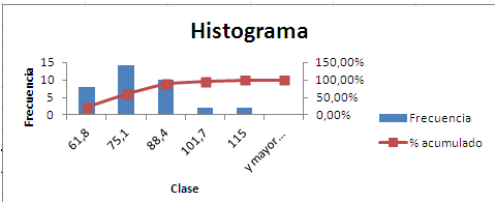
Pareto (Histograma ordenado)

Porcentaje acumulado

Crear gráfico

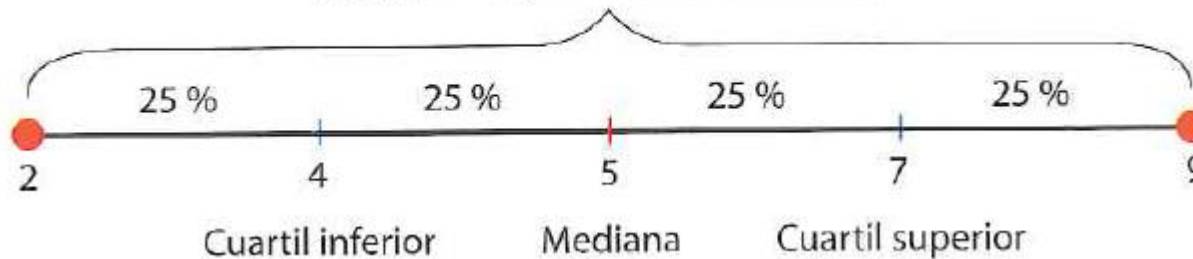
Aceptar Cancelar Ayuda

Columna1	
Media	71,253125
Error típico	2,38216654
Mediana	67,7
Moda	60,6
Desviación e	13,4755689
Varianza de l	181,590958
Curtosis	2,24246142
Coefficiente	1,17009622
Rango	66,5
Mínimo	48,5
Máximo	115
Suma	7780 1



LAS MEDIDAS DE POSICIÓN Y SU REPRESENTACIÓN GRÁFICA

Distribución porcentual de los datos



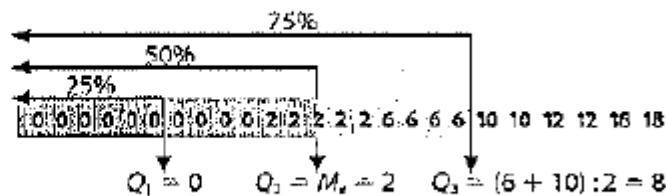
Nota	f_i	f_{hi}
3	3	3
4	4	7
5	6	13
6	6	19
7	4	23
8	2	25
Suma	25	

Cuartil inferior →
Mediana →
Cuartil superior →

EQUIDISTRIBUCIÓN ≠ EQUIDISTANCIA

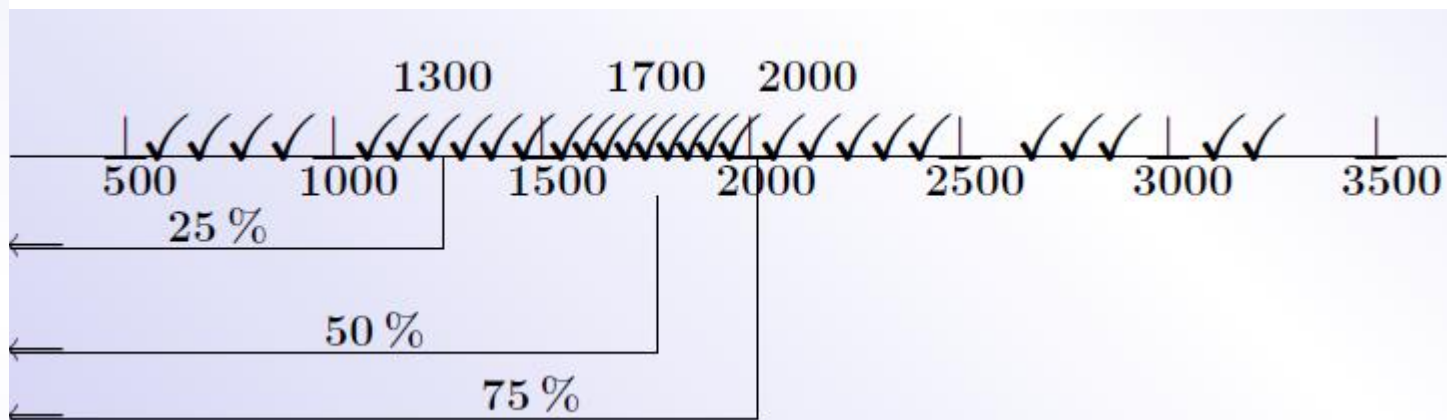
EJEMPLO

Esta lista muestra el número de horas semanales que cada uno de los 25 alumnos de un grupo ha faltado a clase:



Salario de 29 trabajadores

600	700	800	900	1100
1200	1250	1300	1350	1400
1575	1600	1650	1700	1725
1750	1800	1850	1950	2100
2200	2300	2400	2450	2700
2800	2900	3100	3200	



¿Y CÓMO CALCULAR LAS MEDIDAS DE POSICIÓN EN DATOS AGRUPADOS?

Fórmulas

Marca de clase

Software

$$M_e = L_i + \frac{\frac{n}{2} - f_{AMe-1}}{f_{Me}} \cdot c$$

Asume que los datos están uniformemente distribuidos en la clase

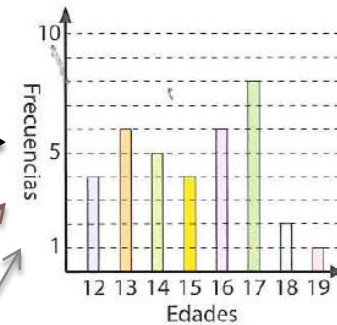
Velocidad (km/h)	x_i	N.º de vehículos f_i	F. absoluta acumulada F_i
[90, 100]	95	16	16
[100, 110]	105	15	31 < 50,5
[110, 120]	115	35	66 > 50,5
[120, 130]	125	25	91
[130, 140]	135	10	101
		101	

Utiliza la definición, pues Siempre tenemos un número n de datos

Asume que en la clase hay tantos datos a la izquierda como a la derecha de la marca de clase

¿Qué SON los diagrama de barras?

Un **diagrama de barras** es la representación en unos ejes cartesianos de las frecuencias de los valores que toma un carácter estadístico.



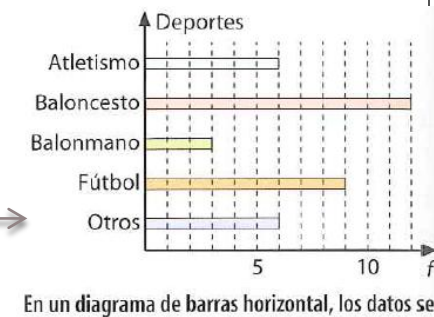
No requieren un eje cartesiano

El **diagrama de barras** es el gráfico que se utiliza para representar mediante barras datos de caracteres cuantitativos no agrupados en intervalos.



No son exclusivos de caracteres cuantitativos

Un **diagrama de barras** es un gráfico estadístico que asocia a cada valor de la variable una barra, vertical u horizontal, cuya longitud es igual, o proporcional, a su frecuencia.



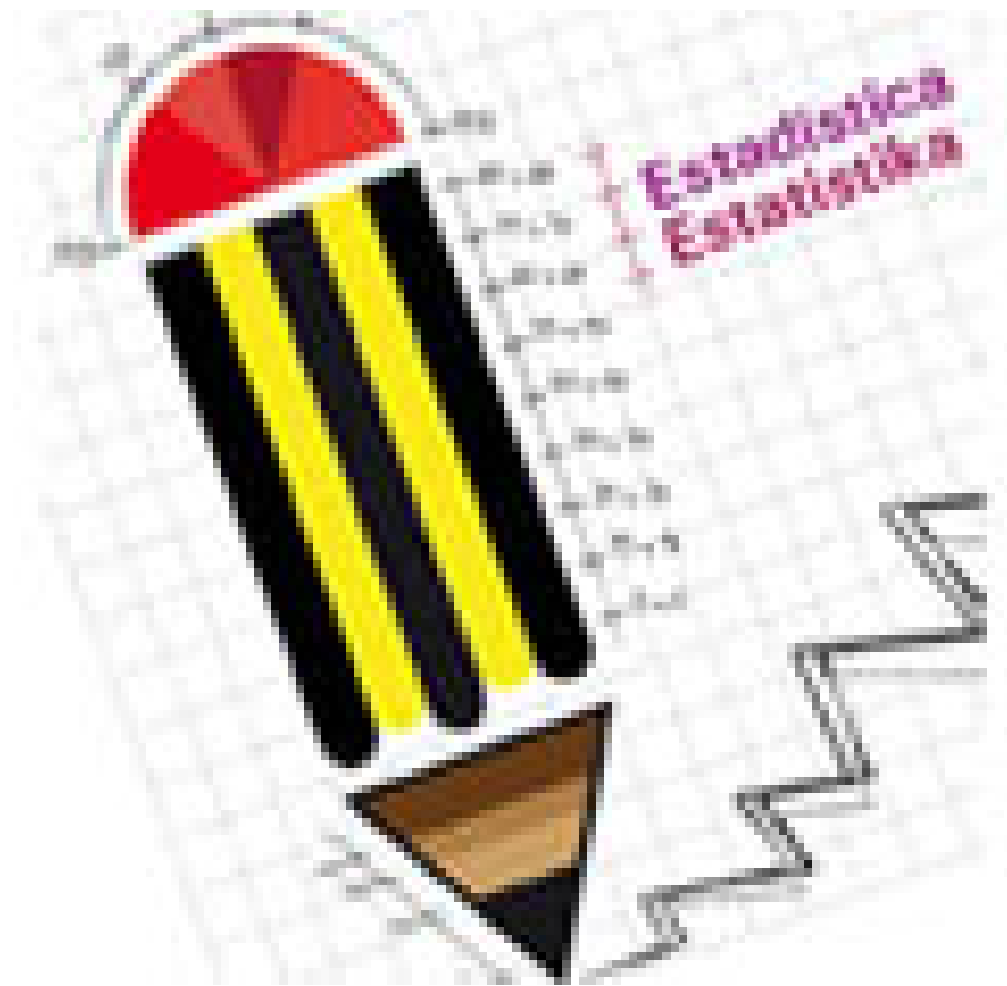
En un diagrama de barras horizontal, los datos se

¿Y LA PROBABILIDAD?

- Saltarse la probabilidad es también saltarse la notación y, también, las operaciones de conjuntos.
- La probabilidad no es un porcentaje, es una medida entre 0 y 1

$$P(A) = \frac{6}{36} = \frac{1}{6} \approx 0,1667 = 16,67\%$$

- El experimento aleatorio más importante de la estadística es “extraer una muestra de una población”, a todas las posibles ocurrencias de este experimento aleatorio se le llame “ESPACIO MUESTRAL” .
- Curiosamente, este es el nombre que se le da a todas las posibles ocurrencias de cualquier experimento aleatorio, pues esencialmente, cada ocurrencia del azar es un muestreo del fenómeno en cuestión.



FIN