



Universidad Inca Garcilaso de la Vega
Facultad de Ciencias de la Comunicación
Turismo y Hotelería

Análisis Estadístico Descriptivo

1. *Conceptos básicos*
2. *Construcción de tablas de frecuencias*
3. *Los gráficos Estadísticos*
4. *El análisis de tendencia Central*
5. *El análisis de dispersión*

Luis Flores Cebrián

1. CONCEPTOS BASICOS

Antes de iniciar el análisis estadístico de un conjunto de datos es necesario establecer algunas precisiones conceptuales:

1.1 Estadística descriptiva y estadística inferencial

El nivel descriptivo está referido al estudio y análisis de los datos obtenidos en una muestra (n) y como su nombre lo indica describen y resumen las observaciones obtenidas sobre un fenómeno un suceso o un hecho.

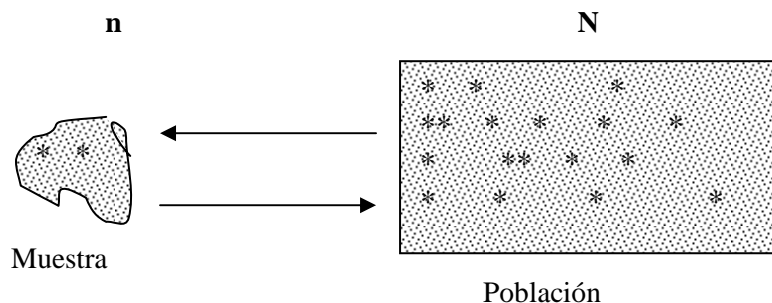
Ejemplos :

- El porcentaje de aprobados en el curso de economía ascendió a un 68%
- El 35% de los visitantes del museo era de género femenino
- La edad promedio de una muestra de compradores en el centro comercial es de 38 años

En cambio la **estadística inferencial** esta referida al procedimiento mediante el cual los resultados de la muestra se trata de hacerlos extensivos a toda la población o universo (N). Procura mostrar relaciones de causa efecto o pruebas de hipótesis.

Ejemplos :

- Los ingresos de los turistas que vienen al país son de \$ 35,000 en promedio
- Si las elecciones fueran hoy día el candidato del partido “ X “ obtendría un 45% de los votos



1.2 Población y muestra

Población es el conjunto completo de individuos, objetos que posee una o más características observables en un ámbito geográfico determinado y en un lapso de tiempo .

Muestra : es un subconjunto representativo de la población, cuando hablamos de representatividad nos referimos a que debe ser lo más parecido a posible a la población desde el punto de vista cuantitativo y cualitativo.

El muestreo puede ser :

- Muestreo al azar simple, cuando el azar es el criterio básico de elección de las unidades de muestreo
- Muestreo estratificado basado en la subdivisión de la población en subconjuntos o estratos mutuamente excluyentes

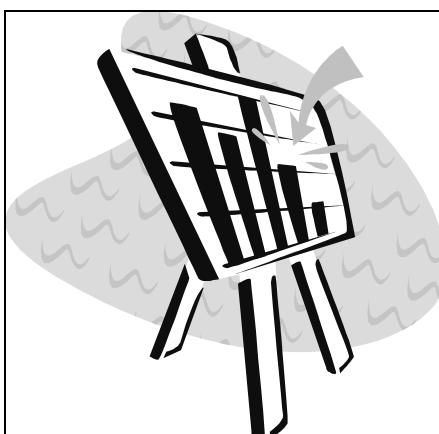
- Muestreo por áreas, cuando se trabaja en función de unidades geográficas o conglomerados
- Muestreo sistemático, cuando se sigue un patrón de elección basado en números aleatorios elegidos al azar o un factor de elección, ejemplo se elige aplicar la encuesta a cada quince visitantes al museo

En todos los casos se debe aplicar el criterio de probabilidad, es decir que cada elemento de la población tenga la misma probabilidad de ser elegido en la muestra. Este es un aspecto fundamental pues hacerlo de otra manera (elección subjetiva) impediría medir el nivel de confianza y / o de error de los resultados obtenidos.

1.3 Las variables

Las variables son las características, atributos o cualidades del objeto de investigación se pueden clasificar de muchas maneras. Ejemplo

CRITERIO	CLASIFICACIÓN	EJEMPLO
Según el número de valores que expresan	Variables dicotómicas, expresan sólo dos valores	Genero : masculino o femenino
	Variables politómicas, expresan mas de dos valores	Grado de instrucción
Según la capacidad de observación	Variables teóricas o no observables de naturaleza conceptual	Gusto, preferencia, opinión
	Variables empíricas observables se pueden medir o cuantificar	Peso, temperatura, ingresos
Según la naturaleza del valor que asumen	Variables discretas asumen valores absolutos	Nacionalidad,
	Variables continuas asumen valores fraccionarios	Ingresos mensuales
Según su función en la hipótesis	Variables independientes	Causas o antecedentes
	Variables dependientes	Consecuencias o efectos
	Variables intervinientes	Influyen pero de manera indirecta
	Variables paramétricas	Medioambientales, de entorno o macro

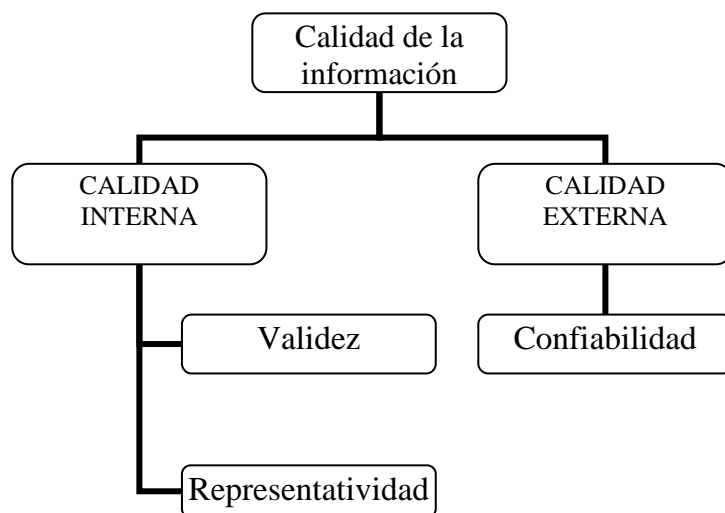


1.4 La medición

Concepto : La medición es un proceso sistemático de asignación de números a individuos , objetos o fenómenos , de tal manera que representen de una manera lógica y racional las características de los mismos

Calidad de la información .

¿ Cómo podemos medir la calidad de la información ?



CONFIABILIDAD :

Es la consistencia de puntajes en mediciones sucesivas, es decir la menor presencia de errores de medición.

Se expresa a través de la estabilidad que consiste en usar el mismo conjunto de preguntas en dos o más momentos y comparar así las respuestas obtenidas.

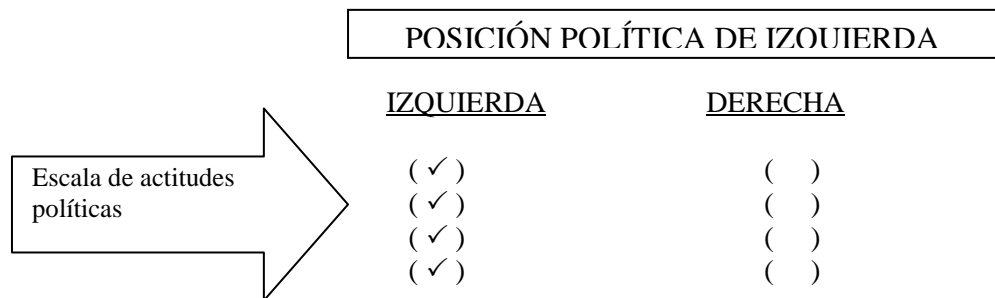
	<u>Período 1</u>	<u>Período 2</u>
Esta a favor del TLC	58%	62%
Esta en contra del TLC	22%	24%
No sabe / no opina	20%	14%

VALIDEZ :

Es el grado en que las mediciones empíricas representan o reflejan la dimensión conceptual que se supone están tratando de medir

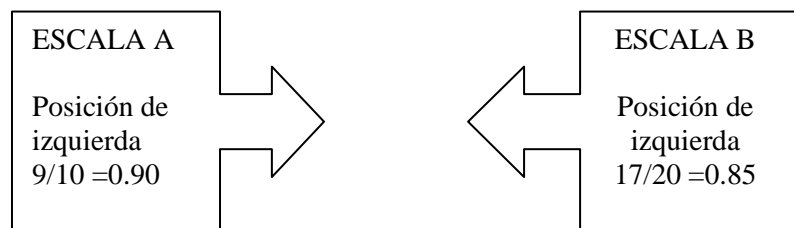
Tiene dos maneras de comprobación:

- a) **Validez predictiva** es aquella que anticipa un comportamiento o una correlación de respuestas congruente. Ejemplo :



Suponiendo que hubiera un proceso electoral y se presentara un conjunto de propuestas políticas , la validez predictiva se manifestaría cuando la persona que exhibe una preferencia por posiciones políticas de izquierda elige a candidatos de una opción política similar o afín.

- b) **Validez concurrente**, se utiliza cuando nuevos indicadores del concepto se correlacionan altamente con indicadores ya establecidos del mismo. Ejemplo :



REPRESENTATIVIDAD

Tiene que ver con el grado en que los resultados de la muestra pueden ser atribuidos a toda la población o universo . Se trabaja con los valores de **Z** (niveles de confianza utilizando las áreas de la curva Normal) y el valor de **e** (máximo error permitido por el investigador).



1.5 Escalas de medición de los datos

Las escalas a considerar son :

- Nivel nominal
- Nivel ordinal
- Nivel de intervalos
- Nivel de razón
-

Nivel Nominal

Escala cualitativa que asigna arbitrariamente un número a cada respuesta de modo que sólo tenga valor como un número de identificación. El número de escala no tiene ningún significado por sí sólo.

Ejemplo : ¿Cuál es la tarjeta de crédito de su preferencia ?

1. VISA ()
2. Mastercard ()
3. American Express ()
4. Diners ()

Estadística permisible :

{ Porcentajes
Moda
Prueba binomial – Ji cuadrado

El número que se asigna en esta escala no representa magnitudes absolutas. Solo sirven para clasificarlos en determinada categoría, en otras palabras 1, no es la mitad de 2 .

Nivel Ordinal

Escala cualitativa que no sólo clasifica , sino establece jerarquías entre los valores. Entre mayor sea el número, mayor (o menor) es la existencia del atributo , pero sin indicar la distancia que hay entre las posiciones , es decir que el número cuatro en preferencia no es 300% superior al número 1, solo indica que es preferido respecto del anterior

Ejemplo : Clasifique en una escala de 1 a 4 las siguientes marcas de gaseosa, en función de su preferencia :

1. Inca Kola ()
2. Coca Cola ()
3. Real Kola ()
4. Pepsi Cola ()

Estadística permisible

{ Percentiles - mediana
Desviación cuartil
Correlación rango-orden

Nivel de intervalo

Escala cuantitativa que clasifica, ordena y establece distancias o intervalos iguales entre las unidades de medida. Asigna un punto de cero en forma arbitraria por convención por los expertos, pero que no implica la ausencia del atributo. Por ejemplo una prueba de coeficiente de inteligencia va tener un punto cero, pero no hay una persona con cero de inteligencia. Otros ejemplos son la medición del calendario, o la medición de la temperatura

Ejemplo : Resultados económicos de empresas de un sector (\$)

de	a
20'000	-10'000
-10'000	0
0	-10'000
10'000	20'000

Estadística permisible

{ Media –Mediana-Moda
Desviación estándar- Varianza
Coeficientes de Correlación
Prueba T – Prueba Z

Nivel de Razón

Escala cuantitativa es igual que las escalas de intervalos, pero poseen un cero absoluto. (origen natural) en el cual hay una ausencia de la propiedad o atributo, ejemplo el peso o los ingresos monetarios de una persona

Ejemplo : Nivel de ingresos mensuales de las familias de un distrito (en soles)

De	A
0	999
1,000	1,999
2,000	2,999
3,999	4,000

Estadística permisible

{ Media geométrica
Media armónica
Coeficiente de variación

La estadística permisible va en sentido acumulativo, así en la escala de razón se pueden estudiar todos los indicadores anteriores a las escalas de intervalos, ordinales y nominales

2. CONSTRUCCION DE TABLAS DE FRECUENCIA

Una primera aproximación al análisis descriptivo es la construcción de la tabla de frecuencias , las cuales presentan la distribución de un conjunto de elementos de acuerdo a las categorías de una variable x .

En la tabla se observa la frecuencia o repetición de cada uno de los valores en el correspondiente intervalo de clase

Se presentan los siguientes casos :

2.1 Variable discreta , es aquella cuyo valor se expresa únicamente por números enteros, adquieren valores absolutos y por lo general son cualitativas.

Ejemplo 1 :

En una muestra de veinte bodegas del distrito X ,se desea conocer la cantidad de marcas de crema dental que ofrecen a sus clientes.

La variable (x_i) es el número de marcas de crema dental ofrecidas.

Hecho el estudio se obtuvieron los siguientes resultados :

Bodega N°	x_i	Bodega N°	x_i
1	6	11	4
2	5	12	5
3	4	13	6
4	4	14	2
5	3	15	4
6	3	16	3
7	4	17	4
8	4	18	6
9	5	19	5
10	6	20	3

N : 20 bodegas

Construyendo la tabla tendríamos :

1° Clasificación : x_i máximo : 6 marcas de crema dental
 x_i mínimo : 2 marcas de crema dental

2° Las clases serían : 2,3,4,5,y 6

3° Tabulación : Se determina cuantas veces de repite cada valor de x_i (frecuencia).
Se denomina frecuencia absoluta (f_i) cuando se contabiliza en valores absolutos (número de bodegas)

Se denomina frecuencia relativa (h_i) cuando se contabiliza en valores relativos (porcentajes)

4° El cuadro de frecuencias quedaría presentado de la siguiente manera :

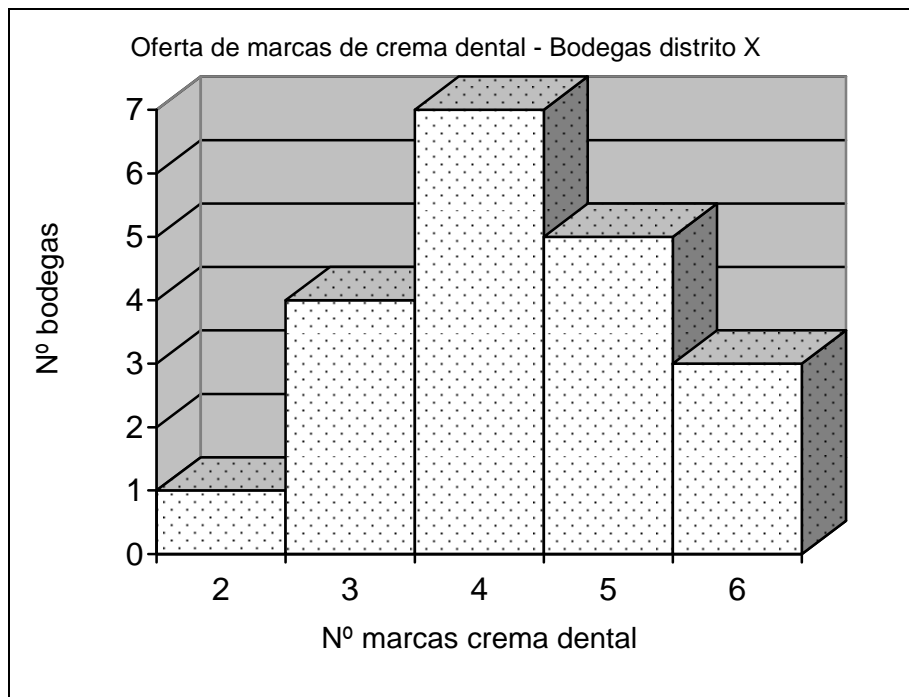
Cuadro N° 1 :

Distribución de 20 bodegas del distrito X en función al número de marcas de crema dental que ofrecen a sus clientes

x_i	tabulación	f_i	h_i (%)
2	/	1	5
3	////	4	20
4	////////	7	35
5	//////	5	25
6	////	3	15
TOTALES		20	100

Fuente : encuesta área de mercadeo

Gráficamente tenemos :



Este gráfico se conoce como Histograma

2.2 Variable continua, es aquella que puede tomar cualquier valor del conjunto de los números racionales (enteros o fraccionarios). Son variables cuantitativas

Ejemplo 2 :

Se desea conocer el ingreso mensual promedio del asentamiento “Galápagos” .

La variable x_i : será ingresos mensuales expresados en Soles

Para tal efecto se ha seleccionado una muestra de 50 familias y se han obtenido los siguientes datos :

<i>Familia Nº</i>	x_i	<i>Familia Nº</i>	x_i
1	730	26	500
2	750	27	870
3	580	28	550
4	430	29	710
5	490	30	750
6	650	31	700
7	670	32	400
8	750	33	610
9	510	34	750
10	970	35	690
11	820	36	540
12	650	37	720
13	890	38	780
14	590	39	850
15	550	40	350
16	700	41	320
17	600	42	830
18	700	43	890
19	380	44	650
20	600	45	450
21	450	46	750
22	750	47	640
23	730	48	930
24	650	49	850
25	760	50	630

n : 50

1º Se calcula el Rango (R)

R = mayor valor de x_i – menor valor de x_i

$$R = x_{10} - x_{41}$$

$$R = 970 - 320 = 650$$

2º Se obtiene el numero de clases e intervalos – para tal efecto se utiliza la Regla de Sturges :

<i>Si el tamaño de la muestra es</i>	<i>Regla de Sturges</i>
<i>Menor de 100</i>	$m = 1 + 3.322 \times \text{Log } n$
<i>Mayor de 100</i>	$m = 3 + 3.322 \times \text{Log } n$

En este caso $n < 100$ entonces :

$$m = 1 + 3.322 \times \text{Log } 50$$

$$m = 1 + 3.322 \times 1.69897$$

$$m = 6.64 \approx 7 \text{ intervalos}$$

3° El tamaño de clase (c) sería : $C = R / m$

$$c = 650 / 7 = 92.8 \approx 93$$

4° Construimos la tabla de frecuencias :

i	Intervalo de clase		Marca de clase - x_i	Frecuencias absolutas		Frecuencias relativas	
	x_i	x_s		f_i	F_i	h_i	H_i
1	320	413	366.5	4	4	8	8
2	413	506	459.5	5	9	10	18
3	506	599	552.5	6	15	12	30
4	599	692	645.5	11	26	22	52
5	692	785	738.5	15	41	30	82
6	785	878	831.5	5	46	10	92
7	878	971	924.5	4	50	8	100
TOTALES				50		100	

Donde :

i : número de intervalo

x_i : intervalo de clase inferior

x_s : intervalo de clase superior

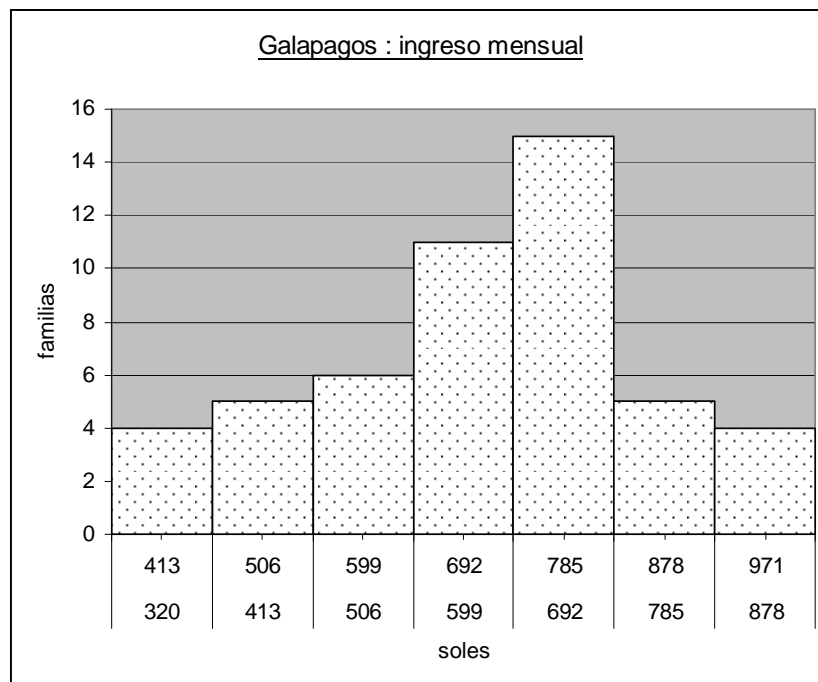
f_i : muestran la repetición de los datos en determinado intervalo de clase- in valores absolutos (familias)

F_i : muestran la acumulación progresiva de las frec.absolutas

h_i : expresan a las frec. absolutas en términos relativos (%)

H_i : muestran la acumulación progresiva de las frec. Relativas

Gráficamente vamos a elaborar el histograma de frecuencias absolutas (f_i) :



Es importante acotar que los gráficos deben de tener las siguientes condiciones básicas :

- Título : descripción abreviada del contenido
- Leyendas y cifras tanto en el eje de las abscisas como de las ordenadas
- Debe ser simétrico, no muy horizontal o vertical
- En el eje de las abscisas se colocan los valores de la variable x
- En el eje de las ordenadas se colocan las frecuencias (f_i, h_i)
- De ser posible se colocan las fuentes de la información

En relación al número apropiado de los intervalos Christensen Howard ¹ plantea lo siguiente :

Número de valores en el conjunto	Número apropiado de intervalos de clase
De 10 a 100	De 4 a 8
De 100 a 1,000	De 8 a 11
De 1,000 a 10,000	De 11 a 14

Avila Acosta ² en cambio define los intervalos de clase en tres categorías :

a. Intervalos de igual amplitud

Alumnos por aula
20 - 29
30 - 39
40 - 49
50 - 69
70 - 69

b. Intervalos de diferente amplitud

Edad de clientes
3 - 5
6- 14
25 - 24
25 - 39

c. Intervalos abiertos por la izquierda y cerrados por la derecha

Sueldos mensuales (]
320 - 370
370 - 420
420 - 470
470 - 520
520 - 570

En este caso NO está incluido el extremo inferior, pero si el extremo superior

¹ CHRISTENSEN Howard. *Estadística Paso a Paso*

² AVILA Acosta . *Estadística Elemental*

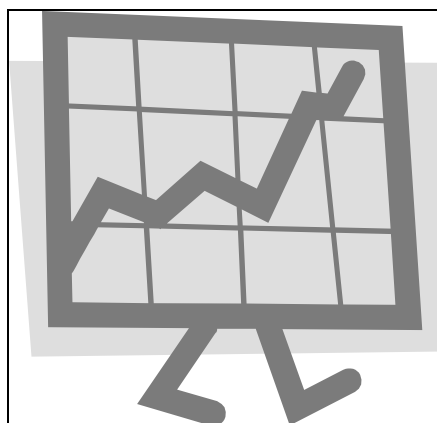
Otra forma de presentar este cuadro es :

Alumnos por aula
320.01 – 370
370.01 – 420
420.01 – 470
470.01 – 520
520.01 - 570

Esta segunda forma de presentación es más práctica ,pues indica directamente los valores comprendidos en cada intervalo.

Por lo general el número de intervalos de clase va depender de :

- La naturaleza de la variable
- El número de valores observados
- El recorrido de la variable
- Los objetivos del estudio

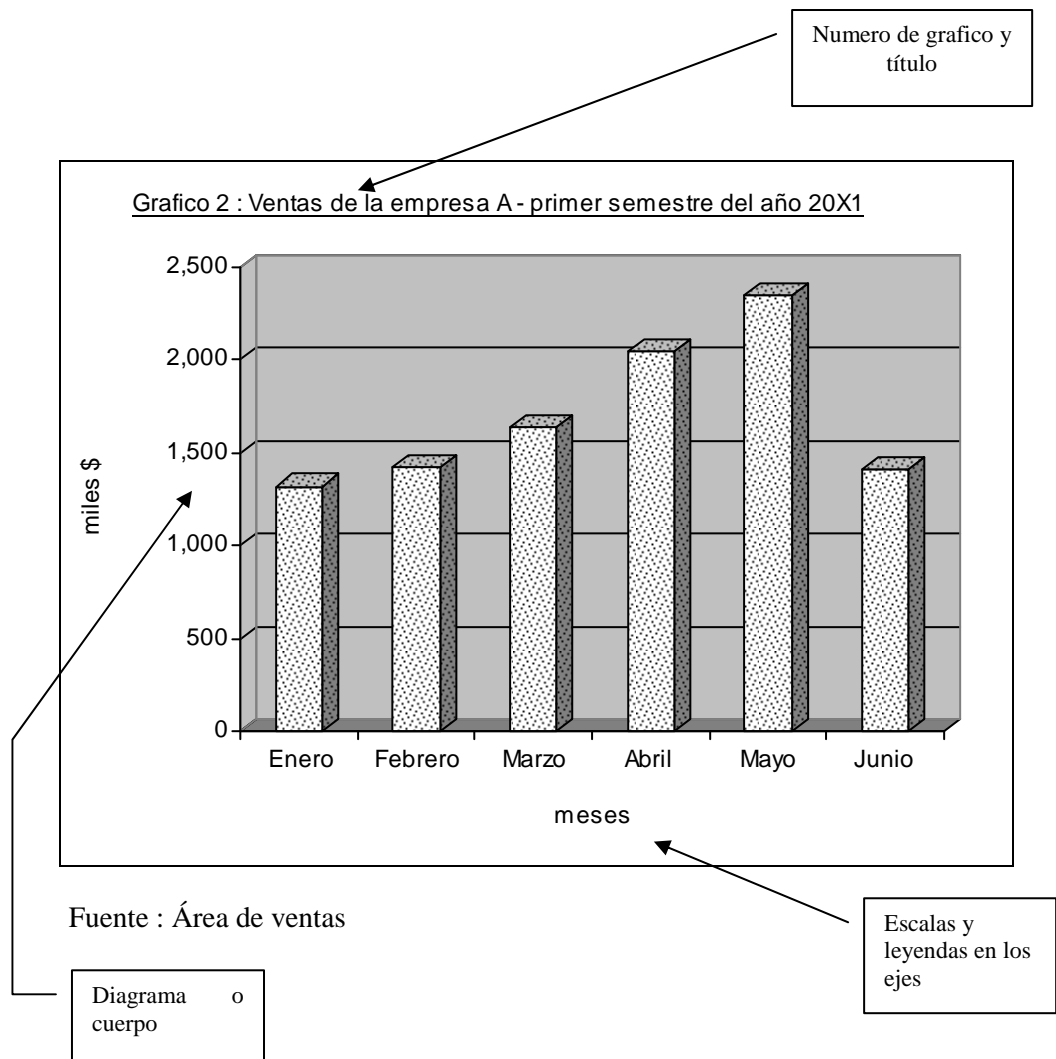


3. LOS GRÁFICOS ESTADÍSTICOS

3.1 Concepto

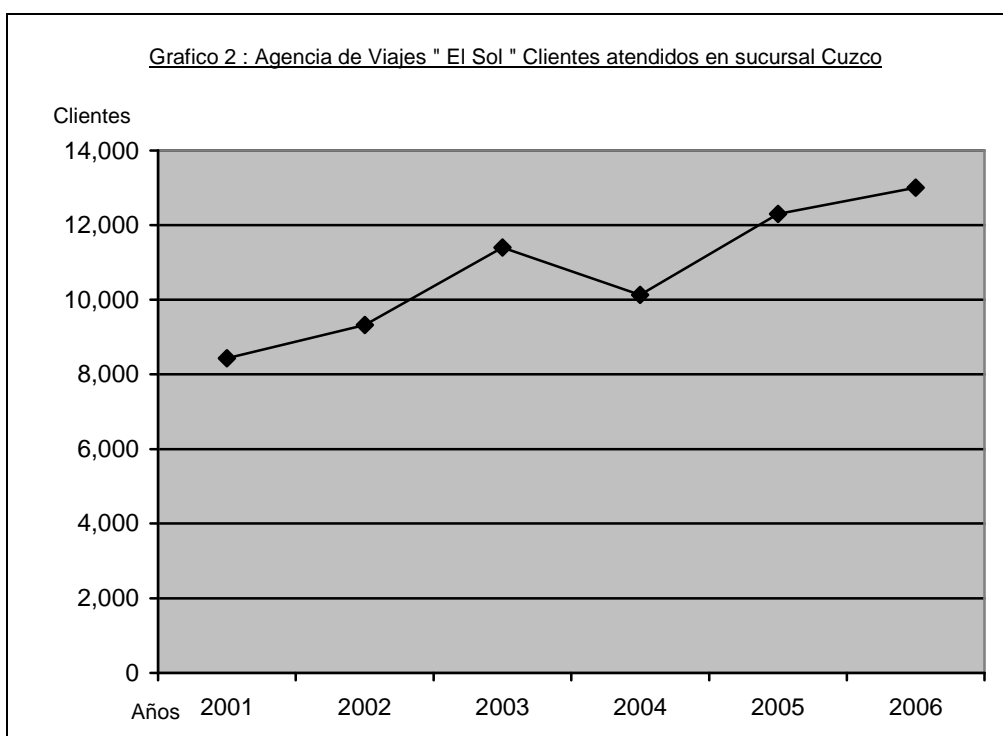
Son representaciones pictóricas (figuras geométricas o de superficie) utilizados con el objeto de mostrar magnitudes , cambios de una variable o comparar dos o más variables relacionadas.

Un gráfico bien elaborado debe tener los siguientes elementos :



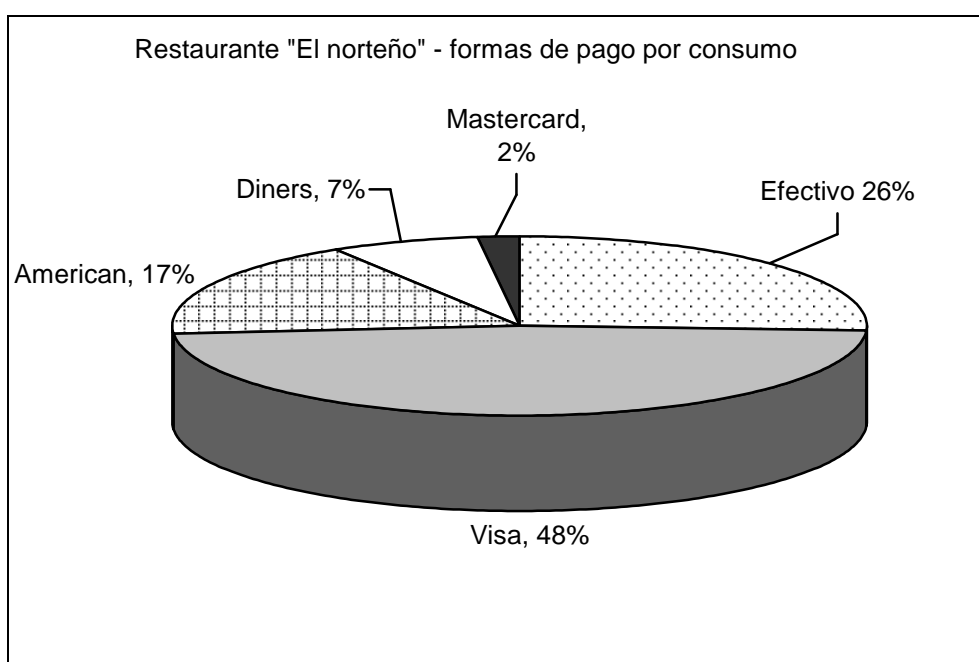
3.2 Tipos de Gráficos

3.2.1 Gráfico Lineal



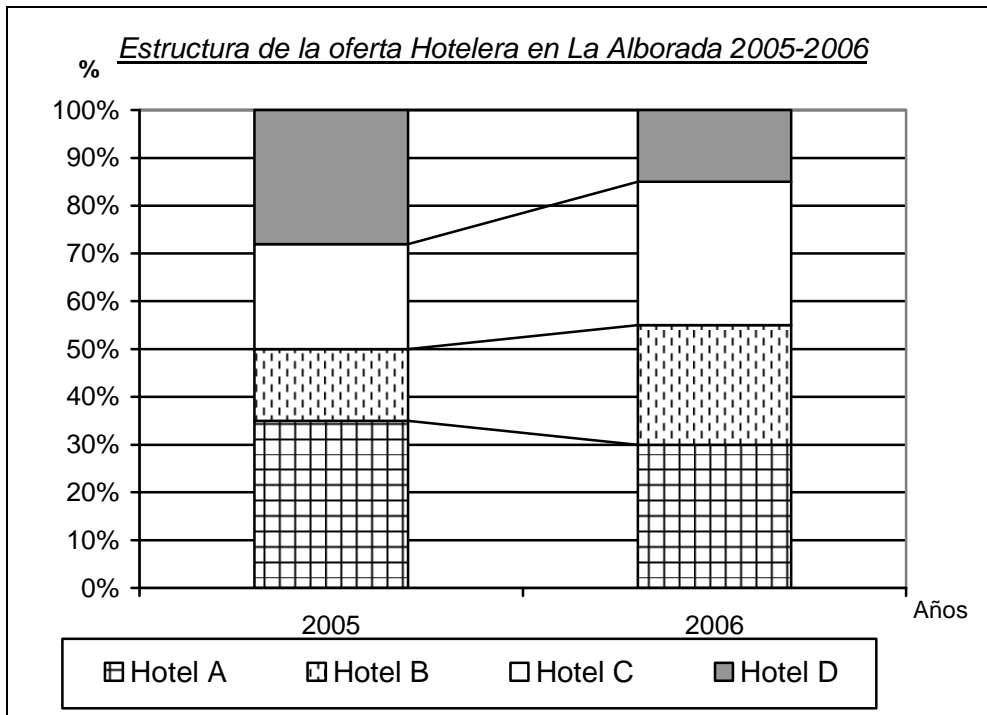
Este gráfico de evolución es útil para representar la evolución de una sola variable en el tiempo (serie de tiempo)

3.2.2 Gráfico circular



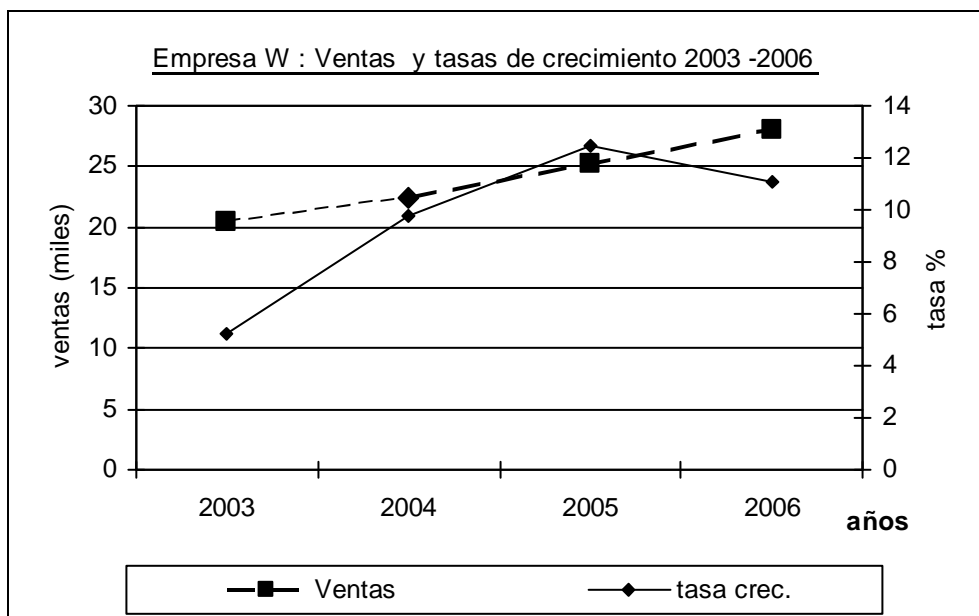
Es utilizado para expresar una variable que esta compuesta de varios subconjuntos es decir es un gráfico de estructura (de una sola variable)

3.2.3 Gráfico de barras comparativas



Se recomienda para comparar estructuras con varios subconjuntos en más de un período de tiempo

3.2.4 Gráfico Combinado (valores en dos abscisas)



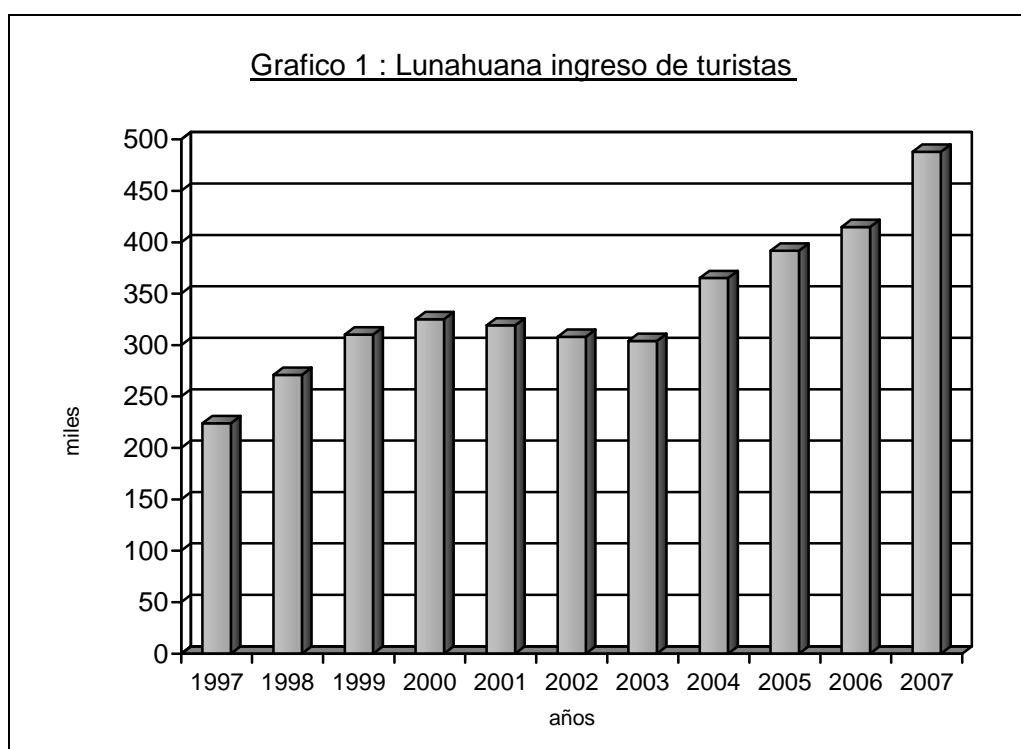
Es muy útil para mostrar dos frecuencias que tienen valores diferentes (dólares y porcentajes por ejemplo); cada una de ellas se ubica en uno de los ejes verticales y su lectura es a través de los valores allí expresados

EJEMPLO PRÁCTICO DE ELABORACIÓN DE UN GRÁFICO ESTADÍSTICO CON EXCEL

Supongamos que estamos estudiando la cantidad de turistas que visitan el valle de Lunahuana y se tienen los siguientes datos estadísticos : (datos supuestos)

1	2	3
Año	Turistas (miles)	Crecim (%)
1997	224	17.28
1998	271	20.98
1999	310	14.39
2000	325	4.84
2001	319	-1.85
2002	308	-3.45
2003	304	-1.30
2004	365	20.07
2005	392	7.40
2006	415	5.87
2007	488	17.59

En primer lugar vamos a utilizar un gráfico de barras para representar las cifras .

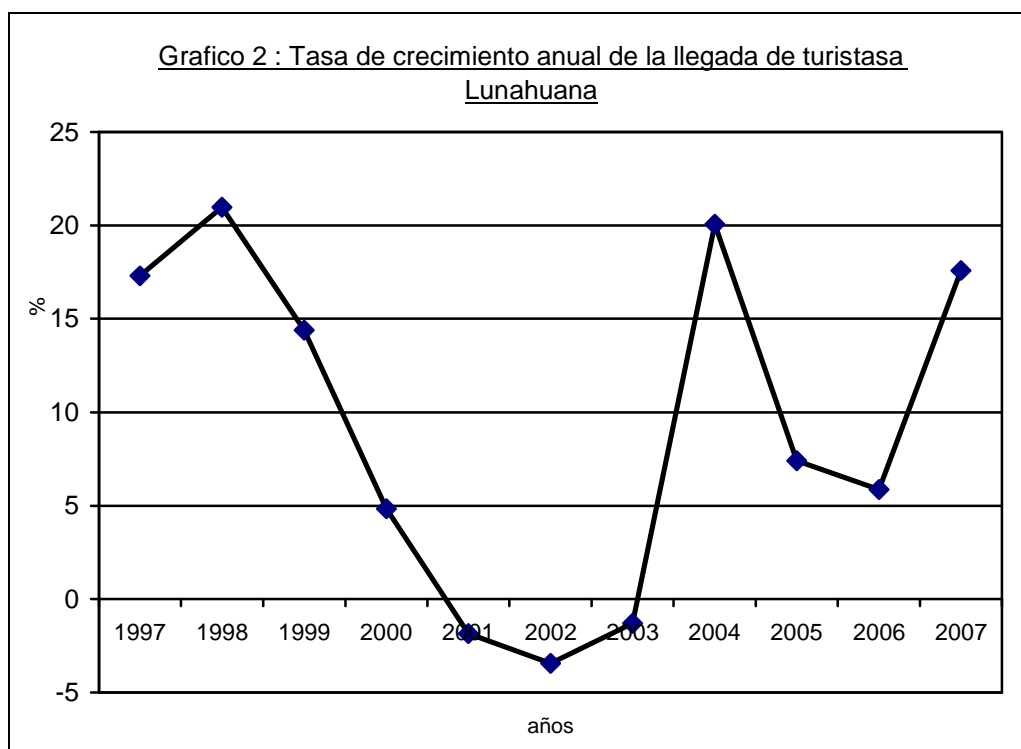


Apreciamos que hay tres etapas claramente definidas:

- 1° De 1997 al año 2000 se aprecia un crecimiento
- 2° De 2001 al año 2003 vemos que se estanca el impulso inicial y cae el número de turistas y visitantes
- 3° A partir de del año 2004 se vuelva a tener un impulsote crecimiento mucho mayor

Estos gráficos tienen la ventaja que nos permiten ver el comportamiento de un variable en un largo lapso de tiempo y podemos apreciar la tendencia de largo plazo (mas de cinco años) que en este caso es de crecimiento.

Pero es importante acompañar el análisis con otro tipo de gráficos, en este caso usaremos el **gráfico lineal** para apreciar cómo es el crecimiento relativo (en porcentajes) de cada año, conforme se aprecia en la columna 3 del cuadro y cuya expresión gráfica es :



Aquí apreciamos más claramente la situación y se pueden graficar los valores negativos como son los años 2001, 2002, 2003 y 2004 , en los cuales no hubo crecimiento sino todo lo contrario se experimento una reducción en la cantidad de visitantes a la localidad

El calculo del crecimiento se efectuó con la fórmula :

$$cresc. = \left(\frac{V_n}{V_{n-1}} - 1 \right) \times 100$$

Donde :

V_n : valor de la variable el año “n”

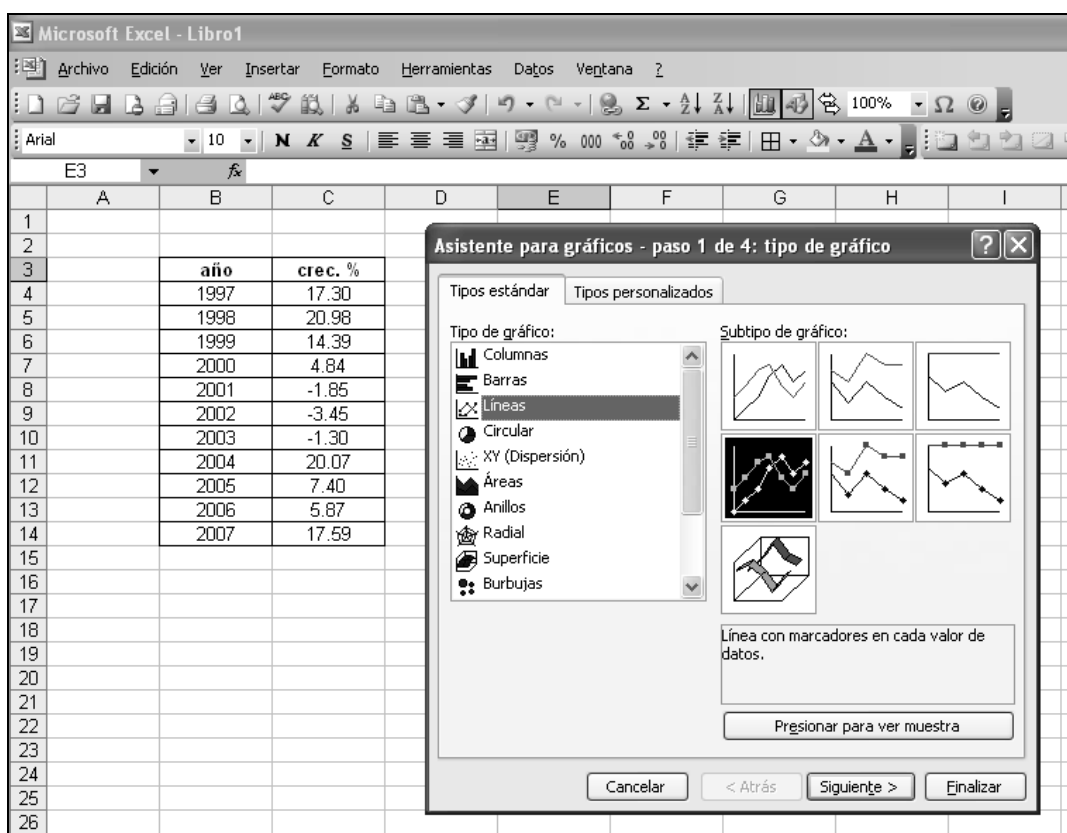
V_{n-1} : Valor de la variable el año “n-1” (año anterior)

Por ejemplo el valor del año 2000 se obtuvo de la siguiente manera :

$$cresc. = \left(\frac{325}{310} - 1 \right) \times 100$$

$$cresc. = 4.84 \%$$

Una dificultad evidente es la elaboración de gráficos con la hoja electrónica EXCEL , vamos a presentar los pasos a continuación con los datos del gráfico 2 :

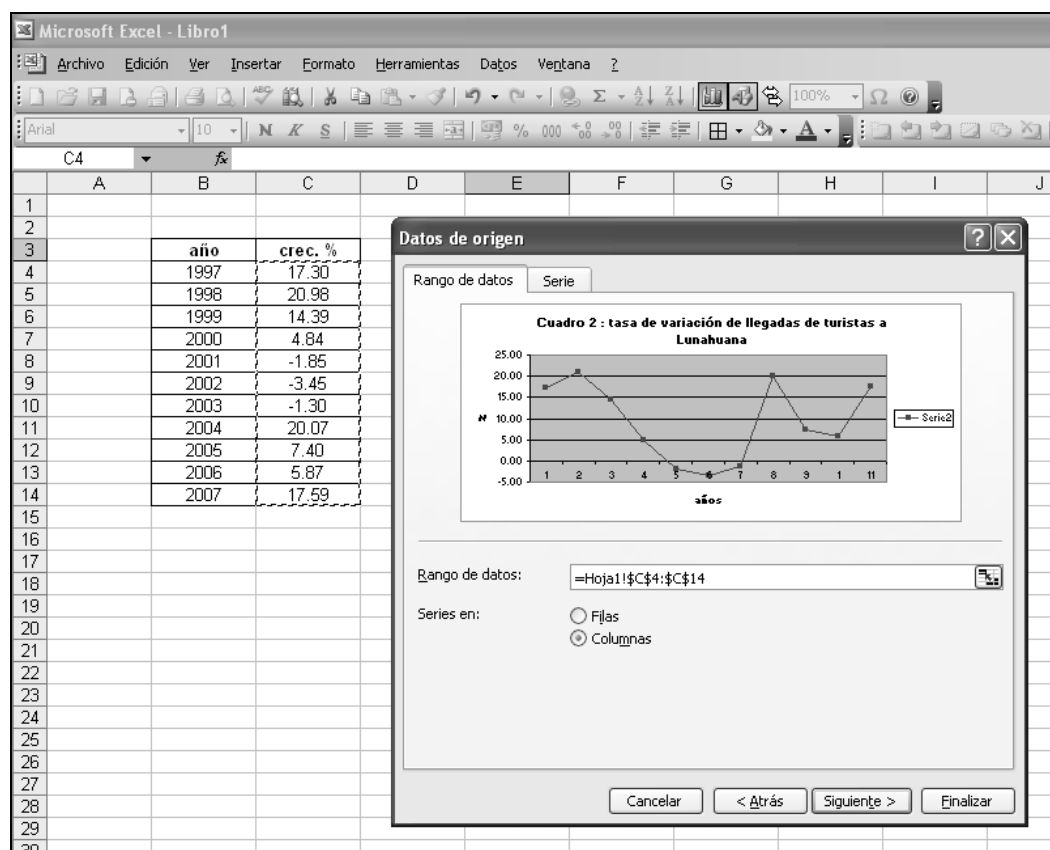


1º paso : ingresamos los datos de los años y las tasas de crecimiento

2º paso : accionamos el icono de gráficos y vamos a tener el asistente para gráficos

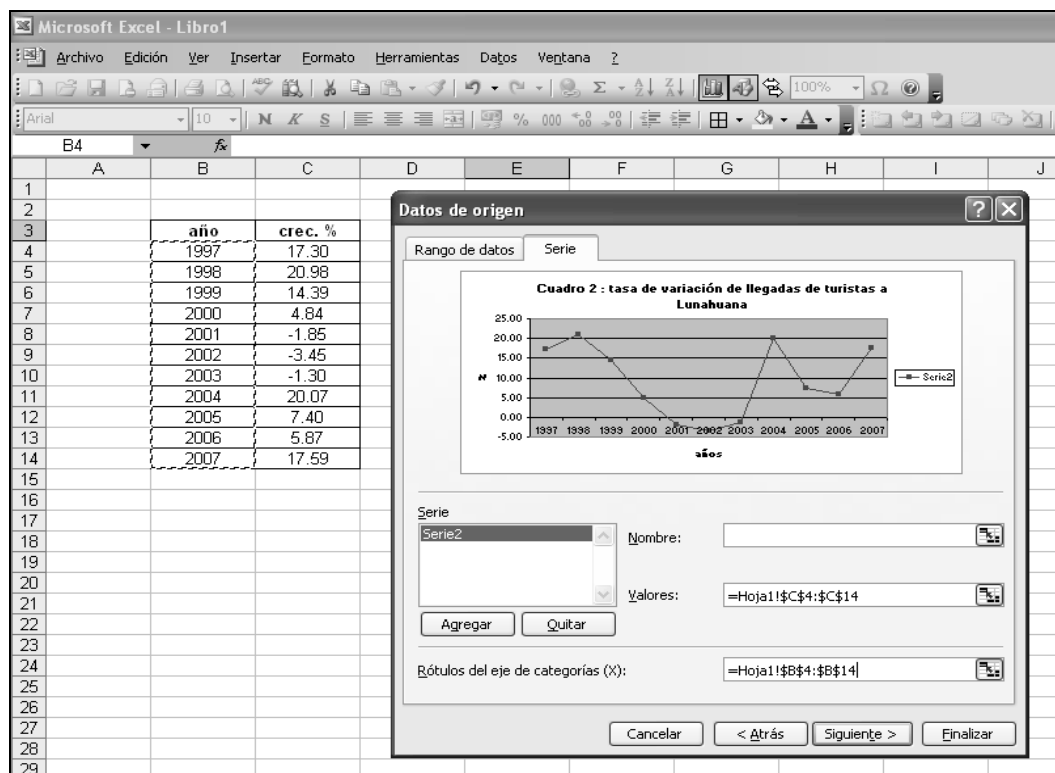
3º paso : elegimos la opción de gráfico lineal- **Líneas**

4º paso : presionamos el comando de **Siguiente >**

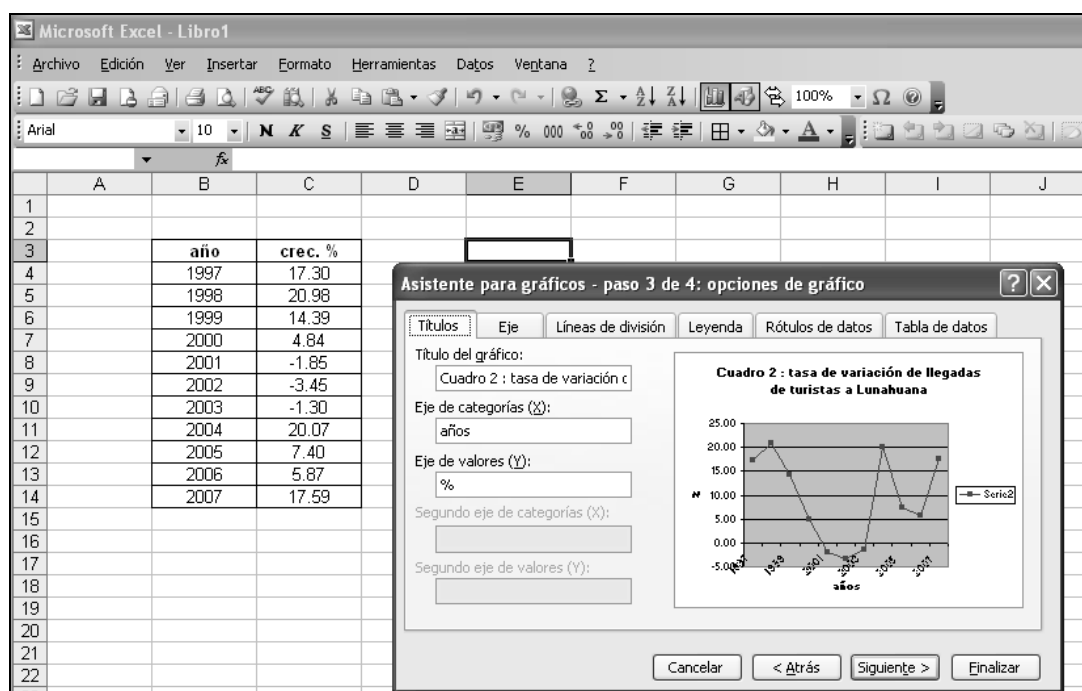


5º paso : Ingresamos el rango de datos : C4:C14, aparece la gráfica de las tasas de crecimiento

6º paso : Se acciona el comando de series para colocar los datos de los años



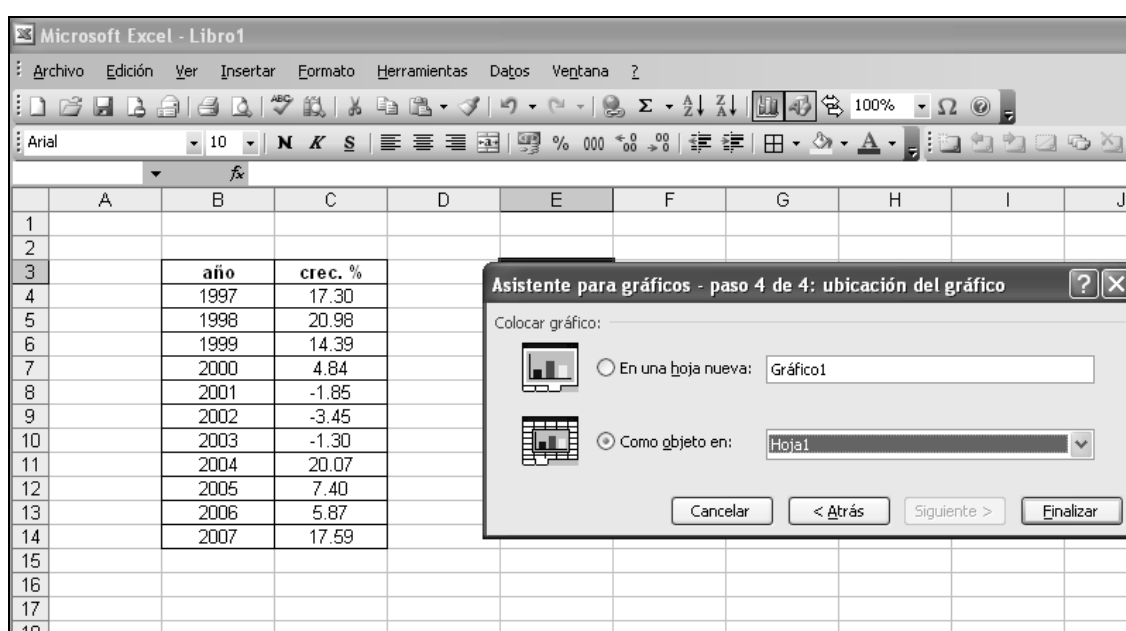
7º paso : Se coloca el rango de los periodos de tiempo B4; B14 y se acciona el comando siguiente



8º paso : Se coloca :

- Título del gráfico : **tasa de variación anual de llegada de turistas a Lunahuana**
- Eje de categorías : **años**
- Eje de valores : **%**

9º paso : Se presiona siguiente y se tiene concluido el gráfico indicando **Finalizar**



4. ANALISIS DE TENDENCIA CENTRAL

Las medidas de tendencia central o de resumen son indicadores que tienden a sintetizar o describir de la manera más representativa las características de un conjunto de datos. Las medidas más importantes son :

- La Media aritmética
- La Mediana
- La Moda

4.1 La Media Aritmética (\bar{x})

La media aritmética es la clase que determina el centro de gravedad de un conjunto de datos, es decir es el valor más representativo

a) Media aritmética de datos no agrupados :

Formula :

$$\bar{x} = \frac{\sum_{i=1}^n xi}{n}$$

Donde :

x_i : clase

n : número de clases

Σ : Sumatoria (desde $i = 1$, hasta $i = n$)

Ejemplo 3 :

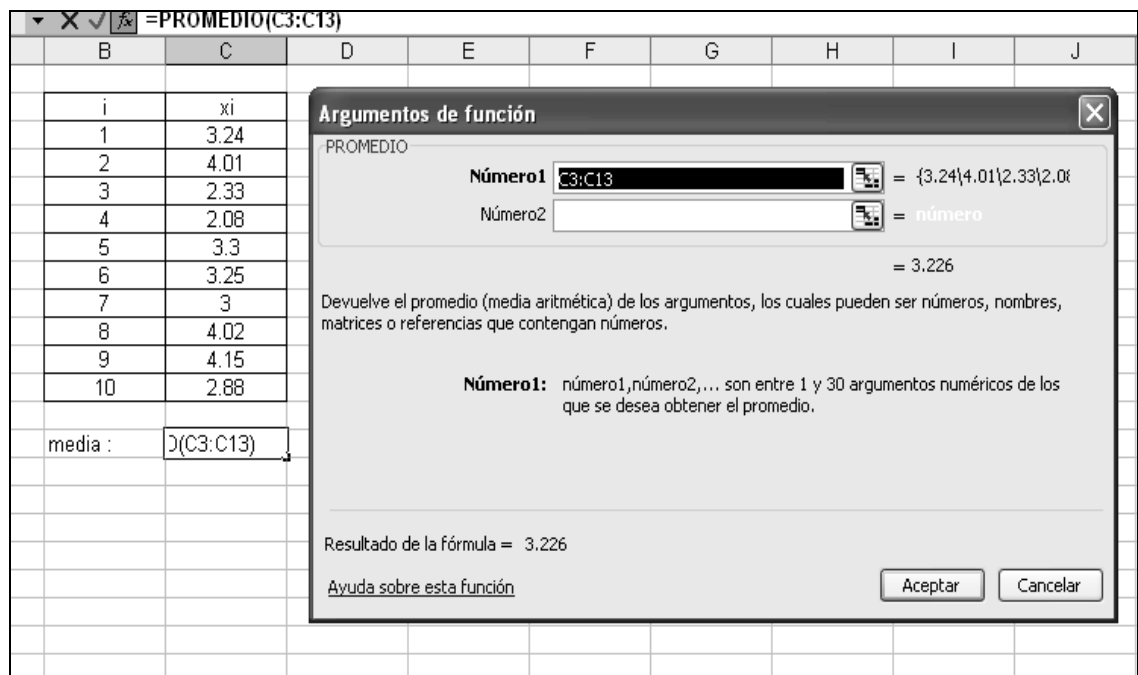
Se ha efectuado la medición de cuanto demora la atención a los clientes en un Supermercado. Se ha tomado una muestra de 10 clientes y los resultados obtenidos son :

Cliente x_i	Duración de la atención
1	3.24
2	4.01
3	2.33
4	2.08
5	3.30
6	3.25
7	3.00
8	4.02
9	4.15
10	2.88
Σ	32.26

$$\bar{x} = \frac{32.26}{10}$$

$\bar{x} = 3.23$ minutos , que es el promedio de duración de la atención a los clientes

Utilizando Excel el procedimiento es el que sigue :



b) Media aritmética de datos agrupados

$$\bar{x} = \frac{\sum_{i=1}^n x_i \times f_i}{N}$$

Donde :

x_i : marca de clase

f_i : frecuencia absoluta

n : total de frecuencias

Ejemplo 4 :

La gerencia de mercadeo de un Hotel ha decidido estudiar un estudio acerca de la edad promedio de los clientes del Café Bar “ El Sol ”. Se ha elegido una muestra de 300 clientes recogida durante todo un mes típico . Aplicada la encuesta se han obtenido los siguientes resultados :

Clase (i)	Intervalo	x_i	f_i	$x_i \times f_i$
1	19 – 23	21	5	105
2	23 – 27	25	9	225
3	27 – 31	29	13	377
4	31 – 35	33	48	1,584
5	35 – 39	37	67	2,479
6	39 - 43	41	58	2,378
7	43 – 47	45	54	2,430
8	47- 51	49	29	1,421
9	51- 55	53	17	901
TOTALES			300	11,900

La media aritmética es igual a :

$$\bar{x} = \frac{11,900}{300}$$

$$\bar{x} = 39.67 \text{ años}$$

El promedio de edad de los clientes del Café Bar “ El Sol “ es de 39. años y medio

LA MEDIA ARITMÉTICA : RESUMEN	
CARACTERISTICAS	<ul style="list-style-type: none"> En su valor influyen todos los componentes de la distribución Puede ser manipulada algebraicamente
VENTAJAS	<ul style="list-style-type: none"> Es la medida más fácil de calcular Es la medida más conocida y utilizada
DESVENTAJAS	<ul style="list-style-type: none"> Su valor puede ser distorsionado por los valores extremos o singulares

4.2 La Mediana (Me)

Es la medida de tendencia central que corresponde al valor de la variable que divide a la frecuencia total en dos partes iguales .

a) Mediana de datos no agrupados

En este caso se procede de la siguiente manera :

- 1° Se ordena el conjunto de valores en orden creciente
- 2° Se halla el valor que ocupa la posición media
- 3° Si el número es impar, el valor central es la mediana
- 4° Si el número es par , el promedio de los dos centrales es la mediana

Ejemplo 5 :

Se tiene el siguiente conjunto de datos :

4	8	5	3	9	7	2
---	---	---	---	---	---	---

Se ordena

2	3	4	5	7	8	9
3			Me	3		

Ejemplo 6 :

Se tiene el siguiente conjunto de datos :

6	8	9	10	11	15
---	---	---	----	----	----

Se ordena

6	8	9	9.5	10	11	15
3			Me = (9+10) / 2 = 9.5			

3.2 Mediana de datos agrupados

Formula :

$$Me = Li + c \times \left(\frac{N/2 - Fa}{f_i} \right)$$

Donde :

Li : limite inferior del intervalo de la clase que contiene a la Me

c : Tamaño del intervalo de clase

n : Total de frecuencias absolutas

Fa : Frecuencia absoluta acumulada **anterior** al la clase que contiene a la Me

fi : frecuencia absoluta de la clase que contiene a la Me

Utilizando el ejercicio desarrollado en el ejemplo N° 4 tenemos :

Clase (i)	Intervalo	f _i	F _i
1	19 – 23	5	5
2	23 – 27	9	14
3	27 – 31	13	27
4	31 – 35	48	75
5	35 – 39	67	142
6	39 - 43	58	200
7	43 – 47	54	254
8	47- 51	29	283
9	51- 55	17	300
TOTALES		300	

El valor de N/2 es = 300/2 = 150, este valor se encuentra ubicado en el 6° intervalo

$$Me = 39 + 4 \times \left(\frac{(300/2) - 142}{58} \right)$$

$$Me = 39 + 0.55$$

$$Me = 39.55 \text{ años}$$

El 50% de los asistentes al Café Bar “ El Sol ” está en el intervalo de 19 a 39.55 años y el 50% restante está en el intervalo de 39.55 a 55 años.

50%	50%
19	55

39.55 años

LA MEDIANA : RESUMEN	
CARACTERISTICAS	<ul style="list-style-type: none"> • Es un promedio de posición • Cuando la agrupación de datos es muy estrecha es el mejor indicador
VENTAJAS	<ul style="list-style-type: none"> • Calculo relativamente fácil de efectuar • No es distorsionada por los valores extremos
DESVENTAJAS	<ul style="list-style-type: none"> • Su interpretación es bastante restringida • No se manejar algebraicamente, la mediana de varios subconjuntos no puede ser promediada para obtener la mediana del total • No es muy conocida ni entendida

4.3 La Moda (Mo)

Es la medida de tendencia central que corresponde al valor de la clase cuya frecuencia es la que más repite (f_i mayor)

No se puede calcular la Moda en datos no agrupados

a. Moda de datos agrupados

Formula :

$$Mo = Li + c \left(\frac{fp}{fp + fa} \right)$$

Donde :

Li : limite inferior del intervalo de la clase que contiene a la Moda

c : Tamaño del intervalo de clase

n : Total de frecuencias absolutas

fp : Frecuencia absoluta **posterior** a la clase que contiene a la Moda

fa : frecuencia absoluta **anterior** de la clase que contiene a la Moda

Utilizando el ejercicio desarrollado en el ejemplo N° 4 tenemos :

Clase (i)	Intervalo	f_i
1	19 – 23	5
2	23 – 27	9
3	27 – 31	13
4	31 – 35	48
5	35 – 39	67
6	39 - 43	58
7	43 – 47	54
8	47- 51	29
9	51- 55	17
TOTALES		300

La frecuencia mayor se encuentra ubicada en el 5° intervalo = 67 clientes

$$Mo = 35 + 4 \times \left(\frac{58}{58 + 48} \right)$$

$$Mo = 35 + 2.19$$

$$Mo = 37.19 \text{ años}$$

La edad más frecuente de los asistentes al Café Bar “ El Sol ” es de 37.19 años.

LA MODA : RESUMEN	
CARACTERISTICAS	<ul style="list-style-type: none"> Es absolutamente independiente de valores extremos Es un valor típico
VENTAJAS	<ul style="list-style-type: none"> Es la medida más descriptiva Cuando el número de valores es pequeño es fácil determinarla por observación
DESVENTAJAS	<ul style="list-style-type: none"> No es posible calcularla en caso de datos no agrupados

Relación empírica entre Media, Mediana y Moda :

DISTRIBUCIONES	Relación
SIMETRICAS	$\bar{x} = Me = Mo$
ASIMETRICAS A LA DERECHA	$Mo > Me > \bar{x}$
ASIMETRICAS A LA IZQUIERDA	$Mo < Me < \bar{x}$

Con los datos del ejercicio 4 :

\bar{x} : 39.67 años

Me : 39.55 años

Mo : 37.19 años

Asimetría a la izquierda

La asimetría también se puede calcular de la siguiente ,manera :

$$As = \frac{(\bar{X} - Mo)}{s}$$

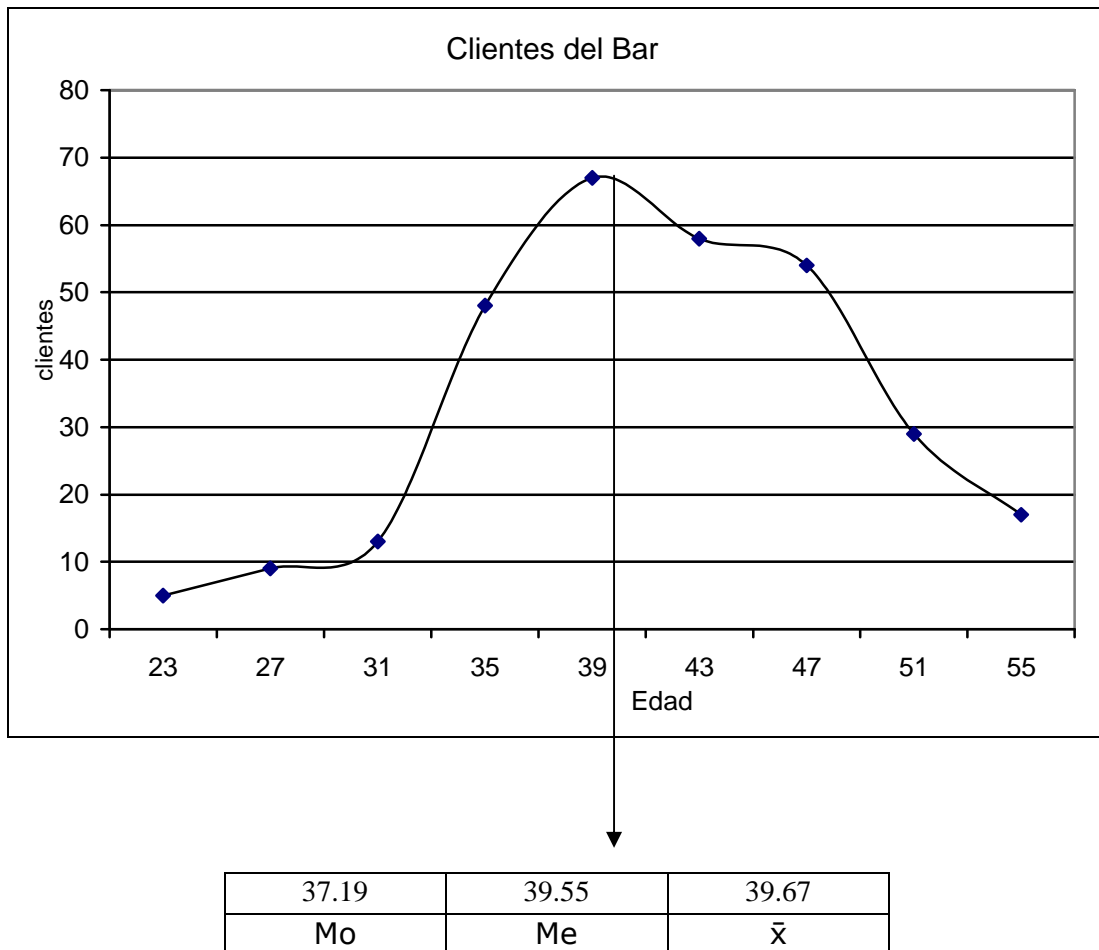
Los resultados obtenidos se pueden clasificar de la siguiente manera :

AS > 0	As = 0	As < 0
Asimetría positiva Sesgo hacia la izquierda Cola hacia la derecha	Simetría	Asimetría negativa Sesgo hacia la derecha Cola hacia la izquierda

Utilizando los datos del ejemplo tenemos :

$$As = \frac{(39.64 - 37.19)}{7.12}$$

As = 0.017 que es una asimetría positiva o a la izquierda



5. ANÁLISIS DE DISPERSIÓN

Las medidas de dispersión son un conjunto de indicadores que nos expresan el grado de concentración o alejamiento de los datos respecto de la media aritmética.

Ejemplo 7 :

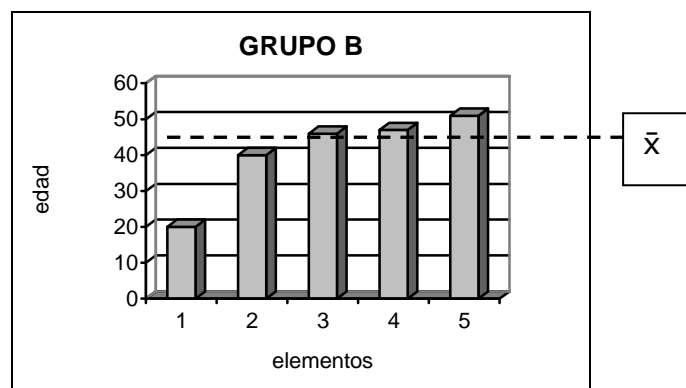
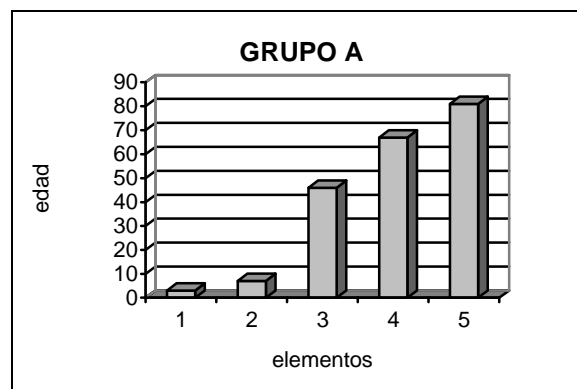
Tenemos las siguientes distribuciones de datos :

x_i	A	B
1	3	20
2	7	40
3	46	46
4	67	47
5	81	51

Hallamos la Media y la mediana :

Media	40.8	40.8
mediana	46	46

Aparentemente ambas distribuciones son iguales, pero ¿ esto es así? : veamos los Los gráficos :



A pesar que ambos grupos tienen los mismos indicadores de tendencia central , las distribuciones de los datos muestran que el grupo B es más homogéneo que el grupo A, pues los datos están más cerca del valor de la edad promedio (40.6 años) , en cambio el grupo A está más disperso o menos concentrado..

Para poder medir el grado de concentración o dispersión de los datos , respecto de la media aritmético se tienen las siguientes medidas de dispersión :

- El Rango - R
- La desviación media - DM
- La desviación estándar - s
- El coeficiente de variación – CV

5.1 El Rango (R)

Es la medida de dispersión que mide la amplitud o recorrido de la distribución y se obtiene de la siguiente manera :

$$R = \text{Mayor Valor} - \text{Menor Valor}$$

Utilizando el ejemplo anterior tenemos :

$$\text{Rango A} = 81 - 3 = 78$$

$$\text{Rango B} = 51 - 20 = 31$$

La distribución B tiene un ,menor recorrido que la distribución A

La utilización del Rango es muy limitada pues sólo considera los valores extremos y no indica como se dispersan los valores intermedios.

5.2 La Desviación Media (DM)

Es una medida de dispersión que es el promedio aritmético de las desviaciones de las clases respecto de la media aritmética

a) Desviación Media de datos no agrupados :

Formula :

$$DM = \frac{\sum_{i=1}^n |xi - \bar{x}|}{n}$$

Donde :

x_i : clase

\bar{x} : media aritmética

n : número de clases

Σ : Sumatoria (desde i = 1 , hasta i = n)

Utilizando los datos del ejemplo 7 tenemos :

x_i	A	B
1	3	20
2	7	40
3	46	46
4	67	47
5	81	51
X	40.6	40.6

La desviación media del primer grupo es :

$$DMA = \frac{|3 - 40.6| + |7 - 40.6| + |46 - 40.6| + |67 - 40.6| + |81 - 40.6|}{5}$$

$$DMA = 28.68$$

La desviación media del grupo B :

$$DMB = \frac{|20 - 40.6| + |40 - 40.6| + |46 - 40.6| + |47 - 40.6| + |51 - 40.6|}{5}$$

$$DMB = 8.68 \text{ años}$$

En otras palabras la dispersión del grupo B 2.3 veces menor que la del grupo A, por tanto este grupo **es más homogéneo o más concentrado**

b) Datos agrupados

Fórmula :

$$DM = \frac{\sum_{i=1}^n |x_i - \bar{x}| \times f_i}{n}$$

Donde :

x_i : clase

\bar{x} : media aritmética

n : número de frecuencias absolutas

f_i : frecuencia absoluta

$||$: Valor absoluto (la resta debe ser siempre positiva)

Utilizamos el ejemplo N° 4 - edad promedio de los clientes del Café Bar

" El Sol

Clase (i)	Intervalo	x_i	f_i	$ x_i - \bar{x} $	$ x_i - \bar{x} \times f_i$
1	19 – 23	21	5	18.67	93.35
2	23 – 27	25	9	14.67	132.03
3	27 – 31	29	13	10.67	138.71
4	31 – 35	33	48	6.67	320.16
5	35 – 39	37	67	2.67	178.89
6	39 - 43	41	58	1.33	77.14
7	43 – 47	45	54	5.33	287.82
8	47- 51	49	29	9.33	270.57
9	51- 55	53	17	13.33	226.61
Totales			300		1,625.28

Nota : La media aritmética es $\bar{x} = 39.67$ años

La desviación media sería :

$$DM = \frac{1,625.28}{300}$$

$$DM = 5.42 \text{ años}$$

El promedio de las desviaciones de los datos respecto a la media aritmética es de 5.42 años

5.3 La Desviación Estándar (s)

Es una medida de dispersión más utilizada y confiable es igualmente un promedio de las desviaciones de los datos pero elevados al cuadrado.

a) Desviación Estándar de datos no agrupados :

Formula :

$$s = \sqrt{\frac{\sum_{i=1}^n (xi - \bar{x})^2}{n}}$$

Donde :

x_i : clase

\bar{x} : media aritmética

n : número de clases

Σ : Sumatoria (desde $i = 1$, hasta $i = n$)

Utilizando los datos del ejemplo 7 tenemos :

x_i	A	B
1	3	20
2	7	40
3	46	46
4	67	47
5	81	51
\bar{x}	40.6	40.6

La desviación estándar del grupo A es :

$$s_A = \sqrt{\frac{(3 - 40.6)^2 + (7 - 40.6)^2 + (46 - 40.6)^2 + (67 - 40.6)^2 + (81 - 40.6)^2}{5}}$$

$$s_A = 31.31 \text{ años}$$

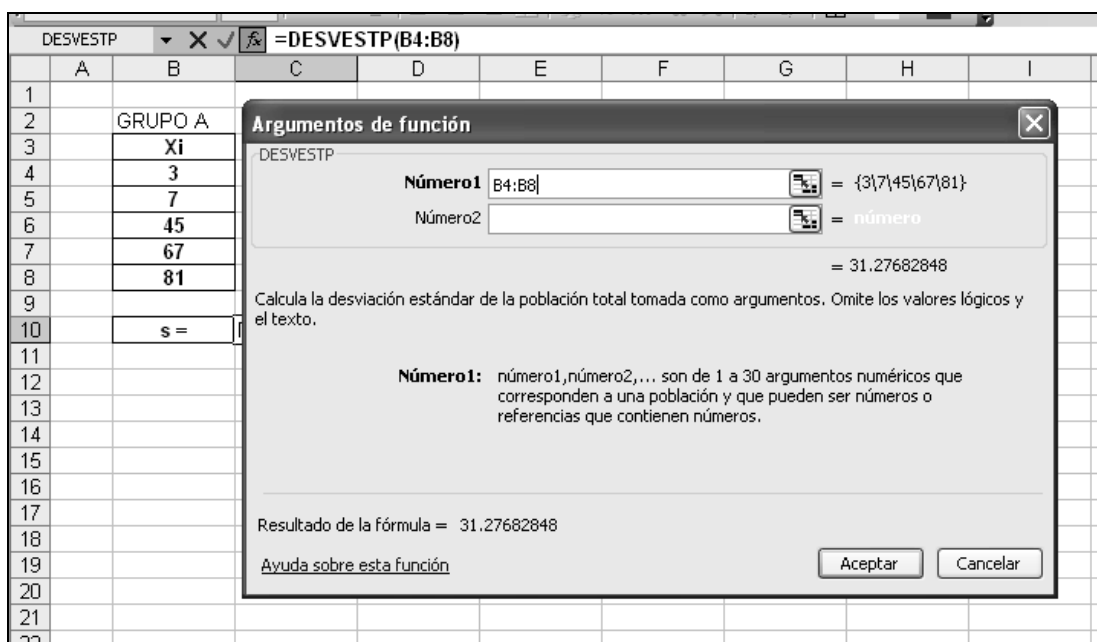
La desviación estándar del grupo B :

$$sB = \sqrt{\frac{(20 - 40.6)^2 + (40 - 40.6)^2 + (46 - 40.6)^2 + (47 - 40.6)^2 + (51 - 40.6)^2}{5}}$$

$$sB = 10.98 \text{ años}$$

Estos resultados ratifican los obtenidos con la desviación media, la diferencia es que son más exactos

Nota : cuando se trata de una muestra (n) en la fórmula se varía el denominador por $n-1$



b) Desviación estándar de Datos agrupados

Fórmula :

$$DM = \sqrt{\frac{\sum_{i=1}^n (xi - \bar{x})^2 \times fi}{n}}$$

Donde :

x_i : clase

\bar{x} : media aritmética

n : número de frecuencias absolutas

f_i : frecuencia absoluta

$\sqrt{}$: raíz cuadrada

Utilizamos el ejemplo N° 4 - edad promedio de los clientes del Café Bar "El Sol"

Clase (i)	Intervalo	x_i	f_i	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 \times f_i$
1	19 – 23	21	5	384.16	1920.80
2	23 – 27	25	9	243.36	2190.24
3	27 – 31	29	13	134.56	1749.28
4	31 – 35	33	48	57.76	2772.48
5	35 – 39	37	67	12.96	868.32
6	39 - 43	41	58	0.16	9.28
7	43 – 47	45	54	19.36	1045.44
8	47- 51	49	29	70.56	2046.24
9	51- 55	53	17	153.76	2613.92
			300		15,216.16

Nota : La media aritmética es $\bar{x} = 39.67$ años. Su desviación estándar es :

$$s = \sqrt{\frac{15,216.16}{300}}$$

$$s = 7.12 \text{ años}$$

El promedio de las desviaciones de los datos respecto a la media aritmética es de 7.12 años

5.4 El Coeficiente de variación (CV)

Es el indicador de dispersión que se expresa en valores independientes de la naturaleza de la variable.

Se utiliza para comparar dos o mas distribuciones cuando las unidades de medida de las variables están expresadas en diferentes unidades o escalas de medida .

Comparando dos o más distribuciones de datos , es más homogénea aquella que tiene el menor CV

Formula :

$$CV = \left(\frac{s}{\bar{x}} \right) \times 100$$

Con los datos del ejemplo N° 7 (edad de dos grupos de personas) tenemos :

indicador	A	B
s	31.3	10.98
\bar{x}	40.8	40.8

El CV seria :

	A	B
CV	$= \left(\frac{31.3}{40.8} \right) \times 100$ $= 76.72\%$	$= \left(\frac{10.98}{40.8} \right) \times 100$ $= 26.91\%$

El grupo de personas B tiene un indicador de dispersión que es casi la tercera parte del grupo A, lo cual significa que es un grupo **más homogéneo, menos disperso o más concentrado**, alrededor del valor representativo, que en este caso es la media aritmética o edad promedio.

Dicho de otro modo, la media aritmética del grupo B es de mejor calidad y representatividad que la media aritmética del grupo A.