



**Nombre de alumno:**

**Guadalupe Nájera  
López**

**Nombre del profesor:**

**Andrés Alejandro reyes  
molina**

**Nombre del trabajo:**

**Súper nota**

**Materia:**

**Estadística descriptiva**

**Grado:**

**3er cuatrimestre**

Comitán de Domínguez Chiapas a 08 de junio de 2022

## ORGANIZACIÓN Y REPRESENTACIÓN GRÁFICA DE LOS DATOS

Una distribución de frecuencias o tabla de frecuencias es una ordenación en forma de tabla de los datos estadísticos, asignando a cada dato su frecuencia correspondiente. A recogida de datos asociada a la realización de un estudio suele representar la obtención de un conjunto más o menos numeroso de datos, ahora bien, la interpretación de los mismos a simple vista suele resultar poco inteligible en la mayoría de los casos

### -La distribución de Frecuencias.

La distribución de frecuencias constituye una de las formas más intuitiva de organizar los datos de una variable: se basa en el conteo del número de entidades (casos, sujetos) que tienen cada uno de los valores con que la variable se ha manifestado (modalidades). Es una técnica estadística básica pero, a la vez, muy informativa y relevante en la práctica del análisis de datos

- El número de veces que aparece una determinada modalidad de una variable (X) es lo que se conoce como la frecuencia absoluta ( $n_i$ ) de esa modalidad o valor.

- Derivadas de las frecuencias absolutas se pueden obtener las frecuencias relativas o proporciones ( $p_i$ ):

$$P_i = n_i / n$$

**Ejemplo de Distribución de Frecuencias**

Tabla 2. Clasificación de un grupo de Estudiantes de la Facultad de Humanidades y Educación, según su opinión con respecto a la calidad del servicio del comedor universitario. Universidad de los Andes, Mérida, Venezuela. Semestre A-2002.

CALIDAD	Frecuencia	Porcentaje	Porcentaje acumulado
Mala	4	14,3	14,3
Regular	5	17,9	32,1
Buena	11	39,3	71,4
Excelente	9	29,8	100,0
Total	29	100,0	

Fuente: Desconocida.

**Distribución de frecuencias**

Intervalo de clase	Conteo	f
95-99		4
90-94		5
85-89		5
80-84		5
75-79		5
70-74		5
65-69		5
60-64		4
55-59		4
50-54		3
45-49		2
		1
		70

### Tipos de Frecuencia.

#### Frecuencia absoluta

La frecuencia absoluta es el número de veces que aparece un determinado valor en un estudio estadístico. Se representa por  $f_i$ .

La suma de las frecuencias absolutas es igual al número total de datos, que se representa por  $N$ .

$$f_1 + f_2 + f_3 + \dots + f_n = N$$

## Frecuencia relativa

La frecuencia relativa es el cociente entre la frecuencia absoluta de un determinado valor y el número total de datos. Se puede expresar en tantos por ciento y se representa por  $n_i$ .

$$n_i = \frac{f_i}{N}$$

## Frecuencia acumulada

La frecuencia acumulada es la suma de las frecuencias absolutas de todos los valores inferiores o iguales al valor considerado. Se representa por  $F_i$ .

## Frecuencia relativa acumulada

La frecuencia relativa acumulada es el cociente entre la frecuencia acumulada de un determinado valor y el número total de datos. Se puede expresar en tantos por ciento.

Intervalos	Marca de clase $x$	Frecuencia absoluta $f$	Frecuencia acumulada $F$
[0 - 4)	2	3	3
[4 - 8)	6	5	8
[8 - 12)	10	6	14
[12 - 16)	14	4	18
[16 - 20)	18	3	21
Total		21	

ESTADÍSTICA DESCRIPTIVA			
Nota	Frecuencia Absoluta	Frecuencia Relativa	Frecuencia Relativa Porcentual (%)
2,8	1	0,041	4,166
3,2	4	0,166	16,666
3,9	3	0,125	12,500
4,2	5	0,208	20,833
5,0	4	0,166	16,666
5,6	3	0,125	12,500
6,0	4	0,166	16,666

## Marca de clase:

La marca de clase es el punto medio de cada intervalo y es el valor que representa a todo el intervalo para el cálculo de algunos parámetros.

Para calcular varianzas y desviación típica, necesitamos añadir una nueva columna a la tabla:

Intervalo	Marca de clase $x$	Frecuencia absoluta $f$	$x \cdot f$	$x^2 \cdot f$
[340,5 < x < 350,5]	349	2	298	(2497,2 = 44.402)
[350,5 < x < 360,5]	354	1	154	23.716
[360,5 < x < 365,5]	359	4	636	101.124
[365,5 < x < 370,5]	364	12	1968	322.752
[370,5 < x < 375,5]	369	10	1690	285.610
[375,5 < x < 380,5]	374	1	174	30.276

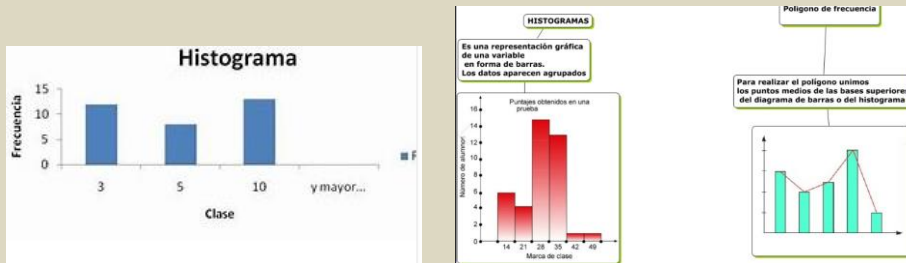
Ingreso Familiar (miles)	$n_i$	$y_i$	$y_i \times n_i$	$(y_i - \bar{y})$	$(y_i - \bar{y})^2$	$(y_i - \bar{y})^2 \times n_i$	$A_i$
80 - 99	8	90	720	-49,13	2.413,76	19.310,08	20
100 - 119	10	110	1100	-29,13	848,56	8.485,60	20
120 - 159	11	140	1540	0,87	0,76	8,36	40
160 - 199	6	180	1080	40,87	1.670,36	10.022,16	40
200 - 249	5	225	1125	85,87	7.373,66	36.868,30	50
Total	40		5565			74.694,50	

## La representación gráfica de una distribución de frecuencias

En los análisis estadísticos, es frecuente utilizar representaciones visuales complementarias de las tablas que resumen los datos de estudio. Con estas representaciones, adaptadas en cada caso a la finalidad informativa que se persigue, se transmiten los resultados de los análisis de forma rápida, directa y comprensible para un conjunto amplio de personas.

## Diagramas de barras e histogramas

Los diagramas de barras se usan para representar gráficamente series estadísticas de valores en un sistema de ejes cartesianos, de manera que en las abscisas se indica el valor de la variable estadística y en las ordenadas se señala su frecuencia absoluta.



## Propiedades de la Distribución de Frecuencias

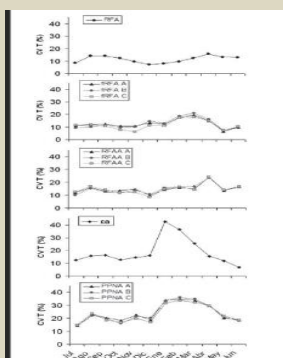
### Tendencia central:

Es un valor de la variable que se encuentra hacia el centro de la distribución de frecuencias. A este valor se le denomina promedio y es un valor que sintetiza a todos los valores de la distribución.



### Variabilidad:

Es un índice o medida que resume el grado de concentración de los valores de una distribución en torno a un promedio. Si los valores están muy cercanos al promedio se habla de una distribución homogénea; si, por el contrario, los valores se alejan del promedio.



## Estadísticos de Posición Grupal

Para una variable discreta, se define el percentil de orden  $p$ , como la observación,  $x_p$ , que deja por debajo de sí el  $p\%$  de la muestra. Esta definición es semejante a la de la mediana, pues como consecuencia de la definición, es evidente que:

$$Med = P_{50}.$$

Primer Cuartil:  $Q_1 = P_{25}.$

Segundo Cuartil:  $Q_2 = P_{50}.$  Es equivalente a la Mediana.

Tercer Cuartil:  $Q_3 = P_{75}.$



## Variables Categóricas: La Moda

Artículo principal: Moda (estadística)

La moda es el dato más repetido, el valor de la variable con mayor frecuencia absoluta. En cierto sentido se corresponde su definición matemática con la locución "estar de moda", esto es, ser lo que más se lleva.

Sus principales propiedades son:

- Cálculo sencillo.
- Interpretación muy clara.

$$Mo = L_{i-1} + a \cdot \frac{D_1}{D_1 + D_2}; \text{ siendo:}$$

$L_{i-1}$  = Límite inferior del intervalo modal  
 $a$  = Amplitud de los intervalos  
 $D_1$  = Diferencia de la frecuencia absoluta entre Intervalo modal y el anterior  
 $D_2$  = Diferencia de la frecuencia absoluta entre Intervalo modal y la siguiente



## VARIABLES ORDINALES: LA MEDIANA, EL MÍNIMO, Y EL MÁXIMO LOS CUANTILES.

Artículo principal: Mediana (estadística)

La mediana es un valor de la variable que deja por debajo de sí a la mitad de los datos, una vez que estos están ordenados de menor a mayor.<sup>17</sup> Por ejemplo, la mediana del número de hijos de un conjunto de trece familias, cuyos respectivos hijos son: 3, 4, 2, 3, 2, 1, 1, 2, 1, 1, 2, 1 y 1, es 2, puesto que, una vez ordenados los datos: 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 3, 3, 4, el que ocupa la posición central es 2:

$$Me = L_{i-1} + \frac{\frac{N}{2} - F_{i-1}}{f_i} \cdot a$$

$L_{i-1}$  = Límite inferior del intervalo mediana  
 $a$  = Amplitud del intervalo mediana  
 $F_{i-1}$  = Frecuencia acumulada anterior al intervalo  
 $f_i$  = Frecuencia absoluta del intervalo mediana  
 $N$  = Total de datos

**Ejemplo:**

Calcular la mediana de una distribución estadística que viene dada por la siguiente tabla:

	$f_i$	$F_i$
[66, 82)	5	5
[82, 94)	18	23
[94, 99)	42	65
[99, 121)	27	92
[121, 130)	8	100
	100	

$100/2 = 50$   
 Clase de la mediana: [66, 99)  
 $Me = 66 + \frac{50 - 23}{42} \cdot 3 = 67,93$

## VARIABLES CUANTITATIVAS:

La media y sus alternativas robustas.

La media. Es una medida de dispersión que tiene, por su propia definición, las mismas propiedades que la mediana. Por ejemplo, no se ve afectada por valores extremos o atípicos.

Media y varianza de una variable aleatoria continua

Variable estadística	Variable aleatoria discreta	Variable aleatoria continua
$\bar{x} = \sum_{i=1}^n x_i \cdot f_i$	$\mu = \sum_{i=1}^n x_i \cdot p(x_i)$	$\mu = \int_{-\infty}^{\infty} x \cdot f(x) \cdot dx$
$\sigma^2 = \frac{1}{n} \sum_{i=1}^n f_i (x_i - \bar{x})^2$	$\sigma^2 = \sum_{i=1}^n (x_i - \mu)^2 \cdot p(x_i)$	$\sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 \cdot f(x) \cdot dx$

$$\mu = E(X) = x_1 \cdot f(x_1) + x_2 \cdot f(x_2) + x_3 \cdot f(x_3) + x_4 \cdot f(x_4)$$

$$\mu = E(X) = 0 \cdot 0,60 + 1 \cdot 0,25 + 2 \cdot 0,10 + 3 \cdot 0,05$$

$$\mu = E(X) = 0 + 0,25 + 0,20 + 0,15$$

$$\mu = E(X) = \mathbf{0,60}$$

## Coeficiente de variación:

Se define como, donde  $\sigma$  es la desviación típica y  $\mu$  es la media aritmética. Se interpreta como el número de veces que la media está contenida en la desviación típica. Suele darse su valor en tanto por ciento, multiplicando el resultado anterior por 100. De este modo se obtiene un porcentaje de la variabilidad.

**COEFICIENTE DE VARIACIÓN**

- Es una medida de variabilidad relativa (sin unidades de medida) cuyo uso es para cuantificar en términos porcentuales la variabilidad de las unidades experimentales frente a la aplicación de un determinado tratamiento.
- En experimentación no controlada (condiciones de campo) se considera que un coeficiente de variabilidad mayor a 35% es elevado por lo que se debe tener especial cuidado en las interpretaciones y conclusiones.
- La expresión estimada del coeficiente de variabilidad es:

$$CV = \frac{\sqrt{CME}}{\bar{x}} \times 100$$

COEFICIENTE DE VARIACIÓN		CV
El coeficiente de variación indica que tan grande es la desviación estándar en relación al promedio.		
Como se Calcula	Ejemplo	
$CV = \frac{S}{\bar{x}} 100 \%$	$CV = \frac{9,13}{443,39} 100\% = 2,06 \%$	
Interpretación		
La distribución promedio de los precios de un apartamento de una habitación presenta menor variación o es menos heterogénea.		

- Variables ordinales: el rango y el rango intercuartil

Parámetros de dispersión.

Son datos que informan de la concentración o dispersión de los datos respecto de los parámetros de centralización.

Por ejemplo, vamos a suponer que hemos realizado el mismo examen en dos grupos distintos. En uno, todos los alumnos han sacado la misma nota, un 5; en otro, la mitad de los alumnos ha sacado un 0 y la otra mitad un 10.

**-Rango-**

Es la diferencia entre el dato mayor y el dato menor de un conjunto de datos numéricos.

Edad de los profesores de un colegio		
23	52	50
36	38	23
45	28	45
28	43	28
28	32	36

Rango= Dato mayor - Dato menor  
Rango= 52 - 23 = 29

**Rango**

- El **RANGO** es la diferencia entre el valor máximo y el valor mínimo de una serie de números o valores.
- **Ejemplo:**
- En {4, 6, 9, 3, 7} el menor valor es 3, y el mayor es 9, entonces el rango es 9-3 igual a 6.