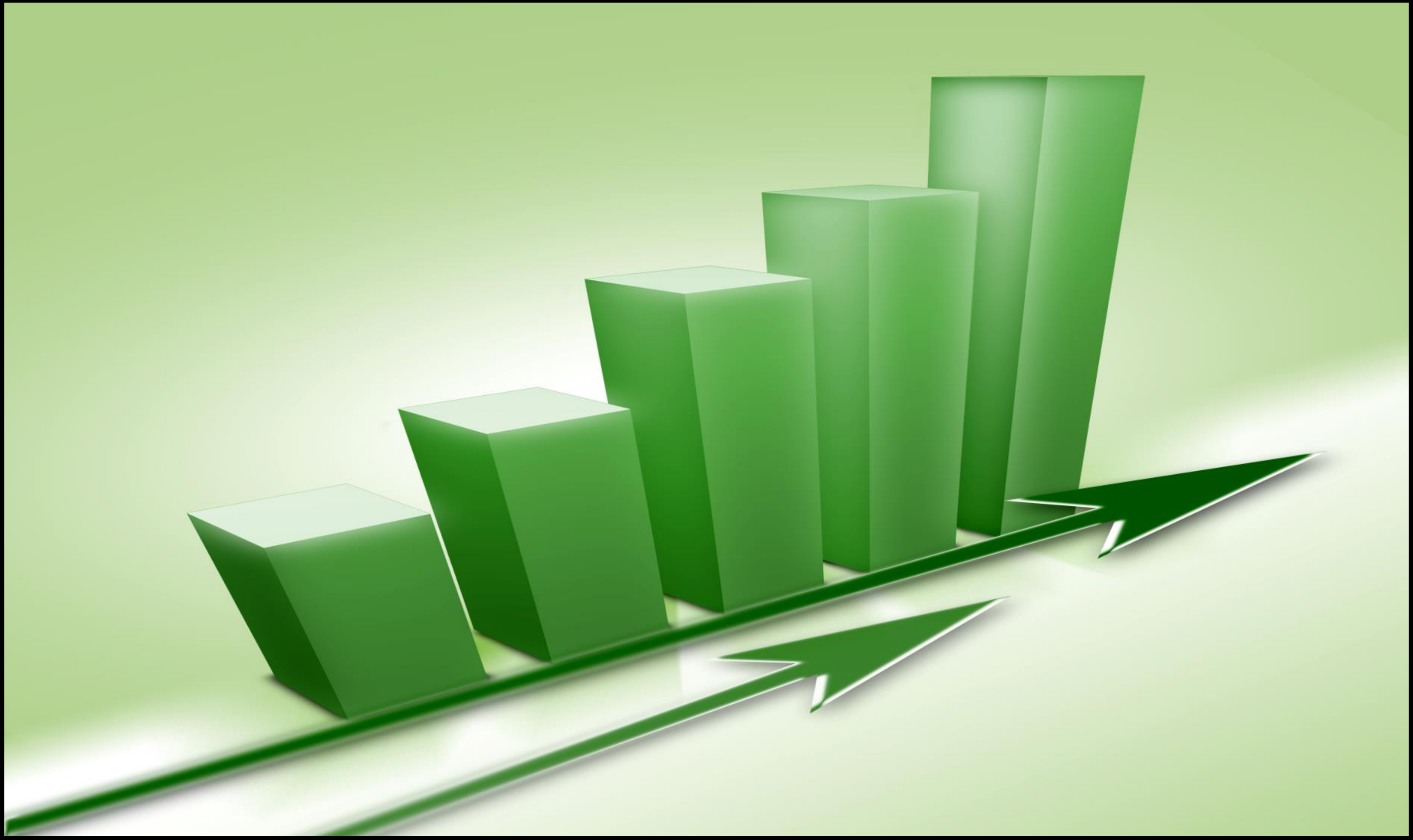
MODULO INSTRUCCIONAL PARA ESTADISTICA DESCRIPTIVA



ELABORADO POR PORFESOR: VICTOR ANTONIO VEGAS R

INDICE DE CONTENIDO

ESTADÍSTICA: La estadística es una metodología científica, la cual tiene por objeto la obtención, agrupación, clasificación, presentación y análisis de datos de forma sistemática, a fin de poder interpretarlos a través de la aplicación de métodos que faciliten la toma de decisiones acertadas, en función de los fenómenos que se estudian.

CLASIFICACION DE LA ESTADÍSTICA DESDE EL PUNTO DE VISTA DE LOS FINES Y LOS NIVELES DE LAS INVESTIGACIONES:

Estadística Descriptiva: Se centra en la recolección, clasificación, ordenación, tabulación y graficación de los datos obtenidos a través de los procesos de medición. (PERMITE DESCRIBIR NO INFERIR)

Estadística Inferencial: En esta la atención se centra en la obtención de conclusiones o inferencias válidas para una población o universo, a partir de los resultados obtenidos en el estudio de muestras representativas de dicha población o universo, y va desde los modelos paramétricos hasta los no paramétricos o contrastes de distribución libre.

POBLACIÓN: Es un conjunto finito o infinito de personas u objetos que presentan características comunes, en la cual se pueden someter a estudio o medición una o más variables.

MUESTRA: Se entiende por muestra desde el punto de vista estadístico, a un subconjunto o parte representativa de la población en la cual se pueden someter a estudio o medición una o más variables, a fin de investigar las propiedades de la población de la cual procede.

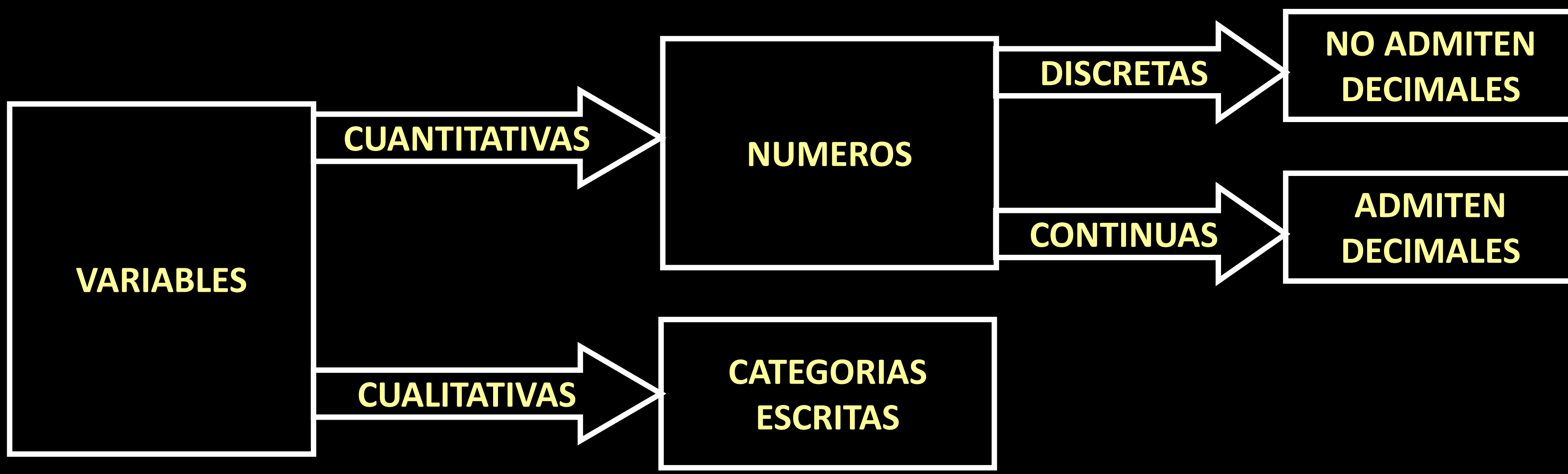
POBLACION
DE
ESTUDIANTES
DE LA UNEG
CIUDAD BOLIVAR

MUESTRA

ESTUDIANTES
DE LA UNEG
CIUDAD BOLIVAR
CARRERA
DE ADMINISTRACION

VARIABLES. Son los elementos de observación, que varían en alguna medida entre ellos clasifican en Cuantitativas y Cualitativas. Las cuantitativas son aquellas que se expresan en números, y las cualitativas o atributos, se expresan en categorías escritas.

Las variables cuantitativas, a su vez se clasifican en discretas y continuas, las discretas son aquellas que desde el punto de vista lógico , no admiten decimales, tales como: el número de personas en esta aula, la cantidad de vehículos que pasan por un sector, etc.



RAZONES PARA ESTUDIAR ESTADÍSTICA

Dentro de la amplia gama de aplicaciones de los modelos y métodos estadísticos tenemos :

-El aprendizaje de procedimiento y métodos que permiten el procesamiento adecuado de información.

-Cuantificar y evaluar la importancia de los resultados estadísticos obtenidos a la luz del proceso de investigación.

- Facilidad en la toma de decisiones.

- Comprender mejor los fenómenos investigados.

Entre otras cosas la Estadística se ocupa del manejo de la información que pueda ser cuantificada. Implica esto la descripción de conjuntos de datos y la inferencia a partir de la información recolectada de un fenómeno de interés. La función principal de la estadística abarca:

Resumir, Simplificar, Comparar, Relacionar, Proyectar.

DENTRO DE LAS ACTIVIDADES QUE SE REALIZAN TENEMOS:

1.- Delimitar con precisión la población y muestra de referencia o el conjunto de datos en estudio, las unidades que deben ser observadas, las características o variables que serán medidas u observadas.

2.- Estrategias de Observación: Muestra, Muestreo, Diseño del estudio.

3.-Recolección, Registro y procesamiento de la información.

4.- Construcción de Matrices, Tablas y Gráficos.

5.- Análisis Estadístico

6.- Interpretación de Resultados

7.- Toma de Decisiones.

ORDENAMIENTO DE DATOS NO AGRUPADOS

A continuación se presentan las calificaciones obtenidas por 50 estudiantes de la UCV, en finales de estadística 1. Evaluados en escala de 1 a 20 puntos.

6	12	20	17	3	6	8	10	10	10
9	14	15	16	18	17	13	11	11	9
9	4	7	14	13	12	10	9	13	15
19	5	19	20	3	5	7	11	18	17
6	7	8	14	12	18	10	9	14	16

(X)	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
Fi	2	1	2	3	3	2	5	5	3	3	3	4	2	2	3	3	2	2	n=50

PASOS A SEGUIR:

1.- Se busca el Mínimo y el Máximo valor de la serie, en esta caso Mínimo=3 y Máximo=20

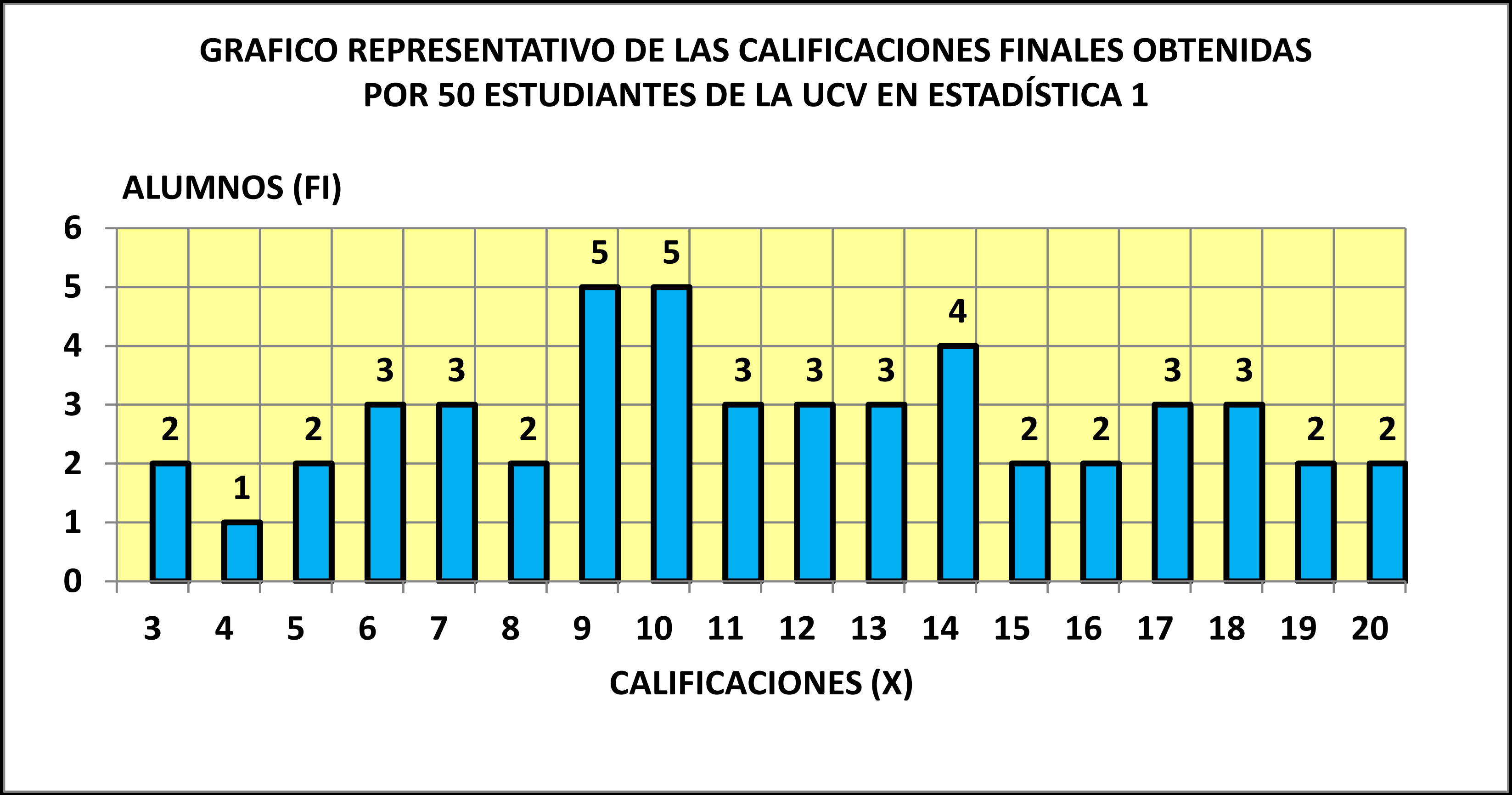
2.- Se ordenan los valores de la variable de menor a mayor

3.- Se contabilizan las frecuencias correspondientes a cada valor de la serie.

En la línea se puede apreciar el conteo de frecuencias para la calificación de 3 puntos.

REPRESENTACIÓN GRÁFICA DE LOS DATOS ORDENADOS NO AGRUPADOS

(X)NOTAS	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
(Fi) ALUMNOS	2	1	2	3	3	2	5	5	3	3	3	4	2	2	3	3	2	2	n=50



ORDENAMIENTO DE DATOS AGRUPADOS EN INTERVALOS

En este curso como la intención no es la de establecer una polémica interminable acerca de cuál debe ser el criterio para el agrupamiento de los datos, en tal sentido utilizaremos la formula se STURGES:

$$IC = \frac{\text{Valor Máximo} - \text{Valor Mínimo}}{1 + (3{,}32 * \text{Log } n)}$$

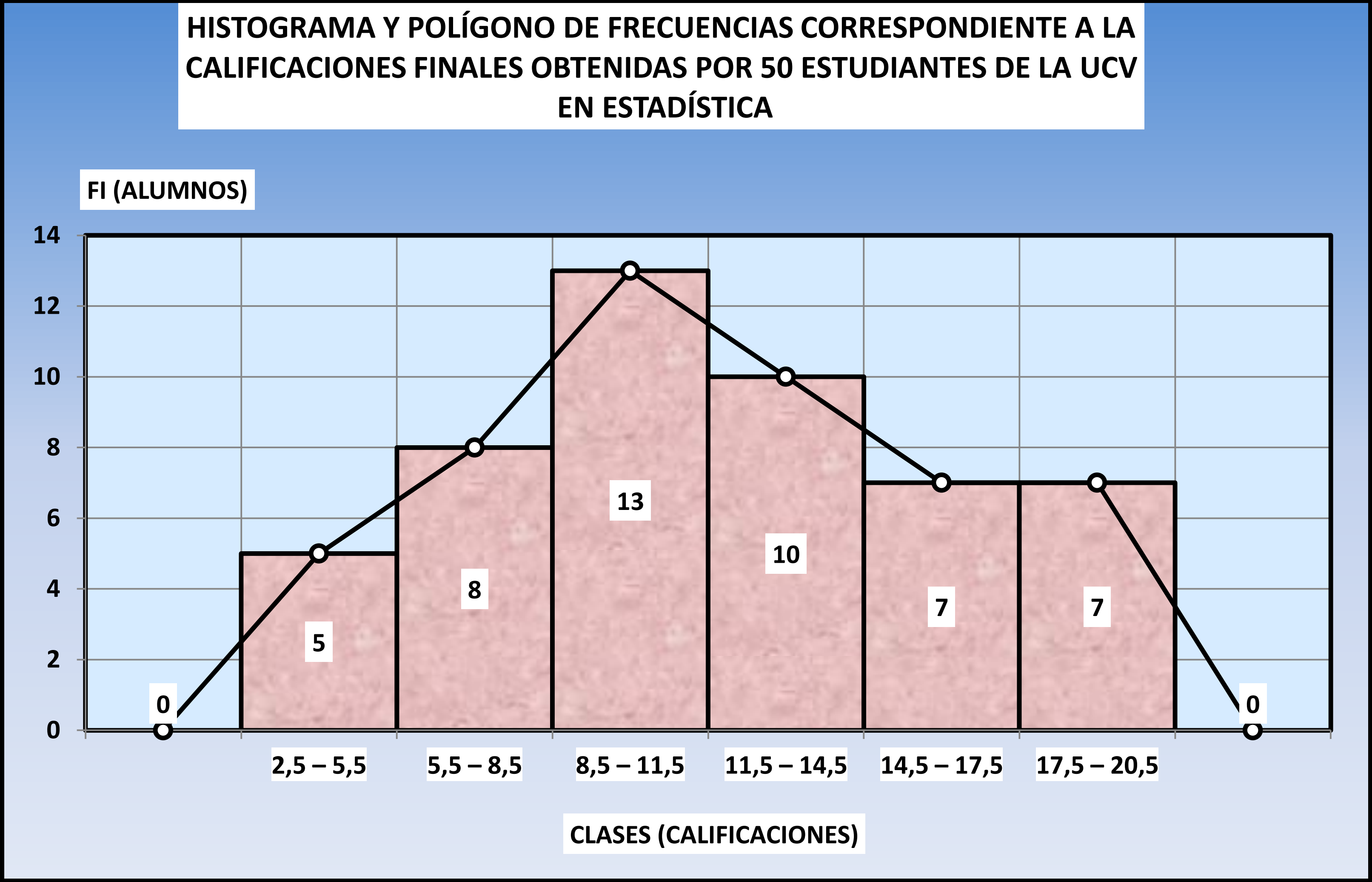
$$IC = \frac{(20 - 3)}{1 + (3{,}32 * \text{Log } 50)} = \frac{17}{6{,}6} = 2{,}57 \cong 3$$

Es importante señalar, que cuando los datos están expresados en números enteros como en este caso el intervalo se toma como un valor entero, por eso se aproximó el intervalo obtenido de 2,57 a 3. Sí los datos estuvieran expresados con decimales en el intervalo se toman tantos decimales como tengan los datos de la muestra con la cual se etá trabajando

(X)NOTAS	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
(Fi) ALUMNOS	2	1	2	3	3	2	5	5	3	3	3	4	2	2	3	3	2	2	n=50

CLASES	Fi	Fr	F%	FA	FAr	FA%	Xi	LIMITES REALES
3 – 5	5	5 / 50=0.1	10	5	5 / 50=0.1	10	4	2,5 – 5,5
6 – 8	8	8 / 50=0.16	16	13	13 / 50=0.26	26	7	5,5 – 8,5
9 – 11	13	13 / 50=0.26	26	26	26 / 50=0.52	52	10	8,5 – 11,5
12 – 14	10	10 / 50=0.2	20	36	36 / 50=0.72	72	13	11,5 – 14,5
15 – 17	7	7 / 50=0.14	14	43	43 / 50=0.86	86	16	14,5 – 17,5
18 - 20	7	7 / 50=0.14	14	50	50 / 50=1.00	100	19	17,5 – 20,5
∑	50	1.00	100					LI - LS

REPRESENTACIÓN GRÁFICA DE LOS DATOS AGRUPADOS (HISTOGRAMA Y POLÍGONO DE FRECUENCIAS)



OTRAS TABLAS Y REPRESENTACIONES GRÁFICAS Existen otras formas de presentación de resultados y representaciones gráficas que se suelen utilizar en la Estadística Descriptiva, tales como las tablas de frecuencias y los gráficos de torta. Veamos Unos ejemplos:

Un estudio acerca del número de estudiantes inscritos en la Sede de la UNEG, Región Ciudad Bolívar por proyectos de carrera, arrojó los siguientes resultados: (CIFRAS HIPOTETICAS): Educación 480 , Administración 320, Contaduría 265 y Turismo 183. Elabore una presentación por tabla y gráficamente de estos resultados:

CUADRO Nº 1

ALUMNOS POR PROYECTO DE CARRERA UNEG CIUDAD BOLIVAR

CARRERA	Fi	%
EDUCACION	480	38,46
AMINISTRACION	320	25,64
CONTADURIA	265	21,23
TURISMO	183	14,66
TOTAL	1248	100

Fuente: OFOCINA DE CONTROL DE ESTUDIOS. UNEG. CD BOLIVAR

Número del Cuadro

Título del Cuadro

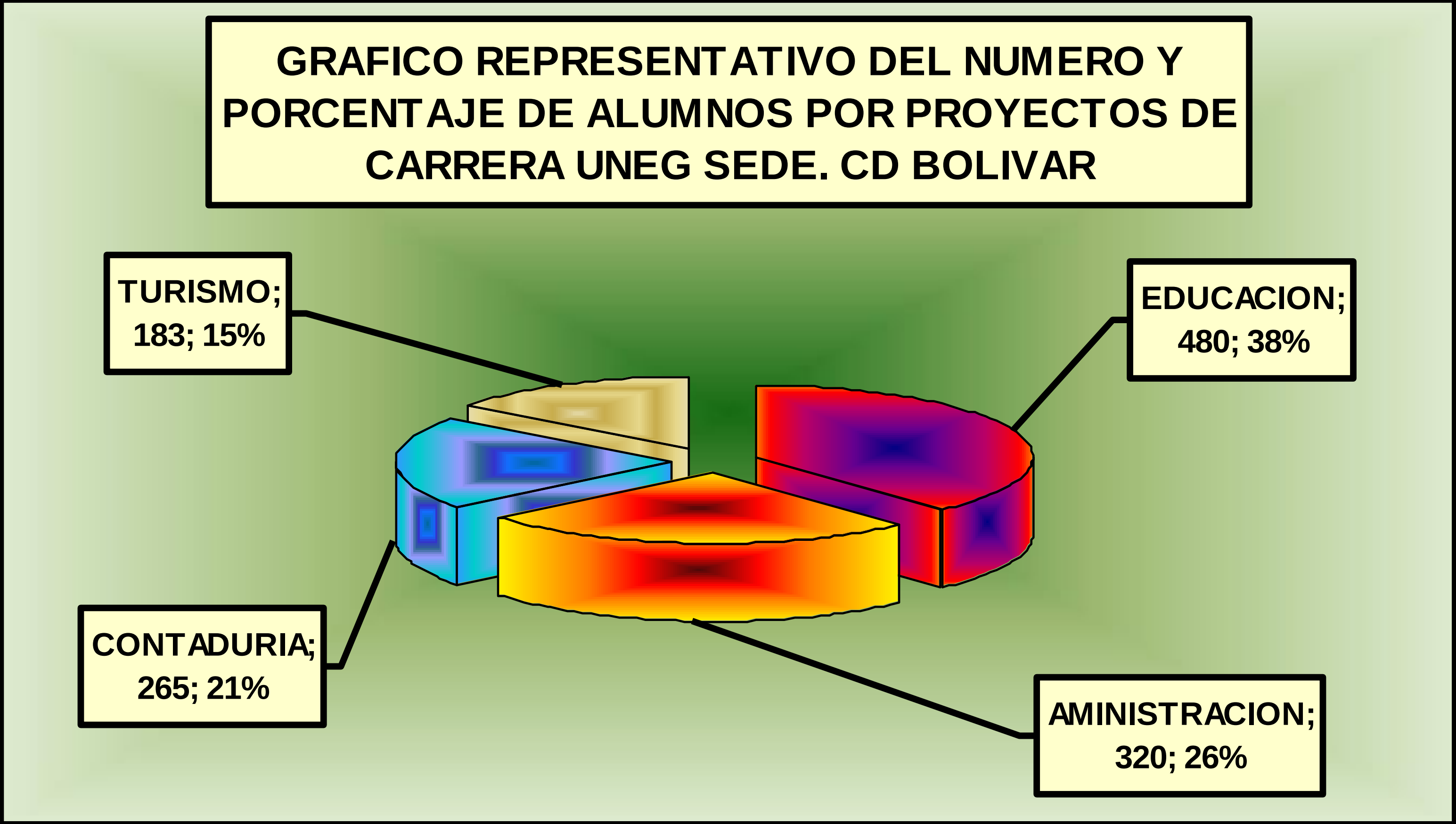
Encabezamiento

Columna Matriz

Fuente de los Datos

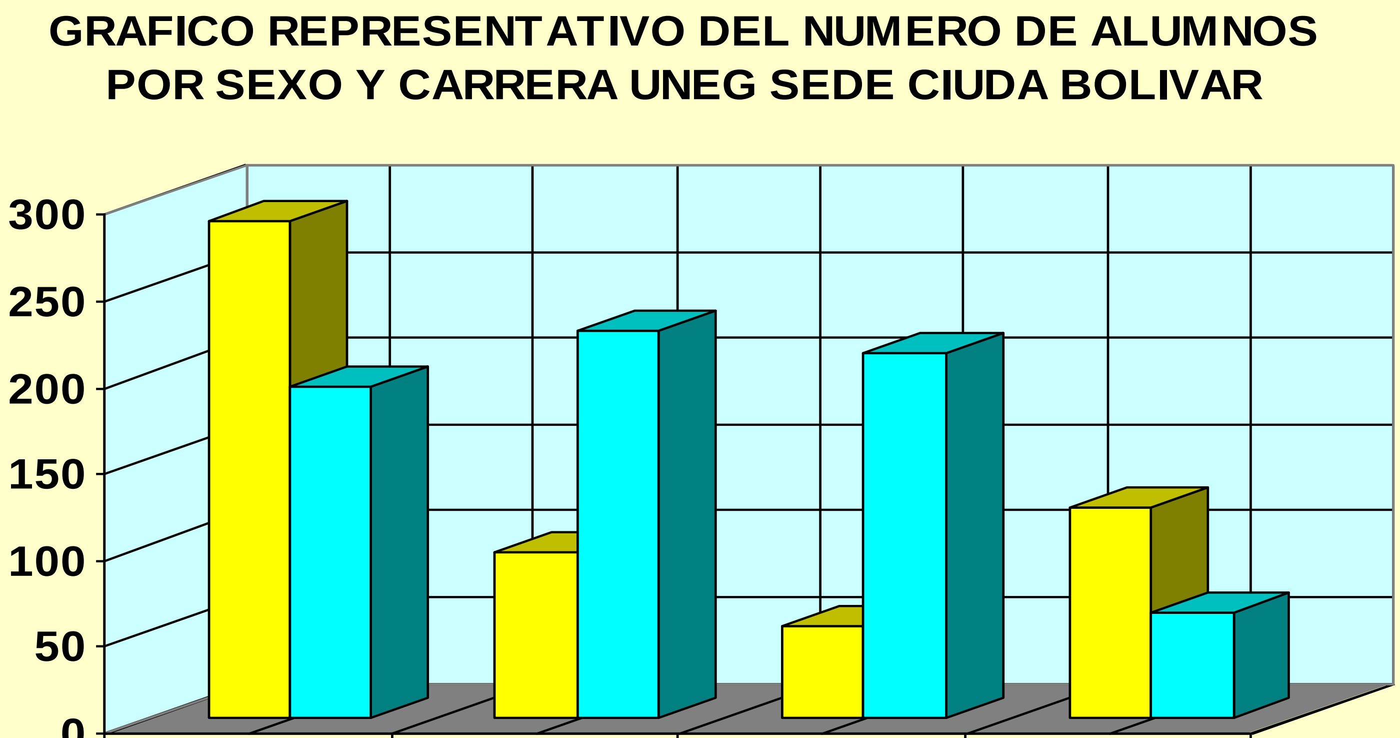
GRÁFICO DE TORTA O ÁREA

Si los datos de la tabla anterior, se desearan representar de forma gráfica tendríamos:



REPRESENTACIÓN GRÁFICA

En este caso utilizaremos el gráfico de columnas con dos categorías, recordemos que los alumnos han sido discriminados por sexo.



	EDUCACION	AMINISTRACION	CONTADURIA	TURISMO
H	288	96	53	122
V	192	224	212	61

LAS MEDIDAS DE TENDENCIA CENTRAL:

Las Medidas de Tendencia Central son aquellas que se utilizan para describir el comportamiento de una serie de datos en conjunto, ya que estos por separado no dan un valor o medida central alrededor del cual se concentre la mayoría. En este sentido existen medidas estadísticas que se usan como promedio. Las basadas en las propiedades matemáticas que se dividen en dos categorías: Promedios computables o matemáticos, y las medidas de posición. Las primeras comprenden: La media aritmética, la media geométrica, la media armónica y la media cuadrática, las últimas comprenden: la mediana y la moda.

LA MODA O MODO:

La moda o modo esta representada por el o los valores de una distribución, que tienen la mayor frecuencia. En otras palabras es el valor o los valores más recurrentes en una distribución de datos en estudio. Si los datos no están agrupados la moda se consigue por simple inspección ocular, y será el valor o los valores más recurrentes de la distribución.

EJEMPLO DE CÁLCULO PARA DATOS NO AGRUPADOS:

En este caso se tomarán como valores de referencia, las notas de los 50 estudiantes de la UCV en finales de Estadística I. Y Procederíamos de la forma siguiente:

(X)	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
Fi	2	1	2	3	3	2	5	5	3	3	3	4	2	2	3	3	2	2	n=50

En esta distribución se puede apreciar claramente que los valores más recurrentes o que tienen la mayor frecuencia son el 9 y el 10 con una frecuencia absoluta de 5 para cada valor, de manera que se presenta una distribución de calificaciones con 2 modas por lo tanto es bimodal.

CÁLCULO DE LAS MEDIDAS DE TENDENCIA CENTRAL CON DATOS AGRUPADOS

LA MODA O MODA: Está representada por el o los valores de una distribución, que tienen la mayor frecuencia. En otras palabras que más se repiten en una distribución. Si l os datos no están agrupados la moda se consigue por simple inspección ocular, tal como ya se explicó, y cuando los datos están agrupados en intervalos, la moda se calcula aplicando la siguiente formula:

$$MO = LI + \left(\frac{D1}{D1 + D2}\right) * IC$$

DONDE:

LI = Límite Inferior de la clase que tiene la mayor frecuencia (Clase modal)

D1 = Diferencia entre la mayor frecuencia y la inmediata anterior a ella.

D2 = Diferencia entre la mayor frecuencia y la inmediata posterior a ella.

IC = Intervalo de clase previamente calculado.

EJEMPLO DE CÁLCULO DE LA MODA CON DATOS AGRUPADOS CON LA MISMA DISTRIBUCIÓN DE DATOS CON LA QUE SE ESTA TRABAJANDO

CLASES	FI	LÍMITES REALES
3 – 5	5	2,5 – 5,5
6 – 8	8	5,5 – 8,5
9 – 11	13	8,5 – 11,5
12 – 14	10	11,5 – 14,5
15 – 17	7	14,5 – 17,5
18 - 20	7	17,5 – 20,5
$\sum$	50	LI - LS

PASOS A SEGUIR:

1.- Se ubica la mayor frecuencia absoluta (FI), que en este caso es (13)

2.- Se ubica el límite real inferior de la clase que contiene la mayor frecuencia absoluta, en este caso (8,5)

3.- Se saca la diferencia 1 (D1) = 13-8 = (5)

4.- Se saca la diferencia 2 (D2) = 13-10 = (3)

5.- Se calcula la moda aplicando la formula respectiva.

CONTINUACIÓN

DATOS:

Li = (8,5)

(D1) = (13-8) = (5)

(D2) = (13-10) = (3)

$$Mo = Li + \left(\frac{D1}{D1 + D2} \right) * Ic$$

$$MO = 8{,}5 + \left(\frac{5}{5 + 3} \right) * 3 = (10{,}37)\text{Puntos}$$

Este resultado indica que la calificación más recurrente en los datos agrupados es de aproximadamente 10,37 puntos.

LA MEDIANA: La mediana esta representada por el valor central de una distribución, y es definida como el valor que divide una distribución de datos en dos partes iguales. Esta medida de tendencia central es más una medida de posición central que de tendencia central, ya que ella coincide por deducción lógica con el cuartil 2, el decil 5 y el percentil 50.

Al igual que la MODA, la MEDIANA puede calcularse tanto para datos no agrupados como para datos agrupados. Cuando los datos no están agrupados, hay que considerar si la distribución de datos es par o impar. Si es par la MEDIANA vendría a ser la semi suma de los valores centrales o el promedio de los valores centrales, pero si la distribución es impar la MEDIANA vendría a ser el valor central de la distribución.

EJEMPLO DE CÁLCULO DE LA MEDIANA CON DATOS NO AGRUPADOS EN UNA MUESTRA PAR

(SE TIENEN LAS SIGUIENTES CALIFICACIONES)

8	10	9	7	7
5	6	11	11	12

PASOS A SEGUIR:

1.- Se ordenan los datos.

2.- Se buscan los dos valores centrales, se suman y se dividen entre 2

5	6	7	7	8	9	10	11	11	12

(8 + 9)/2 = (8,5) PUNTOS SERÁ LA MEDIANA

EJEMPLO DE CÁLCULO DE LA MEDIANA CON DATOS NO AGRUPADOS EN UNA MUESTRA IMPAR

(SE TIENEN LAS SIGUIENTES CALIFICACIONES)

8	10	9	7	7	5	6	11
11	12	13	10	5			

PASOS A SEGUIR:

1.- Se ordenan los datos.

2.- Se buscan el valor central.

5	5	6	7	7	8	9	10	10	11	11	12	13

(9) PUNTOS SERÁ LA MEDIANA

EJEMPLO DE CÁLCULO DE LA MEDIANA CON DATOS AGRUPADOS
(CON LA DISTRIBUCIÓN CON LA CUAL SE HA VENIDO TRABAJANDO)

Como ya se dijo, la MEDIANA puede calcularse tanto para datos no agrupados como para datos agrupados en intervalos. En los casos anteriores había que tomar en cuenta el tipo de distribución (PAR o IMPAR), sin embargo cuando los datos están agrupados en intervalos de clase, esta particularidad no hay que tomarla en cuenta, ya que la MEDIANA se calcula a través de la siguiente ecuación:

$$\boxed{Me = Li + \left(\frac{(n/2) - FAant}{Fi} \right) * Ic}$$

PASOS A SEGUIR:

1.- Se divide n/2 (posición de la mediana), este valor se busca en la frecuencias acumuladas (FA), si no aparece debemos ubicarnos en la primera frecuencia acumulada que contenga el valor, y a partir de allí obtenemos los datos restantes de la formula.

2.- Se busca el límite real inferior que corresponda a la mayor frecuencia acumulada en la cual nos ubicamos.

3.- Se ubica la frecuencia acumulada anterior a donde nos ubicamos.

4.- Se ubica la frecuencia absoluta (Fi) correspondiente.

5.- Seguidamente se aplica la formula correspondiente.

CLASES	FI	FA	LÍMITES REALES
3 – 5	5	5	2,5 – 5,5
6 – 8	8	13	5,5 – 8,5
9 – 11	13	26	8,5 – 11,5
12 – 14	10	36	11,5 – 14,5
15 – 17	7	43	14,5 – 17,5
18 - 20	7	50	17,5 – 20,5
∑	50		LI - LS

PASOS A SEGUIR:

1.- n/2 = 50/2 = 25. Nótese que este valor no se encuentra en la FA, por lo tanto lo ubicamos en (26)

2.- Se ubica El límite real inferior a 26 es igual a (8,5)

3.- Se busca La frecuencia acumulada anterior (FA) es igual a (13)

4.- Se busca La frecuencia absoluta correspondiente a 26 es igual a (13)

5.- Se coloca que es igual a (3)

6.- Se calcula la Median aplicando la formula respectiva.

CONTINUACIÓN

DATOS:

Li = (8,5)

FAant = (13)

Fi = (13)

IC = (3)

$$Me = Li + \left(\frac{(n/2) - FAant}{Fi} \right) * Ic$$

$$Me = 8{,}5 + \left(\frac{25 - 13}{13} \right) * 3 = 11{,}26 \text{ Puntos}$$

Este resultado indica que la Mediana o el valor central de la distribución de calificaciones es de 11,26 puntos, aproximadamente.

EL PROMEDIO ARITMÉTICO O MEDIA ARITMÉTICA

La media aritmética o promedio aritmético, es la medida de tendencia central más conocida y utilizada. La misma no es otra cosa que el cociente de dividir la sumatoria de todos los valores que toma una variable entre el número de ellos (n), por lo tanto esta medida depende en gran medida de cada uno de los valores que forman la serie.

Hay que señalar, que esta medida de tendencia central, se encuentra afectada por valores extremos de la distribución, es decir, aquellos que están situados lejos del centro de la serie, y por lo tanto no es recomendable como medida representativa, en distribuciones en las cuales los datos se encuentran muy dispersos; o en aquellas que son asimétricas.

Cuando los datos no están agrupados en intervalos, la madia se puede calcular utilizando la siguiente formula:

$$\mu = \frac{\Sigma X}{n}$$

EJEMPLO DE CÁLCULO DE LA MEDIA ARITMÉTICA CON DATOS NO AGRUPADOS

Sean las calificaciones obtenidas por un grupo de 50 estudiantes de la Universidad Central de Venezuela en finales de Estadística I (Evaluados en escala de 1 a 20 puntos)

6	12	20	17	3	6	8	10	10	10
9	14	15	16	18	17	13	11	11	9
9	4	7	14	13	12	10	9	13	15
19	5	19	20	3	5	7	11	18	17
6	7	8	14	12	18	10	9	14	16
TOTALES POR COLUMNAS									
49	42	69	81	49	58	48	50	66	67

$\Sigma = \Sigma$ 49 + 42 + 69 + 81 + 49 + 58 + 48 + 50 + 66 + 67 = 579 / 50 = (11,58) Puntos

De manera que el promedio correspondiente a las calificaciones del grupo de estudiantes es de aproximadamente 11,58 puntos

EJEMPLO DE CÁLCULO DE LA MEDIA ARITMÉTICA CON LA MISMA DISTRIBUCION DE DATOS AGRUPADOS QUE SE VIENE TRABAJANDO

PASOS A SEGUIR:

1.- Se selecciona de la tabla de distribución de frecuencias los Puntos Medios (Xi)

2.- Se selecciona de la tabla de distribución de frecuencias las Frecuencias Absolutas (Fi)

3.- Se multiplica cada (Xi) por su correspondiente (Fi)

4.- Se suman los productos y se totalizan

4.- Se calcula la Media Aritmética aplicando la formula respectiva

$$\mu = \frac{\Sigma Xi * Fi}{n}$$

Fi	Xi	Límites Reales	XI*FI
5	4	2,5 – 5,5	20
8	7	5,5 – 8,5	56
13	10	8,5 – 11,5	130
10	13	11,5 – 14,5	130
7	16	14,5 – 17,5	112
7	19	17,5 – 20,5	133
50		LI - LS	Σ561

$$\mu = \frac{\Sigma Xi * Fi}{n}$$

$$\mu = \frac{561}{50} = (11,22)\text{Puntos}$$

Este resultado indica que la calificación promedio correspondiente a esta distribución de datos agrupados es de 11,22 puntos aproximadamente.

LAS MEDIDAS DE POSICIÓN

Son medidas o valores estadísticos, que determinan una posición dentro de una distribución ordenada de datos. Las medidas de posición más comúnmente utilizadas son los: Cuartiles, Deciles y Percentiles. Sin embargo es conveniente señalar que los percentiles por ser más abarcadores como medida de posición, pueden sustituir a cualquier cuartil o decil. En este sentido sólo nos avocaremos a los procedimientos de cálculo e interpretación de estos.

LOS PERCENTILES

Son medidas estadísticas que permiten dividir una distribución ordenada de datos en 100 partes iguales, son 99 y cada percentil representa un 1% de la distribución. Por otra parte es necesario señalar, que las medidas de posición son útiles para el análisis y estudio de distribuciones de datos muy numerosas, y en las cuales resulta muy complicado este trabajo tomando todos los datos en su conjunto.

Los percentiles se pueden calcular aplicando la siguiente formula:

$$Px = Li + \left(\frac{(n * x / 100) - FAant}{Fi} \right) * Ic$$

EJEMPLO DE CÁLCULO DE LOS PERCENTILES CON DATOS AGRUPADOS ANTERIORMENTE

PASOS A SEGUIR:

1.- Se multiplica el valor del percentil que se desea calcular por (n) y se divide entre (100) (posición del percentil), este valor se busca en la frecuencias acumuladas (FA), si no aparece debemos ubicarnos en la primera frecuencia acumulada que contenga el valor, y a partir de allí obtenemos los datos restantes de la formula.

2.- Se busca el límite real inferior que corresponda a la mayor frecuencia acumulada en la cual nos ubicamos.

3.- Se ubica la frecuencia acumulada anterior a donde nos ubicamos.

4.- Se ubica la frecuencia absoluta (Fi) correspondiente.

5.- Seguidamente se aplica la formula correspondiente.

EJEMPLO DE CÁLCULO PARA EL PERCENTIL 25 DE LA DISTRIBUCIÓN

CLASES	FI	FA	LÍMITES REALES
3 – 5	5	5	2,5 – 5,5
6 – 8	8	13	5,5 – 8,5
9 – 11	13	26	8,5 – 11,5
12 – 14	10	36	11,5 – 14,5
15 – 17	7	43	14,5 – 17,5
18 - 20	7	50	17,5 – 20,5
∑	50		LI - LS

PASOS A SEGUIR:

1.- n*25/100 = 50*25/100 = (12,5). Nótese que este valor no se encuentra en la FA, por lo tanto lo ubicamos en (13)

2.- El límite real inferior correspondiente a (13) es igual a (5,5)

3.- La frecuencia acumulada anterior a (13) es igual a (5)

4.- La frecuencia absoluta correspondiente a (13) es igual a (8)

5.- El intervalo de clase o agrupación recordemos que es igual a (3)

CONTINUACIÓN

DATOS:

Li= (5,5)

Faant= (5)

Fi= (8)

Ic= (3)

$$Px = Li + \left(\frac{(n * x / 100) - FAant}{Fi} \right) * Ic$$

$$Px = 5,5 + \left(\frac{12,5 - 5}{8} \right) * 3 = (8,31) Puntos$$

Este resultado indica que entre las calificaciones de 2,5 y 8,31 puntos, se encuentra el 25% inferior del grupo de estudiantes.

VEAMOS OTRO EJEMPLO DE CÁLCULO EN ESTE CASO PARA EL PERCENTIL 75

CLASES	FI	FA	LÍMITES REALES
3 – 5	5	5	2,5 – 5,5
6 – 8	8	13	5,5 – 8,5
9 – 11	13	26	8,5 – 11,5
12 – 14	10	36	11,5 – 14,5
15 – 17	7	43	14,5 – 17,5
18 - 20	7	50	17,5 – 20,5
∑	50		LI - LS

PASOS:

1.- n * 75/100 = 50*75/100 = (37,5). Nótese que este valor no se encuentra en la FA, por lo tanto lo ubicamos en (43)

2.- El límite real inferior correspondiente a (43) es igual a (14,5)

3.- La frecuencia acumulada anterior a (43) es igual a (36)

4.- La frecuencia absoluta correspondiente a (43) es igual a (7)

5.- El intervalo de clase o agrupación recordemos que es igual a (3)

CONTINUACIÓN

DATOS:

Li = (14,5)

Faant = (36)

Fi = (7)

Ic = (3)

$$Px = Li + \left(\frac{(n * x / 100) - FAant}{Fi} \right) * Ic$$

$$Px = 14{,}5 + \left(\frac{(50 * 75 / 100) - 36}{7} \right) * 3 = (15{,}14) Puntos$$

Este resultado indica que entre las calificaciones de 2,5 y 15,14 puntos, se encuentra el 75% del grupo de estudiantes.

LAS MEDIDAS DE VARIABILIDAD Y DISPERSIÓN.

Como ya hemos explicado en este curso, existe un valor significativo que se adopta como medida de tendencia central, sin embargo esto plantea un problema, ya que todas las observaciones de una serie no tienen el mismo valor, y por lo tanto se presentan variaciones entre ellos.

Casi sin excepción se puede decir que los valores de una distribución de datos difieren del promedio o la medida de tendencia central que se adopte, y por esto la utilidad de un promedio depende en gran medida de su poder de representatividad con respecto al conjunto de observaciones o valores de una serie.

En tal sentido se puede asumir, que cuando los valores de una serie presentan pocas diferencias o están muy concentrados alrededor de un promedio o medida de tendencia central, este promedio será muy representativo del conjunto de valores, pero si por el contrario están muy dispersos con relación a este u otra medida de tendencia central que se haya adoptado, el promedio o medida será poco representativo del conjunto de la serie.

A continuación pasaremos a estudiar las más utilizadas de las medidas de dispersión las cuales están representadas por la varianza y la desviación típica. Cuando se desarrolla el cálculo de la desviación media, hay que prescindir de los signos de los desvíos, esto debido a que la sumatoria de ellos con respecto al promedio aritmético da como resultado cero (0).

En esos casos, para prescindir de los signos se tomó el valor absoluto de los desvíos. Cuando se calcula tanto la varianza como la desviación típica esta particularidad se resuelve elevando al cuadrado cada uno de los desvíos, ya que un valor negativo al elevarlo al cuadrado, se convierte en positivo y se conserva al mismo tiempo la información pertinente a las desviaciones.

Procediendo de esta forma se obtienen los valores de estas dos medidas que son las más útiles para determinar la dispersión que presentan los dato s de una distribución o serie.

LA VARIANZA

Cuando los datos no están agrupados la varianza se calcula aplicando la siguiente formula:

$$\boxed{S^2 = \frac{\Sigma(Xi - \mu)^2}{n-1}}$$

EJEMPLO DE CÁLCULO (CON DATOS NO AGRUPADOS)

Sean las calificaciones obtenidas por un grupo de 5 estudiantes de la Universidad de Carabobo en finales de Matemática I (Evaluados en escala de 1 a 20 puntos)

12	14	6	9	20

PASOS A SEGUIR:

1.- Se calcula la Media Aritmética de forma igual a como se explicó anteriormente.

2.- Se resta cada valor menos la Media Aritmética

3.- Se eleva el producto de la resta al cuadrado

4.- Se calcula la varianza.

CONTINUACIÓN

NOTAS	μ	(x – μ)	(x – μ)2
12	12,2	-0,2	0,04
14	12,2	1,8	3,24
6	12,2	-6,2	38,44
9	12,2	-3,2	10,24
20	12,2	7,8	60,84
Σ 61			Σ 112,8

$$\mu = \frac{\Sigma X}{n}$$

$$\mu = \frac{61}{5} = (12{,}2)\text{Puntos}$$

$$S^2 = \frac{\Sigma(Xi - \mu)^2}{n-1}$$

$$S^2 = \frac{112{,}8}{5-1} = (28{,}2)\text{Puntos}$$

Luego podemos decir que la distribución de datos presenta una variabilidad respecto del promedio aritmético de 28,2 puntos

EJEMPLO DE CÁLCULO DE LA VARIANZA CON DATOS AGRUPADOS

Asumiendo la distribución de datos agrupados con la cual se ha venido trabajando en este curso, y el valor obtenido para la Media Aritmética.

PASOS A SEGUIR:

1.- Se calcula la Media Aritmética de forma igual a como se explicó anteriormente.

2.- Se resta cada valor menos la Media Aritmética

3.- Se eleva el producto de la resta al cuadrado

4.- Se calcula la varianza.

CONTINUACIÓN

xi	μ	(xi – μ)	(xi – μ)2	XI*FI	Fi	(xi – μ)2 * Fi
4	11,6	-7,62	58,06	20	5	290,3
7	11,6	-4,62	21,34	56	8	170,72
10	11,6	-1,62	2,62	130	13	34,06
13	11,6	1,38	1,9	130	10	19
16	11,6	4,38	19,18	112	7	134,26
19	11,6	7,38	54,46	133	7	381,22
				Σ561	50	1029,56

$$\mu = \frac{\Sigma Xi * Fi}{n}$$

$$\mu = \frac{561}{50} = (11{,}62)\text{Puntos}$$

$$S^2 = \frac{\Sigma(Xi-\mu)^2 * FI}{n-1}$$

$$S^2 = \frac{1029{,}56}{50-1} = (21{,}01)\text{Puntos}$$

Luego podemos decir que la distribución de datos presenta una variabilidad respecto del promedio aritmético de 21,01 puntos

LA DESVIACIÓN TÍPICA:

Está representada por la raíz cuadrada de la varianza, y no representa otra cosa que el desvío típico de los valores de una serie con respecto al promedio aritmético o media aritmética. En tal sentido se puede asumir, que una distribución de datos está más dispersa que otra en la medida que su desviación típica sea mayor, se denota por la letra (S) cuando es de una muestra , o la letra (σ) cuando se trata de una población. Viene dada por la siguientes formulas.

PARA DATOS NO AGRUPADOS

$$S = \sqrt{\frac{\Sigma(Xi - \mu)^2}{n - 1}}$$

PARA DATOS AGRUPADOS

$$S = \sqrt{\frac{\Sigma(Xi - \mu)^2 * Fi}{n - 1}}$$

EJEMPLO DE CÁLCULO CON DATOS NO AGRUPADOS Y AGRUPADOS CON LOS DATOS QUE SE HAN UTILIZADO EN AMBOS CASOS

$$S = \sqrt{\frac{\Sigma(Xi - \mu)^2}{n-1}}$$

$$S = \sqrt{28{,}2} = (5{,}31)\text{Puntos}$$

Este resultado indica que los datos no agrupados presentan un desvío típico de los datos con respecto a la Media Aritmética de 5,31 Puntos

$$S = \sqrt{\frac{\Sigma(Xi - \mu)^2 * Fi}{n-1}}$$

$$S = \sqrt{21{,}01} = (4{,}58)\text{Puntos}$$

Este resultado indica que los datos agrupados presentan un desvío típico de los datos con respecto a la Media Aritmética de 4,58 Puntos

COEFICIENTE DE VARIACIÓN

El coeficiente de variación es una medida estadística utilizada para determinar el grado de heterogeneidad que presentan los datos de una serie. Por lo tanto una distribución de datos será más heterogénea que otra en la medida que su coeficiente de variación sea mayor. El coeficiente de variación se expresa porcentualmente, viene dado por la siguiente formula de cálculo:

$$\boxed{C.V = \frac{S}{\mu} * 100}$$

CÁCULO CON LOS DATOS DE LA MEDIA Y DESVIACION ANTES CALCULADOS CON LOS DATOS AGRUPADOS

$$\boxed{C.V = \frac{4{,}58}{11{,}62} * 100 = (39{,}41\%)}$$

De esta forma podemos decir que la distribución de calificaciones presenta una heterogeneidad del 39,41% y una homogeneidad del 60,59%.

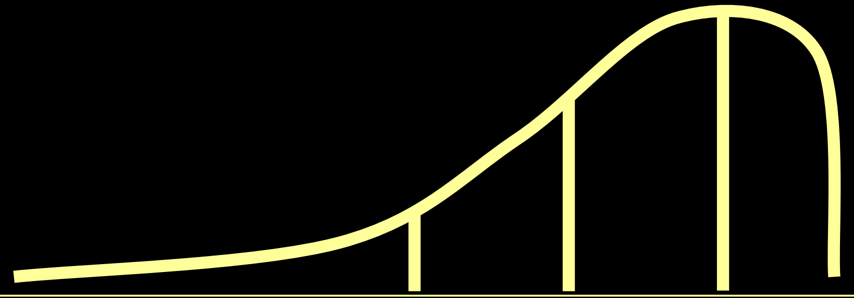
COEFICIENTE DE ASIMETRÍA

El coeficiente de asimetría es una medida estadística utilizada para describir la forma como se encuentran distribuidas las frecuencias correspondientes a los diferentes valores de la variable en estudio.

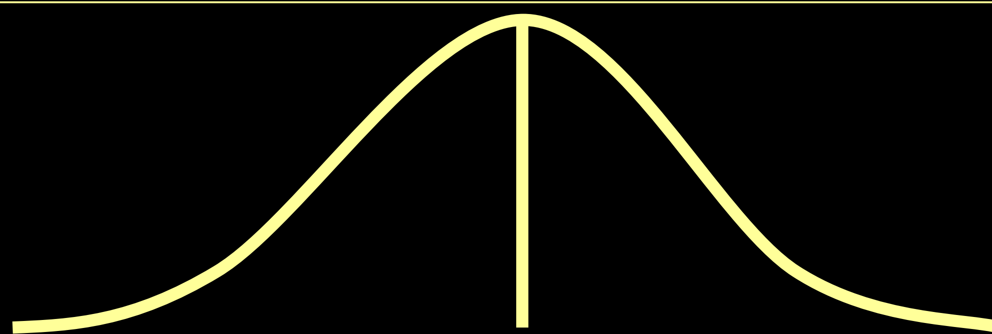
Este coeficiente permite medir la falta de simetría de una distribución, entendiéndose por distribución simétrica, aquella en la cual existe un punto o valor central que divide una distribución en dos partes exactamente iguales, en otras palabras, la distribución de las frecuencias correspondientes a los valores menores es igual a la distribución de las frecuencias de los valores mayores.

En este sentido cuando una distribución es asimétrica negativa existe mayor concentración de valores a la izquierda de la media que a su derecha, y cuando una distribución es asimétrica positiva se presenta todo lo contrario.

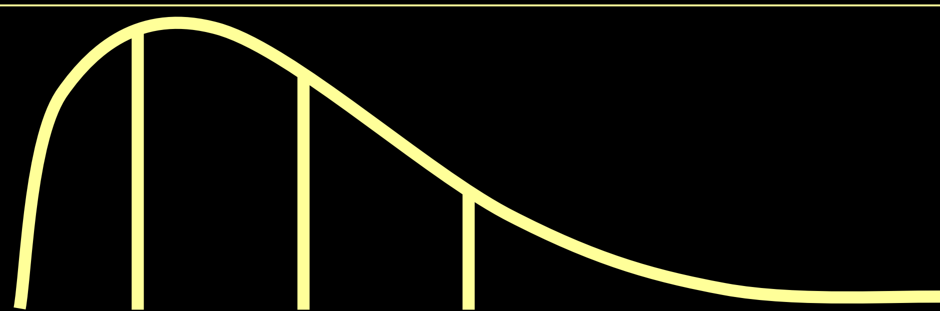
VEAMOS GRÁFICAMENTE



ASIMETRIA NEGATIVA: PROMEDIO < MEDIANA < MODA



SIMETRIA: PROMEDIO = MEDIANA = MODA



ASIMETRIA NEGATIVA: PROMEDIO > MEDIANA > MODA

El coeficiente de asimetría, se puede calcular aplicando la formula de Pearson, en función de la media y la mediana, aun cuando existen otras

$$AP = \frac{3 * (\mu - Med)}{S}$$

EJEMPLO DE CÁLCULO

DATOS:

μ = 11,62

Med = 11,26

S = 4,58

$$AP = \frac{3 * (11{,}62 - 11{,}26)}{4{,}58} = (0{,}23)$$

INTERPRETACIÓN DEL RESULTADO

TABLA DE VALORES PARA INTERPRETACION DEL COEFICIENTE DE ASIMETRIA DE PEARSON	
COEFICIENTE NEGATIVO O POSITIVO IGUAL O MENOR A 0,37	DISTRIBUCIÓN CUASI-SIMÉTRICA
COEFICIENTE NEGATIVO O POSITIVO ENTRE 0,38 Y 1,00	DISTRIBUCIÓN ASIMÉTRICA ACEPTABLE
COEFICIENTE NEGATIVO O POSITIVO MAYOR QUE 1,00	DISTRIBUCIÓN ASIMÉTRICA TOTAL

El resultado de (0,24), indica que la distribución de calificaciones en estudio es Cuasi-simétrica positiva. Nótese que el resultado es menor a 0,37.

CURTOSIS

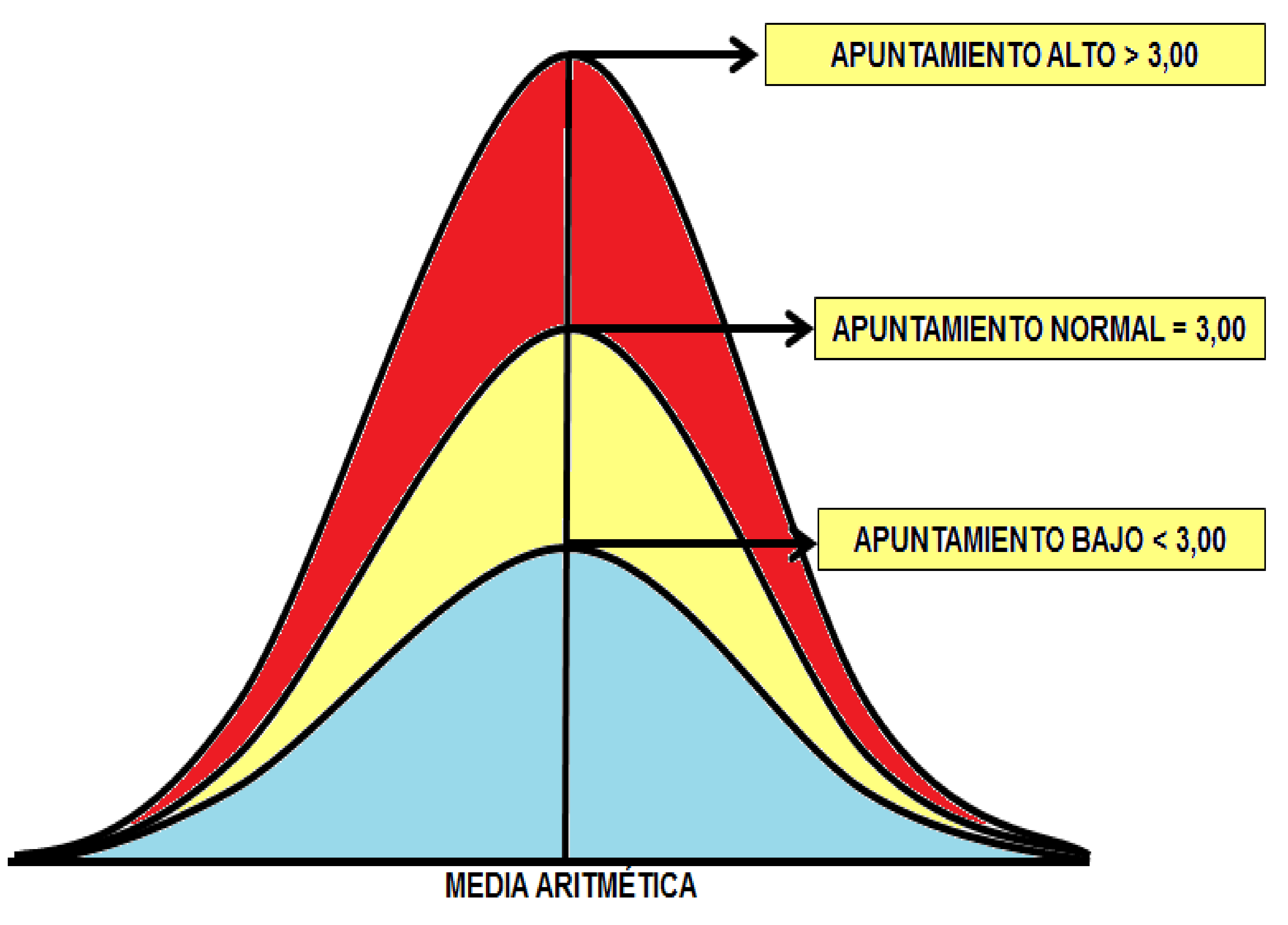
El coeficiente de curtosis, representa el grado de apuntamiento o altura de una distribución con resp ecto a la distribución normal. En este sentido una distribución de datos puede ser:

1.- LEPTOCURTICA: Apuntamiento Alto (Alta concentración de los datos en torno al centro de la serie)

2.- MESOCURTICA: Apuntamiento Normal (Concentración media de los datos en torno al centro de la serie)

3.- PLATICURTICA: Apuntamiento Bajo (Baja concentración de los datos en torno al centro de la serie)

El valor de comparación para el coeficiente de curtosis por los momentos es (3), el cual representa el valor de la curtosis normal o (Mesocurtica) , en este sentido cuando una distribución presenta un coeficiente mayor a (3), se dice que es de apuntamiento alto con respecto a la normal (Leptocúrtica) en cuyo caso los valores de la variable en estudio son muy similares, pero si por el contrario su valor es menor que (3), se dice que es de apuntamiento bajo con respecto a la normal (Platicúrtica) en cuyo caso los valores de la variable en estudio son de gran variación.


APUNTAMIENTO ALTO > 3,00
APUNTAMIENTO NORMAL = 3,00
APUNTAMIENTO BAJO < 3,00
MEDIA ARITMÉTICA

Existen varias formulas para calcular el Coeficiente de Curtosis, sin embargo en este curso utilizaremos la curtosis por los momentos o formula de Fischer, la cual es:

$$Cm^4 = \frac{\Sigma(Xi - \mu)^4 * Fi}{(n) * (S)^4}$$

El problema de la curtosis, no radica en calcular su valor, sino en determinar a través del resultado que tipo de apuntamiento se presenta en la serie de datos que se está analizando.

En tal sentido, a continuación presentamos una tabla que facilitará la labor de determinar el tipo de apuntamiento de acuerdo con el resultado obtenido

CONTINUACIÓN

TABLA DE INTERPRETACION PARA CURTOSIS POR LOS MOMENTOS	
CURTOSIS < 3	BAJA (PLATICURTICA)
CURTOSIS = 3	NORMAL (MESOCURTICA)
CURTOSIS >3	ALTA (LEPTOCURTICA)

VEAMOS UN EJEMPLO DE CÁLCULO: Tomando como referencia los datos agrupados anteriormente, y los resultados obtenidos tanto de la Media Aritmético, como de la Desviación Típica.

PASOS A SEGUIR:

1.- Si se va a trabajar con datos agrupados, hay que elaborar la tabla de distribución de frecuencias.

2.- Se calcula la Media Aritmética

3.- Se calcula la Desviación Típica

4.- Se calcule el Coeficiente de Curtosis aplicando la formula respectiva.

5.- Se interpreta el resultado

DATOS:

μ = (11,62)

S = (4,58)

n = (50)

CONTINUACIÓN

Xi	μ	Xi – μ	(Xi – μ)4	Fi	(Xi – μ)4 * Fi
4	11,6	-7,62	3371,47	5	16857,37
7	11,6	-4,62	455,58	8	3644,66
10	11,6	-1.62	6,88	13	89,53
13	11,6	1,38	3,62	10	36,26
16	11,6	4,38	368,04	7	2576,28
19	11,6	7,38	2966,37	7	20764,59
Σ				50	43968,69

$$Cm^4 = \frac{\Sigma(Xi-\mu)^4 * Fi}{(n) * (S)^4}$$

$$Cm^4 = \frac{43968{,}69}{(50) * (4{,}58)^4} = (1{,}99)$$

Este resultado indica que estamos en presencia de una curtosis menor que (3), que es el valor de referencia, por lo tanto es una curtosis PLATICURTICA. Por otro lado vemos que al ser PLATICURTICA, los valores de la distribución de datos en estudio son de gran variación