UNIVERSIDAD DE CANTABRIA

DEPARTAMENTO DE EDUCACIÓN

FACULTAD DE EDUCACIÓN
MÉTODOS DE INVESTIGACIÓN Y DIAGNÓSTICO EN EDUCACIÓN
Edificio Interfacultativo Tfno.: (942) 201281. Fax : (942) 201173
Avda. de los Castros s/n
39005 - Santander
e-mail: laurentino.salvador@unican.es

ASIGNATURA:

«MÉTODOS EN PSICOLOGÍA»

Laurentino SALVADOR BLANCO

INDICE

INTRODUCCIÓN A LA MEDIDA

Medir es asignar un número (en general, un símbolo) a un fenómeno, objeto o relación. La medida tiene tres características fundamentales:

- Es **relativa**, es decir, comparativa, ya que es necesario utilizar un patrón de medida que sirva como unidad: metro, kilo, etc.; o, en cualquier caso, se le atribuye arbitrariamente un número.
- Es **probabilística**, dado que nunca se puede conseguir una medida exacta. Toda medida oscilará entre dos extremos que serán los límites reales superior e inferior. 62,5 cm., por ejemplo, oscilará entre 62,45 y 62,55.
- Es **indirecta**; en las ciencias humanas, la mayor parte de las medidas tienen carácter indirecto ya que las características humanas, hechos sociales, etc., no son medibles en sí mismos.

Medir significa, en términos generales, asignar números a objetos o a relaciones empíricas de acuerdo con ciertas reglas. La estadística se ocupa de la medida para intentar analizar a nivel operatorio las relaciones existentes entre las propiedades de los objetos que hemos medido.

En líneas generales el proceso estadístico sería el siguiente:

Λ Casos empíricos o elementos de una muestra correctamente elegida/extraída de la población de los que elegimos una o varias características (variables).

Λ Atribución de medida según las características de los datos (asignación de número) y ver qué relaciones matemáticas se pueden establecer (identidad-distinción), ordenación de mayor a menor, determinar una unidad constante por la que sabemos que la igualdad de diferencias entre varios datos es la misma, escalamiento con cero absoluto o sin él). Tipos de escalas de medida.

Λ Dependiendo de las características operatorias de los datos aplicaremos los estadísticos adecuados.

Λ Intentamos generalizar los datos de la muestra a la población, en este punto pretendemos volver a los datos empíricos y generalizar las inferencias desprendidas del procesamiento estadístico de los datos.

La estadística se ocupa de ordenar, describir e interpretar conjuntos de datos. Hemos apuntado anteriormente los dos grandes apartados de la estadística:

* La **descriptiva** que describe las características de una muestra. Para ello hay que seguir los pasos siguientes:

1. Ordenar los datos
2. Hallar los valores estadísticos fundamentales del conjunto de datos ordenados:
    - Estadísticos de posicionamiento o tendencia central (media, mediana...)
    - Dispersión de los datos o variabilidad con arreglo a los estadísticos de posicionamiento (desviación típica).
    - Análisis de las relaciones existentes entre distintas variables (correlaciones).

En términos generales, podríamos decir que la estadística representa en un plano los datos y estudia sus características (posiciones, distancias…).

* La **muestral o inferencial**
1. Problemas de muestreo: cómo hay que elegir las muestras para que sean representativas de la población.
2. Fiabilidad o precisión con que un estadístico representa a su parámetro. El error muestral es la diferencia entre el estadístico y el parámetro. Niveles de confianza. Significatividad de los estadísticos. Muestras pequeñas.

Cabrían otros dos apartados como la estadística no paramétrica (las variables no se distribuyen según los parámetros de curva normal) y los diseños experimentales.

En definitiva, conviene comprender la lógica de la estadística (adquirir mentalidad estadística), saber cuando y cómo aplicar los diversos métodos dependiendo de la muestra con la que estemos trabajando y, en función de los tipos de variables, qué estadísticos podemos aplicar para describir esa muestra, cómo hemos de elegir las muestras y en qué límites o condiciones hemos de movernos a la hora de hacer inferencias o extrapolaciones de la muestra a la población.

ESCALAS DE MEDIDA

Las escalas de medida son una norma o conjunto de normas para la asignación de números a los objetos, fenómenos o relaciones. Es necesario que exista **isomorfismo**, o sea, que con los objetos, fenómenos o relaciones puedan hacerse las mismas operaciones que con los números que representan aquéllos.

ESTUDIO DE LAS VARIABLES

DEFINICIÓN	Una característica que puede adoptar diversos valores y que diferencia a los sujetos. Se opone a constante. Constructos: variables latentes no observables directamente. Ej.: inteligencia, motivación, salud... Las variables observables son indicadores de lo no observable.	
ESCALAS	Nominales ------> nominal parcialmente ordenada Ordinales ------> métrica ordenada De intervalo De razón	
CLASIFICACIÓN	CRITERIO DE CLASIFI-CACIÓN	CLASES DE VARIABLES
	Teórico-explicativo	Estímulo (ej. temperatura ambiental) Respuesta Intermediarias u organísmicas (ej. sexo)
	Metodológico	Independientes: activa (manipulable) y asignada (no manipulable) Dependientes Extrañas o intervinientes (a controlar las relevantes)
	Medición	Cualitativas o categóricas (bien definidas, mutuamente excluyentes y exhaustivas): Dicotómicas (ej. sexo) y policotómicas (ej. clase social) y dicotomizada. Cuantitativas: discretas (valor entero) y continuas (entero o fraccionado, por ej. peso)
	Control	Aleatorias (coinciden con las dependientes) Controladas (coinciden con las independientes)

NIVELES DE MEDIDA

NIVELES	DESCRIPCIÓN	PROPIEDADES QUE LO CARACTERIZAN	ESTADÍSTICA Y OPERACIONES APLICABLES	EJEMPLOS
NOMINAL	**La más primitiva Clasificar el fenómeno que estudiamos en base a poseer o no una determinada característica**	**a) Equivalencia: = # b) Clasificación arbitraria c) Las categorías deben ser variables discretas: bien definidas, mutuamente excluyentes y exhaustivas**	**Descriptiva: Frec. absolutas Frec. relativas Porcentajes Proporción Coeficiente de contingencia para establecer relaciones entre variables o hip. de nulidad entre los observado y lo esperado Gráficos: ciclogramas, diagramas de barras**	**Clasificar individuos por su origen geográfico, sexo, religión, estado civil, etc.**
ORDINAL	**Además de la anterior, las características del fenómeno pueden ser ordenadas en función de una dimensión determinada, ya que ésta posee propiedades cuantitativas**	**Además de las nominales determinación de MAYOR QUE y MENOR QUE**	**Además de las nominales, Descriptiva: mediana percentiles Relacionar y comprobar hip.: coef. correlac. Spearman y Kendall**	**Clasificación de individuos por clases sociales, por orden de méritos, intensidad de actitud o preferencias. Por cualificación profesional,etc**
INTERVALO	**Además de las anteriores las distancias entre las diferentes categorías se conocen**	**Además de las anteriores: la distancia entre cada orden y el siguiente es la misma, es decir, se establece una unidad de diferencia constante La primera que se puede considerar cuantitativa: es posible la suma y resta, pero no la multiplicación y división por ser el 0 arbitrario**	**Además de las anteriores: la media, las medidas de variabilidad, análisis de varianza, significatividad diferencias de porcentajes, medias y desviaciones, correlación de Pearson, polígono e histogramas de frecuencias.**	**Puntuaciones de pruebas estandarizadas. Pretenden serlo las escalas de THURSTONE.**
RAZÓN	**Cuando tiene todas las características de una escala de intervalo, pero además tiene un punto 0 real (absoluto) en su origen**	**Además de los anteriores el 0 absoluto: significa la total carencia de una característica. Es posible la suma, resta, multiplicación y división.**	**Cualquier prueba estadística media geométrica**	**Medidas de longitud, talla, peso, edad... Se encuentran en psicofísica donde se ha utilizado para medir los umbrales sensoriales, la intensidad de los estímulos, etc.**

ESTABLECIMIENTO DE NIVELES DE MEDIDA

EJEMPLOS DE VARIABLES					CARACTERÍSTICAS	NIVEL
Sexo	Clase social	Puntuac. CI	Edad			
SI	SI	SI	SI		= <> Clasificación arbitraria Categorías	NOMINAL
NO	SI	SI	SI	V	> <	ORDINAL
NO	NO	SI	SI	V	Unidad constante 0 arbitrario Posible suma y resta	INTERVALO
NO	NO	NO	SI	V	0 absoluto Posible suma, resta, multiplicación y división	RAZÓN

DATOS MÁS HABITUALES EN PSICOLOGÍA/PEDAGOGÍA

<table>
<tr><td rowspan="7">D
A
T
O
S</td><td>CONSTANTE</td><td colspan="2">Característica que sólo puede adoptar un valor</td></tr>
<tr><td rowspan="5">VARIABLE

Característica que puede adoptar diversos valores o categorías</td><td rowspan="2">CUALITATIVA
(nominal)</td><td>Dicotómica</td></tr>
<tr><td>Policotómica</td></tr>
<tr><td colspan="2">CUASICUANTITATIVA
(ordinal)</td></tr>
<tr><td rowspan="2">CUANTITATIVA
(intervalo ---> razón)</td><td>Discreta</td></tr>
<tr><td>Continua</td></tr>
<tr><td colspan="3">PREFERENCIAS O PRIORIDADES:

CUANTITATIVA (continua ---> discreta) ---> CUALITATIVA (ordenada ---> multicotómica ---> dicotómica)</td></tr>
</table>

EJERCICIOS:

1.- Establecer en cada una de las variables los siguientes elementos: Nivel o tipo de escala (nominal...), la catalogación de la misma desde el punto de vista de la medición (cualitativa -dicotómica, multicotómica, ordenada- o cuantitativa -discreta, continua-) y definir su operativización o categorización.

VARIABLES	ESCALA	MEDICION	OPERATIVIZACION
Origen geográfico			
Religión			
Sexo			
Estado civil			
Profesión			
Clase social			
Expediente académico			
Altura (talla)			
Edad			
Peso			
Notas			
Cociente intelectual			
Puntuación en una prueba objetiva			
Dedicación al estudio			
Capacidad para el estudio			
Estado de salud			
Grupo sanguíneo			
Tensión arterial			
Hermanos			

2.-Determínese qué tipo de escala de medida es la más adecuada para cada una de las siguientes variables:

	Variable	**Escala**
a)	Nuestro sistema de numeración cronológica de los años, por ejemplo: 1492, 1650, 1949, 1985, 1991...	
b)	La edad de los sujetos (entendiendo por edad el tiempo de vida desde su nacimiento)	
c)	La escala de dureza de los minerales.	
d)	Los diferentes números de las camisetas de los jugadores de equipos de fútbol.	
e)	La lista de éxitos discográficos del verano	
f)	El tiempo empleado por los pilotos de automóviles en recorrer diez veces un circuito.	
g)	Las marcas de paquetes de cigarrillos.	
h)	Las puntuaciones de veinte estudiantes en una prueba objetiva de rendimiento, donde se valora como un punto cada acierto en las diez preguntas de que consta.	
i)	Los pesos de un conjunto de cuerpos.	
j)	Los apellidos de una lista telefónica.	
k)	El número de pulsaciones por minuto.	
l)	Las calificaciones medias de los expedientes	
m)	Las puntuaciones en un torneo de golf (par, uno bajo par, etc.).	
n)	Los resultados, en número de sets ganados, en un partido de tenis.	
ñ)	Las posiciones de los atletas en el podium, al recibir sus medallas.	
o)	La denominación, por grados, de los meridianos del globo terráqueo.	

3.-En la relación de variables que se muestra a continuación, especifique si se tratan de variables cuantitativas continuas o cuantitativas discretas

	Variable	**Tipo**
a)	El número de hijos de una familia.	
b)	La estatura de los reclutas en un reemplazo.	
c)	El número de piezas defectuosas en un lote de cien unidades.	
d)	La proporción de coches con los neumáticos en mal estado de una ciudad.	
e)	La velocidad media empleada por automovilistas en recorrer una cierta distancia	
f)	La edad de los individuos.	
g)	El número de matrimonios en la población española	
h)	La temperatura corporal de los animales.	
i)	El número de infracciones automovilísticas.	
j)	La cantidad de páginas que contienen los libros.	
k)	El número de huesos que componen los esqueletos.	
l)	El perímetro de los polígonos.	
m)	El sistema de numeración de las casas en las calles y plazas.	
n)	El índice de precios al consumo (IPC).	
ñ)	La cantidad total de asignaturas cursadas a lo largo de una carrera.	
o)	Las distancias entre dos puntos en un mapa.	

4.-Ejercicio de equivalencia entre escalas:
Un grupo de siete sujetos ha sido examinado, en una cierta característica psicológica, por dos psicólogos. Los valores adjudicados a cada sujeto son los indicados en la tabla siguiente:

Sujeto	**A**	**B**	**C**	**D**	**E**	**F**	**G**
Psicólogo A	**4,0**	**1,0**	**6,0**	**2,0**	**3,0**	**6,0**	**5,0**
Psicólogo B	**90**	**100**	**120**	**110**	**150**	**120**	**80**

Indique si ambas escalas empleadas son equivalentes en algún nivel de medida.

SOLUCIÓN:
Se comprueban las propiedades respectivas de las escalas de medida

Escala de razón: Proporción conocida entre valores.
Se toman, por ejemplo, los casos A y B:
Psicólogo A: 4,0 / 1,0 = 4
Psicólogo B: 90 / 100 = 0,9
No se cumple la proporción conocida entre valores.

Escala de intervalo: Proporción conocida entre diferencias.
Se toman, por ejemplo, los casos A, B, C y D.
Psicólogo A: (4,0-1,0) / (6,0-2.0) = 0.75
Psicólogo B: (90-100) / (120-110) = -1
No se cumple la proporción conocida entre diferencias

Escala de rangos: Orden
Se ordenan los sujetos de acuerdo con los valores ofrecidos por los dos psicólogos:
Psicólogo A: B <D <E <A <G<(C=F)
Psicólogo B: G <A <B <D<(C=F) <E
No son equivalentes ambas ordenaciones.

Escala nominal: Ambas escalas son equivalentes a nivel nominal pues a las mismas modalidades (casos C y F) le corresponden los mismos valores en ambas escalas y a modalidades distintas (casos A, B, D, E y G) le corresponden valores distintos en cada escala.

ESTADÍSTICA DESCRIPTIVA

ORDENACIÓN DE DATOS Y DISTRIBUCIÓN DE FRECUENCIAS

C U A L I T	Clasificación por categorías: bien definidas, mutuamente excluyentes, exhaustivas. Ej. Sexo, estado civil	Cuentan el n1 de casos dentro de cada categoría (n1 enteros)	Tabla de frecuencias					Diagrama de barras Ciclograma Pictograma. Los datos o categorías pueden colocarse en cualquier orden
			Categorías	f	P	P_c	P_a	
			...					
			...					
C U A N T I T	Clasificación por intervalos: medidas continuas. Límites teóricos, límites reales (redondeo). Intervalo: categorías numéricas en las que se incluyen las frecuencias.	Cuentan el número de casos en cada intervalo	Distribución de frecuencias					Polígono de frecuencias Histograma Los datos han de colocarse en orden continuo. Suavización de curvas. Comparación de varias muestras en base a los P de las mismas
			Puntuaciones directas	X	f	Xm	f_a	
				Intervalos	Frecuencias	Punto medio del intervalo	Frec. acumuladas	
	OBJETIVO	CARACTERÍSTICAS						REPRESENTACIONES GRÁFICAS

1.- ORDENACIÓN Y REPRESENTACIÓN DE LOS DATOS

Un primer tratamiento de los datos consiste en su ordenamiento y representación gráfica. En este momento hay que decidir si los datos son continuos o discontinuos.

1.1.- Datos cualitativos o discontinuos

Son características que se resisten a la medida. No podemos hacer otra cosa que clasificarlas en un sistema de categorías discontinuas. Las categorías deben reunir los siguientes requisitos:

* *Estar bien definidas:* Ejs. hombre-mujer o Mucha frecuencia-bastante-poca-casi nunca-nunca. Este 2ı ejemplo estaría mal definido. Lo correcto sería: diariamente-5 ó 6 veces por semana-2 a 4 veces por semana-1 vez por semana - 0 veces. Debemos saber dónde incluir a cada sujeto sin ambigüedad.
* *Deben excluirse mutuamente*. El ej.: varones-mujeres-españoles sería incorrecto.
* *Exhaustivas o totalizantes*. Ningún caso puede quedar fuera.

Los datos de las categorías podemos contarlos por frecuencias o número de casos: llamamos N al n total de casos de la muestra. La frecuencias se diferencian en

* absolutas (f): n exacto de casos de una categoría y
* relativas: se usan para representar el tamaño de cada categoría sobre una base común.

Las más usadas son los porcentajes (P):

P = (f/N)A100 y se deben usar siempre que el N sea superior a 100.

Y las proporciones (p). p = (f/N).

Y también podemos representarlas gráficamente mediante diagramas. Los más usados son los de barras y los ciclogramas.

TABLA DE FRECUENCIAS

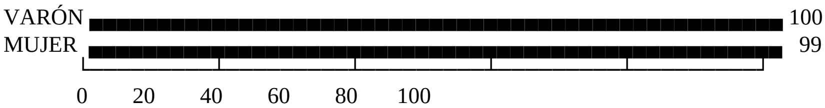
DENOMINACIÓN		DESCRIPCIÓN	CÁLCULO
Frecuencia	f	N de casos en cada categoría	Recuento
Porcentaje	P	Porcentaje de casos en cada categoría	$P = (f/N)·100$
Porcentaje efectivo	P_e	Porcentaje de casos en cada categoría considerando los casos válidos.	$P_e = (f/N_e)·100$
Porcentaje acumulado	P_a	Porcentaje de casos en cada categoría considerando los casos válidos y acumulando los de las categorías previas.	$P_{a1} = P_{e1}$ $P_{a2} = P_{e1} + P_{e2}$ ... $P_{an} = P_{e1} + P_{e2} ... + P_{an}$

Ejemplo: tenemos una muestra de sujetos en los que hemos anotado la variable sexo (1=varón, 2=mujer) obteniendo los siguientes datos:

1;2;1;2;1;2;1;2;1;2;1;2;1;2;1;2;1;1;2;2;1;1;2;2;1;2;1;2;1;2;1;2;1;2;1;2;1;2;1;2;1;2;1;2;2;1;2;1;2;1;2;1;1;2;2;1;2;1;2;2;1;2;1;2;1;2;1;2;1;2;1;2;1;2;1;2;1;2;1;2;1;2;1;2;1
2;1;2;1;2;1;2;1;2;1;2;1;2;1;2;1;2;1;2;1;2;1;2;1;1;2;1;2;1;2;1;2;1;2;1;2;1;2;1;2;1;2;1;1;2;1;2;1;2;1;2;1;2;1;2;1;2;1;2;1;2;1;2;1;2;1;2;1;2;1;2;1;2;1;2;1;2;1;2;1;2;1;2;1;2
1;2;1;2;1;2;1;2;1;2;1;2;1;2;1;2;1;2;1;2;0;1;1;2;1;2

SEXO SEXO

Etiqueta Categoría	Valor X	Frecuenci f	Porcent P	Porcent Efectivo P_e	Porcent Acumulado P_a
VARÓN	1	100	50.0	50.3	50.3
MUJER	2	99	49.5	49.7	100.0
	0	1	0.5	Perdido	
	Total	200	100.0	100.0	



Casos válidos 199 Casos perdidos 1

Hemos obtenido las frecuencias absolutas y relativas. Las hemos representado en un diagrama de barras. Igualmente podemos representarlas en un ciclograma o en un pictograma.

1.2.- Datos cuantitativos discretos

Se pueden medir, aunque los puntos de medida son siempre aproximados, ya que su exactitud depende del instrumento de medida. Por esta razón, se suelen redondear al número exacto más próximo al nivel que deseemos (unidades, décimas, centésimas...). Como norma general, se suele trabajar con tres decimales. Técnica de redondeo:

- La última cifra decimal es superior a 5, se aumenta una unidad a la anterior 98,6->99
- La última cifra decimal es menor que 5, la unidad anterior se deja inalterada 98,4->98
- La última cifra decimal es igual a 5 y la anterior par, ésta se deja inalterada 98,5->98
- La última cifra decimal es igual a 5 y la anterior impar, ésta se aumenta 97,5->98

Hay que tener en cuenta que estas variables se pueden medir y, en vez de hacer tablas de frecuencias, se presentan en distribuciones de frecuencias.

Hay que tener en cuenta que los programas de ordenador a veces redondean tomando la parte entera y esto puede afectar a los cálculos. Por tanto hay que conocerlo cuando decidamos redondear. Y utilizar las funciones adecuadas.

DISTRIBUCIÓN DE FRECUENCIAS

Mediante ellas representamos este tipo de variables y para realizar la distribución podemos dar los siguientes pasos:

1) En función de las puntuaciones de los sujetos determinamos la Amplitud total (A). Buscamos la puntuación máxima y la mínima y calculamos la diferencia entre ambas.

$$A = (Xmáx - Xmin)$$

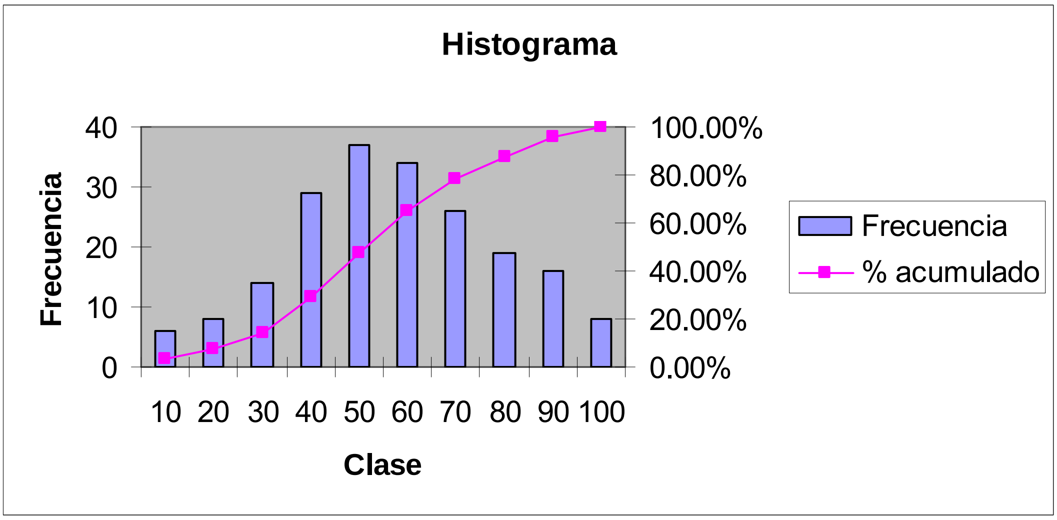
En el ejemplo anteriormente tabulado, la puntuación máxima es 100 y la mínima 1 por lo que la amplitud total es 100.

2) División de la serie de puntuaciones en intervalos o categorías numéricas.
   a) Elección del número de intervalos (n): se suele operar con un número de intervalos entre 7 y 10.
   b) Establecemos la amplitud del intervalo (i) $i = A/n$
   La amplitud del intervalo es igual a la amplitud total dividida por el número de intervalos elegido (n).
   En el ejemplo, vamos a considerar un número de 10 intervalos por lo que
   $$i = 100/10 = 10.$$
   c) Establecemos los límites teóricos o tabulares de los intervalos. 1-10... 91-100

El resto de los elementos (f, P...) se interpretan como en la Tabla de Frecuencias.

Ejemplo: a los mismos sujetos del ejemplo anterior les hemos aplicado un test y hemos obtenido los siguientes datos:
52; 44; 56; 61; 12; 0; 31; 48; 60; 50; 69; 93; 0; 54; 73; 15; 36; 42; 58; 49; 29; 9; 42; 64; 51; 38; 83; 71; 39; 62; 18; 43; 51; 88; 45; 54; 67; 23; 33; 88; 59; 46; 34; 3; 82; 63; 48; 35; 21; 53;100; 97; 25; 49; 55; 37; 66; 75; 35; 38; 48; 57; 86; 33; 72; 13; 87; 41; 81; 51; 61; 24; 76; 1; 95; 32; 61; 84; 71; 56; 62; 26; 75; 53; 63; 44; 65; 43; 50; 45; 30; 80; 47; 70; 59; 37; 40; 54; 79; 77; 38; 15; 54; 63; 37; 13; 32; 31; 83; 66; 87; 61; 65; 40; 48; 71; 42; 23; 35; 60; 39; 58; 33; 21; 35; 81; 71; 50; 79; 99; 38; 51; 56; 45; 44; 29; 69; 18; 46; 53; 48; 24; 62; 30; 56; 9; 93; 43; 59; 97; 86; 2; 26; 80; 61; 42; 41; 51; 34; 25; 57; 68; 75; 47; 12; 64; 54; 88; 3; 49; 33; 76; 53; 70; 35; 51; 73; 45; 82; 55; 72; 95; 63; 59; 44; 37; 36; 67; 48; 75; 41; 84; 43; 54; 52; 50; 49; 62; 88
que hemos agrupado por intervalos o clases

Clase	*Frecuencia*	*% acumulado*
10	6	3.05%
20	8	7.11%
30	14	14.21%
40	29	28.93%
50	37	47.72%
60	34	64.97%
70	26	78.17%
80	19	87.82%
90	16	95.94%
100	8	100.00%
N	197	



Cuando queremos representar varias distribuciones en la misma gráfica, podemos encontrarnos con dos situaciones:

* El tamaño de las dos muestras es el mismo. En este caso, representamos los distintos polígonos con colores distintos o caracteres tipográficos diferentes.

* El tamaño de las muestras es distinto por lo que conviene representar los porcentajes en vez de las frecuencias absolutas y distinguir los polígonos de la misma manera.

Para todo este tipo de trabajos existen al alcance de cualquier usuario numerosos programas de ordenador que permiten la representación gráfica de los datos. Podemos utilizar la hoja dis_frec_ej_p17 del fichero de ejercicios apuntes.xls

2.- RESUMEN DE LOS DATOS: MEDIDAS DE TENDENCIA CENTRAL

Este tipo de medidas ofrecen valores que representan de forma global al conjunto de los valores de una muestra. En definitiva, nos indican el punto -dentro del continuo de los valores de una variable- donde colocar el valor que mejor representa al conjunto de puntuaciones. Por tanto, pretenden resumir en un único valor la tendencia de todo un conjunto de datos de una muestra.

Los que normalmente se consideran son la Moda (Mo), la Mediana (Mdn) y la Media ($\overline{X}$). Vamos a estudiarlos según su menor-mayor relevancia estadística.

En el ejemplo que venimos utilizando hemos obtenido los siguientes estadísticos para la variable X mediante la hoja de Excel citada líneas atrás:

X	
Media	**52.3248731**
Error típico	**1.5485948**
Mediana	**51**
Moda	**48**
Desviación estándar	**21.7355638**
Varianza de la muestra	**472.434735**
Curtosis	**-0.34954399**
Coeficiente de asimetría	**-0.01193432**
Rango	**99**
Mínimo	**1**
Máximo	**100**
Suma	**10308**
Cuenta	**197**
Mayor (1)	**100**
Menor(1)	**1**
Nivel de confianza (95.0%)	**3.05404758**

2.1. LA MODA (Mo)

2.1.1. Concepto: Es el valor que se repite más veces o el más común en un conjunto de datos.

2.1.2. Cálculo:

a) Datos sin agrupar: el valor que más se repite. En el ejemplo que estamos considerando aparecerían varios con las mayores frecuencias de aparición. Es la forma más habitual de encontrarse los datos. Cuando aparecen dos valores con la misma frecuencia, los ordenadores suelen considerar el menor de los valores (por ejemplo el SPSS).

b) Datos agrupados por intervalos o clases: es el punto medio del intervalo de mayor frecuencia. Efectivamente, en los datos del ejemplo con el que estamos trabajando nos encontramos con que el intervalo 41-50 es el de mayor frecuencia (37) y dado que su punto medio es 45,5 ~ 46 este valor puede considerarse como la moda (Mo).

c) Cuando conocemos la Mdn y la $\overline{X}$ se puede calcular un valor aproximado mediante la fórmula

Mo = 3Mdn – 2$\overline{X}$. Ejemplo: Mo = (3·51)-(2·52,325)=153-105=48

2.1.3. Propiedades: es el más sencillo, fácil de calcular y por ello el de menos importancia. Sólo es recomendable su uso cuando no queda más remedio; por ejemplo, cuando tenemos variables cualitativas o categóricas es el único que podemos usar. También puede usarse con datos cuasicuantitativos (ordinales) siendo en este caso el rango con mayor frecuencia. Cuando los datos están agrupados, puede variar en función de la agrupación de intervalos que se haya hecho (mayor o menor número de intervalos... Véase la variación entre las dos formas de cálculo sobre los mismos datos del ejemplo patrón). Es aconsejable su uso cuando en la distribución de intervalos se desconoce el límite de alguno de dichos intervalos o cuando el intervalo de mayor frecuencia coincida con alguno en el que desconozcamos los límites.

2.2. LA MEDIANA (Mdn)

2.2.1. Concepto: es el punto que divide a la muestra en dos partes con el mismo número de casos. O lo que es lo mismo, es el punto que deja al 50% de los casos por debajo en una serie ordenada de valores. En la media intervendrán los datos como valores, en la mediana lo que cuenta es la posición que ocupan en la serie de manera que, cuando el número de datos es impar, la mediana es el valor central y, cuando es par, será la semisuma de los dos centrales.

2.2.2. Cálculo con datos sin agrupar:

La Mdn no es más que un caso particular del cálculo de centiles ya que se corresponde con el centil 50 (C_{50}). Vamos a calcular la Mdn como un caso particular y así podremos aplicarlo al resto de las medidas de posición (cuartiles y deciles en particular).

1) Ordenar los datos de menor a mayor.

2) Calcular la posición (p) en la que se encontraría el centil que buscamos.

$$p = \frac{P \bullet (N+1)}{100}$$ para la serie par

$$p = \frac{P \bullet N}{100}$$ para la serie impar

Serie ordenada impar: 13-19-20-34-44-46-50-54-58-67-98

IMPAR

Centil	Lugar	Elección superior	Valor en esa posición Excel	Valor en esa posición SPSS(*)
C	(C*N)/100			
10	(10*11/100=1,1	2	19	14,2
25	(25*11)/100=2,75	3	20	20
50	(50*11)/100=5,5	6	46	46
75	(75*11)/100=8,25	9	58	58
90	(90*11)/100=9,9	10	67	91,8

(*) Le da igual que la serie sea par o impar y hace todos los cálculos como si fuera impar

Lugar y valor: calculamos el lugar p= (50*11)/100=5,5 redondeamos al entero superior y elegimos el valor que está en esa posición (la 6). En este caso el valor que ocupa esa posición es el 46.

Serie ordenada par:

25-28-29-30-30-31-32-33-34-36-37-37-38-39-40-41-42-45-47-47

PAR

Centil	Lugar	lugar sup	lugar inf	valor sup (Vs)	valor inf (Vi)	diferencia	%	Valor Excel	Valor SPSS
C	C(N+1)/100					D=Vs-Vi	D*(C/100)	Vs-%	Vi+%
10	(10*21/100=2,1	3	2	29	28	1	1*(10/100)=0,1	29-0,10=28,90	28+0,10=26,1
25	(25*21)/100=5,25	6	5	31	30	1	1*(25/100)=0,25	31-0,25=30,75	30+0,25*30,25
50	(50*21)/100=10,5	11	10	37	36	1	1*(50/100)=0,50	37-0,50=36,50	36+0,50=36,5
75	(75*21)/100=15,75	16	15	41	40	1	1*(75/100)=0,75	41-0,75=40,25	40+0,75=40,75
90	(90*21)/100=18,9	19	18	47	45	2	2*(90/100)=1,8	47-1,80=45,20	45+1,8=46,8

Lugar:

Determinamos la posición mediante la fórmula adecuada 50(N+1)/100 que en nuestro ejemplo es (50*21)/100=10,5.

Determinamos el lugar inferior (la parte entera del resultado) y el superior (el lugar inferior + 1). En nuestro ejemplo son 10 y 11 respectivamente. El valor del C_{50} estará entre estas posiciones.

Valor:

A continuación buscamos los valores que están en esas posiciones. Valor superior (Vs=37) y el inferior (Vi=36). Por tanto, entre el 37 y el 36 estaría el valor mediano y podría ser cualquiera de ellos. Un acuerdo para su cálculo es elegir por sistema el punto medio o la semisuma de ambos (37+36)/2=36,5. Otro sistema que utilizan los programas de ordenador, en general, es el siguiente:

- Una vez que hemos determinado el intervalo mediano o los valores entre los que se encontrará el C_{50}. En este caso entre el 36 y el 37.
- Calculamos la longitud del intervalo entre el valor superior y el inferior (37-36)=1.
- Multiplicamos la longitud del intervalo por el porcentaje que corresponde al centil. En este caso, 1*(50/100)=0,5. Es decir repartimos el espacio en del intervalo en 100 partes y vemos el valor que le corresponde a la posición del centil correspondiente.
- Finalmente, si trabajamos con Excel, restamos este dato al valor superior. 37-0,5= 36,5 y, si lo hacemos con SPSS, sumamos el dato al valor inferior (esta solución es la más lógica).

Esta forma de realizar el cálculo coincide con la de Excel (función PERCENTIL) y es la que adoptaremos para los cálculos manuales. El SPSS suma el % del intervalo al valor del límite inferior. Pese a que esta solución es la más lógica, para el trabajo de clase usaremos el Excel dado que, en caso contrario, los cálculos que realicemos manualmente no podremos comprobarlos al no disponer del SPSS.

Como hemos indicado, la Mdn es el punto que divide a la muestra dejando a la mitad de los elementos a cada lado. Podemos dividir la muestra en cuatro partes y así tendremos los cuartiles (Q_1, Q_2 y Q_3). En cinco partes y obtendremos los quintiles (el 1º sería el C_{20}, el 2º el C_{40}, el tercero el C_{60} y cuarto el C_{80}). En diez partes y estaríamos hablando de deciles. Y, finalmente, en cien partes con lo que obtenemos los centiles. El valor de estos índices se calcula como la Mdn que se corresponde con el C_{50} y D_5.

Correspondencia entre los índices:

	D_5	
C_{25}	C_{50}	C_{75}
Q_1	Q_2	Q_3
	Mdn	

En un primer acercamiento a la interpretación de los datos podemos ver cómo están los valores individuales respecto al conjunto. Por ejemplo, si se tratara de las puntuaciones de un test de rendimiento, podríamos hacernos una idea de la posición de cada alumno.

X	Y
9	25
9	28
9	29
9	30
10	30
10	31
10	32
11	33
11	34
12	36
12	37
12	37
12	38
13	39
13	40
13	41
14	42
14	45
14	47
14	47
15	

Manual	**IMPAR**	**PAR**
N	21	20
mdn	12	36.5
Centil	10	10
entero(lugar)+1	3	3
lugar	2.1	2
lugar inferior	3	2
lugar superior	3	3
valor inferior	9	28
valor superior	9	29
diferencia	0	1
%del intervalo	0	0.1
Valor Excel	9	28.9
Valor SPSS	9	28.1
Excel		
C_{10}	9	28.9
C_{25}	10	30.75
C_{50}	12	36.5
C_{75}	13	40.25
C_{90}	14	45.2

C_{10} serie impar:

- lugar = (10*21)/100=2,1=3
- valor= 9 que ocupa el lugar 3

C_{10} serie par:

- lugar :
(10*20)/100=2 y
(10*21)/100=2,1=3
- valor:
En el lugar 2 está el valor 28
En el lugar 3 está el valor 29
La longitud del intervalo es 29-28=1
El % del intervalo que corresponde a 10 (el número de centil) es 1*(10/100)=0,1
El valor será igual al valor del límite superior menos el % del intervalo. 29-0,1=28,9
Opción Excel. SPSS suma el % del intervalo al valor inferior: 28+0.1=28.1

Procedimiento para el cálculo manual
1) Introducimos los datos (si el N es impar en X y si es par en Y)
2) Si no están ordenados, los ordenamos de menor a mayor
3) Introducimos en nº del centil que queremos calcular
serie impar
4) Según el valor que nos salga en lugar, redondeamos al entero superior
5) Elegimos el valor que encontremos en ese punto y lo consideramos como límite superior e inferior
6) Por lo que el resultado sería dicho valor

serie par
4) Tomaremos el valor entero que nos en lugar. Los límites serán esa posición y la siguiente.
5) El ordenador selecciona los valores inferior y superior
6) Resultado en la casilla valor.

Siempre que introduzcamos más datos hay que repasar las celdas que usan el nombre de la variable

Estadísticos obtenidos mediante el SPSS

		X	Y
N	Válidos	21	20
	Perdidos	0	1
Media		11,71	36,05
Mediana		12,00	36,50
Moda		9(a)	30(a)
Desv. típ.		1,953	6,386
Varianza		3,814	40,787
Asimetría		-,045	,196
Error típ. de asimetría		,501	,512
Curtosis		-1,292	-,819
Error típ. de curtosis		,972	,992
Rango		6	22
Mínimo		9	25
Máximo		15	47
Percentiles	10	9,00	28,10
	25	10,00	30,25
	50	12,00	36,50
	75	13,50	40,75
	90	14,00	46,80

a Existen varias modas. Se mostrará el menor de los valores.

Se pueden apreciar las diferencias en los resultados según el programa de cálculo.

2.2.3. Propiedades: indicada cuando no se puede calcular la media por las características de los datos (variables discretas), cuando hay influencia de los valores extremos para el cálculo de la media, cuando la distribución es incompleta o abierta siempre que el intervalo crítico no coincida con el abierto. En función de cómo se construya la distribución de intervalos, el valor de la Mdn puede cambiar. Más aconsejable que la media cuando la distribución es marcadamente asimétrica (puntuaciones concentradas en algún extremo).

2.3. LA MEDIA ARITMÉTICA ($\overline{X}$)

2.3.1. Concepto: es el valor que mejor representa a un grupo de puntuaciones. Es el promedio de un conjunto de valores y es el más deseable a utilizar siempre que se pueda.

2.3.2. Cálculo:

a) *Datos sin agrupar*. $\overline{X}=\Sigma X/N$.
Ej.: 3,3,7,7,7,7,10,10,10,10,10,12,12,12,15,15

$\overline{X} = \Sigma X/N = 150/16 = 9{,}375$

b) *Media total de varios grupos:*[1]

$$\overline{X}_t = \frac{N_1\cdot\overline{X}_1+N_2\cdot\overline{X}_2+...+N_n\cdot\overline{X}_n}{N_1+N_2+...N_n}$$

Por extensión se procedería igual con las proporciones y los porcentajes sustituyendo las X por p ó P según de lo que se trate.

Por ejemplo: las medias de tres muestras son 8, 10 y 7 y sus N respectivos 12, 14 y 20. Si calculamos la media sumando las medias y dividiendo por el número de medias tendríamos de resultaría una media total de 25/3= 8,3333. Mientras que, si aplicamos la fórmula adecuada, tendríamos lo siguiente:

$$\overline{X}_t = \frac{12\cdot 8+10\cdot 14+20\cdot 7}{12+14+20} = \frac{96+140+140}{146} = \frac{3{,}76}{46} = 8{,}1739$$

c) *Media total ponderada:* las medias de los distintos grupos tienen distinto peso (h).

$$\overline{X}_p = \frac{h_1\cdot\overline{X}_1+h_2\cdot\overline{X}_2+...+h_n\cdot\overline{X}_n}{h_1+h_2+...h_n}$$

Ejemplo 1: Calcular la media total de un grupo de alumnos sabiendo que la media de prácticas es 6 puntos, 3 la del examen parcial y 4,5 la del examen final. La Nota de prácticas se valora en un 42%, la del parcial en un 20% y la del final en un 38%.

$\overline{X}_p = \Sigma\,(h_n\cdot \overline{X}_n)/h_t = (42*6+ 20*3+38*4{,}5)/(42+20+38)=4{,}83$

[1] **Conviene recordar las normas para resolver operaciones combinadas: 1º se resuelven los paréntesis, 2º se resuelven las multiplicaciones y divisiones en el orden en que aparecen y 3º se resuelven las sumas y restas. Igualmente conviene manejar adecuadamente las reglas del sumatorio (Σ).**

Ejemplo 2: Calcular la nota final de un alumno sabiendo que obtiene 6 puntos en prácticas, 3 en el examen parcial y 4,5 en el examen final. La Nota de prácticas se valora en un 42%, la del parcial en un 20% y la del final en un 38%. Aplicaríamos la misma fórmula sustituyendo cada media por la puntuación correspondiente del alumno (X).

2.3.3. Propiedades:

* Es función de todas y cada una de las puntuaciones de la distribución. Basta que cambie un valor cualquiera de la misma para que varíe el valor de la media.
* La suma de las diferencias de las puntuaciones de una muestra (X_1,X_2...X_n) con respecto a la media vale 0: $\sum (X_i - \overline{X}) = 0$
* La media es función de los intervalos elegidos para la distribución de frecuencias (número, amplitud...). En el ejemplo patrón se puede ver cómo la con datos sin agrupar es 52,325 y con datos agrupados por intervalos y frecuencias es 52,708.
* Es más sensible que la mediana a la variación de una puntuación. La mediana no tiene que variar necesariamente cuando cambia un valor de la distribución.
* No es aconsejable su uso cuando la distribución es marcadamente asimétrica.
* No es posible su cálculo en distribuciones abiertas. Dado que la media se calcula en base a los puntos medios de los intervalos. Cuando en uno de ellos no se conoce, no es posible calcularla.
* Si a todas las puntuaciones de una distribución se le suma una constante arbitraria, la media de la nueva variable así generada es igual a la media de la variable original más el valor de la constante que se ha sumado a cada puntuación.
* Si todas las puntuaciones de una variable son multiplicadas por una constante, el valor de la media de la nueva distribución es igual al valor de la media de la distribución original multiplicada por el valor de la constante.

3.- VARIABILIDAD DE LOS DATOS: MEDIDAS DE DISPERSIÓN

Seguimos completando la descripción de la muestra cada vez con más precisión. Hasta ahora hemos estudiado las frecuencias por categorías, por intervalos, hemos extraído los valores representativos de la misma, etc. Sin embargo, puede suceder que varias muestras, aún teniendo la misma media, sean muy diferentes debido a la variabilidad o dispersión de los datos respecto a la media. Por esta razón, necesitamos los índices de variabilidad. Si representáramos los valores en el espacio tendríamos a cada caso situado en un punto y un valor representativo del conjunto (media). Además podríamos apreciar cómo se distribuyen los sujetos por el citado espacio: muy juntos o muy dispersos respecto al valor central. Los índices de variabilidad nos van a dar información sobre dicha dispersión.

3.1.- Rango o Amplitud total (A). Ya conocida y explicada con anterioridad. A mayor recorrido, mayor dispersión. Prácticamente, su utilidad se reduce a la correcta realización de la distribución de intervalos. Su mayor problema como índice de dispersión es que se ve muy afectado por las puntuaciones extremas.

3.2.- Desviación semiintercuartil (Q): intenta resolver el principal problema de A. Para ello utiliza solamente el 50% central de los casos (25% por debajo y 25% por encima de la Mdn) desechando el otro 50%. También se denomina amplitud semiintercuartil (ASI) o error probable (EP).

$$Q = (Q_3 - Q_1) / 2$$

Está indicado su uso como índice de dispersión siempre que, por las características de los datos, nos hayamos visto obligados a utilizar la Mdn como única medida de tendencia central. El cálculo de este índice es, como ya hemos indicado en el apartado de la mediana, el correspondiente al C_{75} y C_{25}.

3.3.- Desviaciones medias (D_{Me}, DM):

- *Desviación media respecto a la mediana (D_{Me}):* se define como la media aritmética de los valores absolutos de las diferencias entre los valores de la variable y la mediana. $D_{Me} = \Sigma|X\text{-}Mdn|/N$

- *Desviación media respecto a la media (DM):* es el primer estadístico de este tipo que utiliza todas y cada una de las puntuaciones de la muestra. Se define como el promedio de las desviaciones de cada puntuación respecto a la media. Su cálculo se realiza mediante la siguiente expresión cuando se trata de datos sin agrupar.

$$DM = \Sigma|X - \overline{X}|/N$$

Las propiedades de este índice se pueden comparar con las de la media.

3.4.- Desviación típica (σ,s) y varianza (σ²,s²):

3.4.1. Definición y características: es la más usada y la más fiable. Junto con la media es la base para cálculos posteriores. Se define como la raíz cuadrada de la varianza o, lo que es lo mismo, como la raíz cuadrada positiva de la media de la suma de los cuadrados de las desviaciones (de cada valor respecto a la media). Dado que la varianza se expresa en las unidades de medida de la variable al cuadrado, se utiliza la desviación estándar ya que maneja las mismas unidades de medida que la variable que describe.

3.4.2. Procedimientos de cálculo: datos sin agrupar[2]:

Ejemplo

	$X - \overline{X}$	
X	x	x^2
2	-3	9
3	-2	4
7	2	4
8	3	9
20	0	26

Lo primero que deberemos realizar es una tabla como la anterior y calcular la $\overline{X}$. A continuación calcularemos la puntuación diferencial (x) o diferencia entre la puntuación de cada caso y la $\overline{X}$ (desviaciones respecto ésta). Seguidamente elevamos cada desviación (x) al cuadrado.

N = 4 $\overline{X}$ = 5

$$x = X_i - \overline{X}$$

$$\sigma_n = \sqrt{\frac{\sum x^2}{N}} \qquad \sigma_n = \sqrt{\frac{26}{4}} = 2{,}5495$$

$$\sigma_{n-1} = \sqrt{\frac{\sum x^2}{N-1}} \qquad \sigma_{n-1} = \sqrt{\frac{26}{3}} = 2{,}9439$$

[2] **El SPSS/PC+ utiliza N-1. Muchos programas de cálculo hacen lo mismo. Cuando las muestras son grandes no hay diferencias; sin embargo, con muestras pequeñas el índice obtenido puede ser muy diferente.**

Como ya hemos indicado, lo más habitual es utilizar ésta última fórmula. En nuestro caso siempre indicaremos con qué fórmula debe calcularse σ_n o σ_{n-1}.

3.4.3. Significación y propiedades:

* En una distribución aproximadamente normal la desviación típica tiene las siguientes propiedades:

  a) La zona delimitada por los valores de la $\overline{X} \pm 1\sigma$ incluye al 68% de los casos.

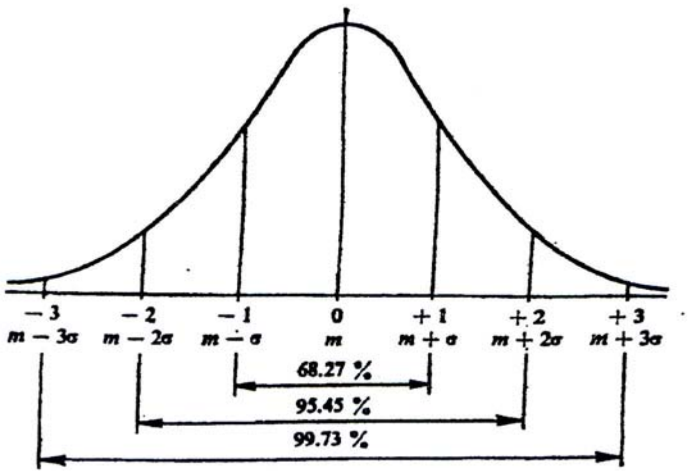
  b) La zona delimitada por los valores de la $\overline{X} \pm 1{,}96\sigma$ incluye al 95% de los casos.

  c) La zona delimitada por los valores de la $\overline{X} \pm 2\sigma$ incluye al 95,45% de los casos.

  d) La zona delimitada por los valores de la $\overline{X} \pm 2{,}58\sigma$ incluye al 99% de los casos.

  e) La zona delimitada por los valores de la $\overline{X} \pm 3\sigma$ incluye al 99,7% de los casos.

De manera que se puede establecer una escala típica (Z), con $\overline{X} = 0$ y σ =1. Este aspecto se tratará más adelante en el capítulo de valoración de las puntuaciones individuales.



Denominaremos σ_{n-1} a la obtenida con N-1 y σ_n a la obtenida con el N total.

* Es función de todas y cada una de las puntuaciones que componen la muestra.
* Es función de los intervalos elegidos para la construcción de la distribución de frecuencias (número, amplitud...).
* Viene expresada en las mismas unidades que la variable a que se refiere.
* Si se le suma una constante a todas las puntuaciones de una muestra, la desviación típica no varía.
* Si multiplicamos por una constante arbitraria a todas las puntuaciones de una muestra, la desviación típica es igual a la original multiplicada por el valor absoluto de la constante.
* No es aconsejable cuando no se puede hallar la media.
* Cálculo de la σ combinada (McNemar 1962):

$$\sigma_t = \sqrt{\frac{N_1(\overline{X}_1^2 + \sigma_1^2) + N_2(\overline{X}_2^2 + \sigma_2^2) + \ldots + N_n(\overline{X}_n^2 + \sigma_n^2)}{N_1 + N_2 + \ldots + N_n} - \overline{X}_t^2}$$

$$\sigma_t = \sqrt{\frac{N_1(\sigma_1^2 + d_1^2) + N_2(\sigma_2^2 + d_2^2) + \ldots + N_n(\sigma_n^2 + d_n^2)}{N_1 + N_2 + \ldots + N_n}}$$

Donde $d_1 = \overline{X}_1 - \overline{X}_t$; $d_2 = \overline{X}_2 - \overline{X}_t$; $d_n = \overline{X}_n - \overline{X}_t$

3.5.- Coeficiente de Variación (CV) de Pearson o dispersión relativa:

Indica el número de veces (%, regla de tres) que la σ contiene a la $\overline{X}$

Como criterio de interpretación se suele adoptar el siguiente: si es ≤0,1 la $\overline{X}$ representa adecuadamente al conjunto de los valores (cuanto más cerca de 0 mejor, cuanto más alejada peor).

Permite la comparación del grado de homogeneidad de dos muestras y exige que la variable analizada provenga de una escala de razón (0 absoluto).

$$CV = \frac{\sigma \cdot 100}{\overline{X}}$$

3.6.- Coeficiente de Variación (V_{Me}) respecto a la mediana:

Se trata de resolver el problema de comparación de medianas de varias distribuciones que pueden utilizar escalas con unidades diferentes. Se define como el cociente entre la desviación media respecto a la mediana (D_{Me}) y la mediana (Mdn).

$$V_{Me} = D_{Me} / Mdn$$

A menor índice de dispersión mejor es la mediana.

4.- ASIMETRÍA Y CURTOSIS: MEDIDAS DE FORMA

Los dos estadísticos más importantes son la media y la desviación típica. Ambos pertenecen a un grupo que se denominan "momentos". Los primeros cuatro momentos respecto a la media de una distribución son los siguientes:

$$m_1 = \frac{\sum (X - \overline{X})}{N} = \frac{\sum x}{N} = 0$$

$$m_2 = \frac{\sum (X - \overline{X})^2}{N} = \frac{\sum x^2}{N} = \sigma^2$$

$$m_3 = \frac{\sum (X - \overline{X})^3}{N} = \frac{\sum x^3}{N} = \sigma^3$$

$$m_4 = \frac{\sum (X - \overline{X})^4}{N} = \frac{\sum x^4}{N} = \sigma^4$$

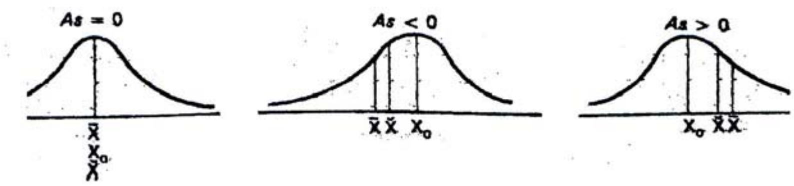
Momento de orden n respecto a la media con datos sin agrupar:

$$m_n = \frac{\sum (X - \overline{X})^n}{N} = \frac{\sum x^n}{N} = \sigma^n$$

La asimetría y la curtosis nos informan de la aproximación de la forma de la distribución que estamos estudiando a la forma de la curva normal por lo que, si fuera similar, podríamos utilizar todas las aplicaciones de ésta. Aunque un buen diseño muestral puede presagiar una distribución normal de los datos, nunca hay que suponer que esto es así y se hace necesario estudiar si nuestra muestra de estudio se distribuye normalmente.

4.1.- Asimetría (As, g_1)

Significado: Cuando las dos partes de una distribución no coinciden se habla de distribución asimétrica (una mitad de la curva no es imagen exacta de la otra). Por tanto, cabe hablar de dos tipos de curvas: las simétricas y las asimétricas y, dentro de éstas, las asimétricas positivas y negativas. La asimetría positiva (>0) se produce cuando hay un predominio de los valores bajos y la mayoría de los valores tienden a situarse a la izquierda. La asimetría negativa (<0) se produce cuando predominan los valores altos y se sitúan gráficamente a la derecha del eje de abscisas.

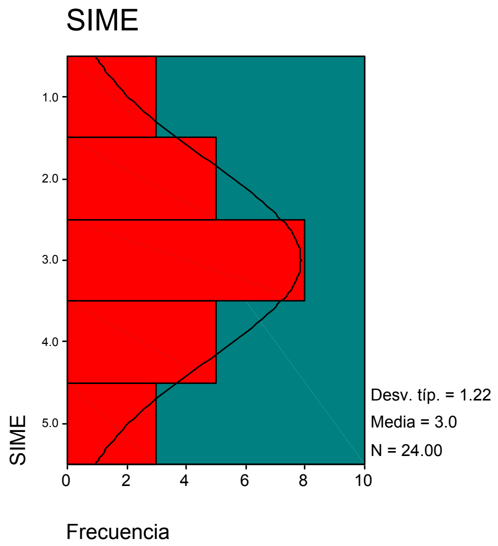


Estadísticos descriptivos SPSS

	N	Media	Desv. típ.	Asimetría	
	Estadístico	Estadístico	Estadístico	Estadístico	Error típico
sime	24	3,00	1,216	,000	,472
asimn	24	3,46	1,351	-,250	,472
asimp	24	2,67	1,308	,302	,472

SIME

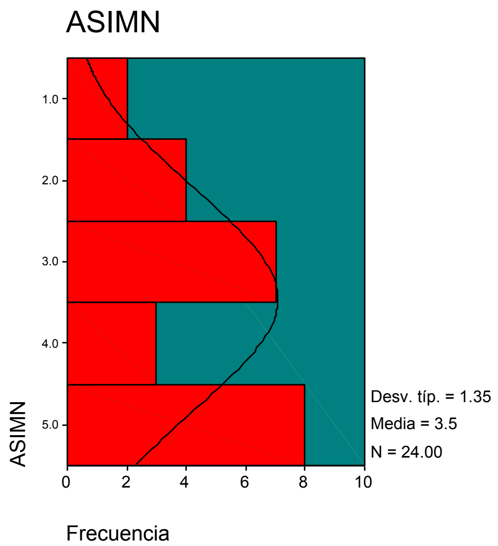
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	1	3	12.5	12.5	12.5
	2	5	20.8	20.8	33.3
	3	8	33.3	33.3	66.7
	4	5	20.8	20.8	87.5
	5	3	12.5	12.5	100.0
	Total	24	100.0	100.0	



As = 0,000

ASIMN

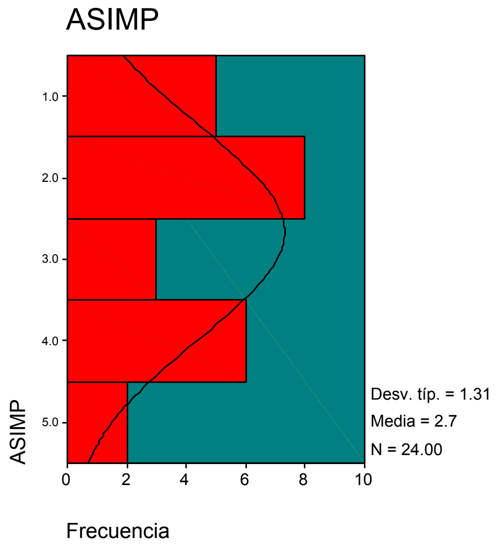
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	1	2	8.3	8.3	8.3
	2	4	16.7	16.7	25.0
	3	7	29.2	29.2	54.2
	4	3	12.5	12.5	66.7
	5	8	33.3	33.3	100.0
	Total	24	100.0	100.0	



As = - 0,250

ASIMP

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	1	5	20.8	20.8	20.8
	2	8	33.3	33.3	54.2
	3	3	12.5	12.5	66.7
	4	6	25.0	25.0	91.7
	5	2	8.3	8.3	100.0
	Total	24	100.0	100.0	



As = 0,302

Procedimiento de cálculo:

- *Índice basado en los cuartiles*: asimetría cuartílica de BOWLEY. Considera el 50% central de la distribución.

$$As_1 = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{(Q_3 - Q_1)} = \frac{Q_1 + Q_3 - 2Mdn}{Q_3 - Q_1}$$

- *Índice basado en los percentiles o coeficiente absoluto de asimetría:* considera el 80%

$$As_2 = \frac{(P_{90} - P_{50}) - (P_{50} - P_{10})}{(P_{90} - P_{10})}$$

Si la distribución fuera absolutamente normal este índice sería 0. Podemos comprobarlo sustituyendo los percentiles por sus porcentajes en la curva normal 90, 10 y 50).

$$As_2 = \frac{(90-50)-(50-10)}{(90-10)} = \frac{40-40}{80} = \frac{0}{80} = 0$$

- *Índice de Pearson:* tiene en cuenta todas las puntuaciones

$$As_3 = \frac{3 \cdot (\overline{X} - Mdn)}{\sigma}$$

- *Índice de Fisher* para datos sin agrupar:

$$As_4 = \frac{m_3}{m_2 \sqrt{m_2}} = \frac{\frac{\sum (X-\overline{X})^3}{N}}{\sigma^3} = \frac{\sum \left(z_x^3\right)}{N}$$

Este coeficiente, cuando el N>150, se puede estandarizar y sería normal con una media de 0 y una varianza de 6/N. Por lo que el **coeficiente de asimetría estandarizado** se calcularía con la siguiente expresión. Este coeficiente es asintóticamente normal (0,1).

$$As_{4s} = \frac{As_4}{\sqrt{\frac{6}{N}}}$$

Los programas estadísticos como el SPSS o el Excel, por ejemplo, utilizan la siguiente fórmula que será la que utilizaremos siempre que dispongamos de todos los datos primarios:

$$As_5 = \frac{N}{(N-1)\cdot(N-2)} \bullet \sum \left(\frac{X_i - \overline{X}}{\sigma} \right)^3$$

Interpretación:

En el caso del índice de Fisher, si una distribución es normal, la suma de los cubos de las desviaciones positivas es igual a la suma de los cubos de las desviaciones negativas, con lo que la suma algebraica de los cubos de las desviaciones es cero. Si es asimétrica positiva, la suma de los cubos de las desviaciones positivas es mayor que la suma de los cubos de las desviaciones negativas con lo que As es positivo. En el caso contrario, As es negativo y cuanto mayor sea el valor de As mayor será la asimetría.

Para los índices As_4 y As_5, si el valor de la asimetría está comprendido en el intervalo que marcan las tablas (ver Tabla 4) se puede decir que la población de la que procede la muestra es simétrica.

Si consultamos la Tabla 4, nos encontramos que, para un N=24 y un α =5%, el valor entre el que deberían estar los índices que hemos obtenido sería ±0,880 y para un α =1% el valor que aparece es ± 1,223 para ser consideradas como simétricas. Ver ARDANUY,R./SOLDEVILLA, Mª M. (1992, p. 37). Los índices obtenidos deberían ser iguales o mayores a los que hemos observado en la tabla. Si el valor que obtenemos está dentro de los límites, podemos decir "que la población de la que proceden los datos puede considerarse simétrica y que la asimetría observada en la muestra es debida al azar". Así ocurre con nuestros datos.

El resto de los índices de asimetría obtenidos mediante otras fórmulas (As_1 a As_3) se interpretan de manera que si el valor es 0 hay simetría, si es superior a 0 hay asimetría positiva y si es inferior a cero hay asimetría negativa. En el caso del índice de Pearson, se interpreta como casi simétrica si el valor obtenido se encuentra entre 0 y 0,37, dudosa aplicación de criterios de normalidad (>0,37-1) y totalmente asimétrica (>1).

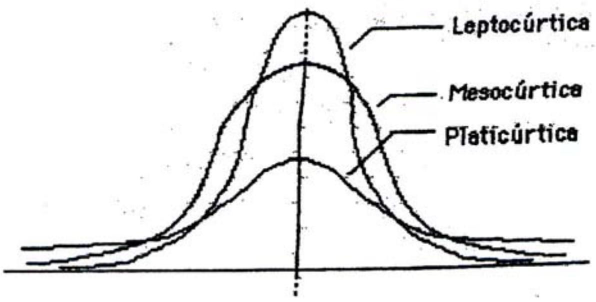
Propiedades: relacionadas con la media y la mediana de la variable

* Si la distribución es simétrica, media y mediana coinciden en un mismo punto del eje de abscisas y es sobre el que se levanta el eje de simetría del polígono de frecuencias.
* Si asimétrica positiva, el valor de la mediana es menor que el de la media.

Esto se debe a que existen valores altos que hacen que la media se eleve mientras que la mediana se mantiene más próxima al centro de los datos.

* Si es asimétrica negativa, el efecto es el contrario. El valor de la media es más bajo que el de la mediana.

4.2.- Curtosis (Curt, K, g_2)



Tipos de curtosis

Una curva puede ser simétrica y, sin embargo, no ser normal. Para saber si una curva es normal, la sometemos a dos criterios: el de la asimetría o sesgo y el de curtosis o apuntamiento. Aún teniendo la misma media, si las puntuaciones se acumulan en torno a la media hay muy poca dispersión (leptocúrtica), si hay mucha dispersión el apuntamiento es mínimo (platicúrtica) y mesocúrtica cuando el apuntamiento es "normal".

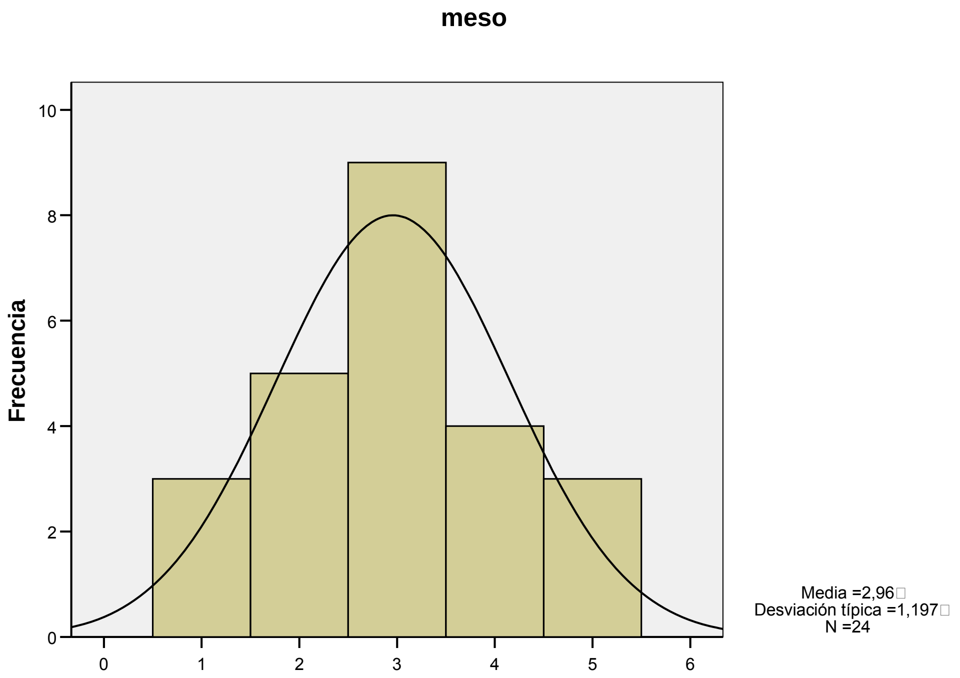
Ejemplos:

Estadísticos descriptivos SPSS

	N	Media	Desv. típ.	Curtosis	
	Estadístico	Estadístico	Estadístico	Estadístico	Error típico
meso	24	2,96	1,197	-,543	,918
plati	24	2,83	1,404	-1,299	,918
lepto	24	2,92	,776	4,044	,918

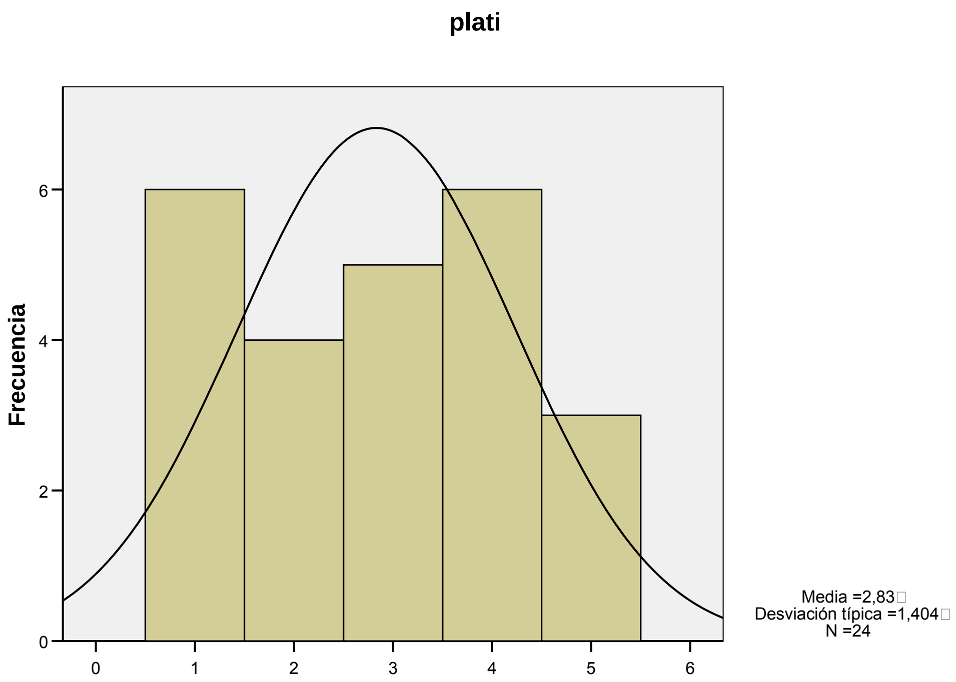
Meso (-0.543)

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	**1**	**3**	**12,5**	**12,5**	**12,5**
	2	**5**	**20,8**	**20,8**	**33,3**
	3	**9**	**37,5**	**37,5**	**70,8**
	4	**4**	**16,7**	**16,7**	**87,5**
	5	**3**	**12,5**	**12,5**	**100,0**
	Total	**24**	**100,0**	**100,0**	



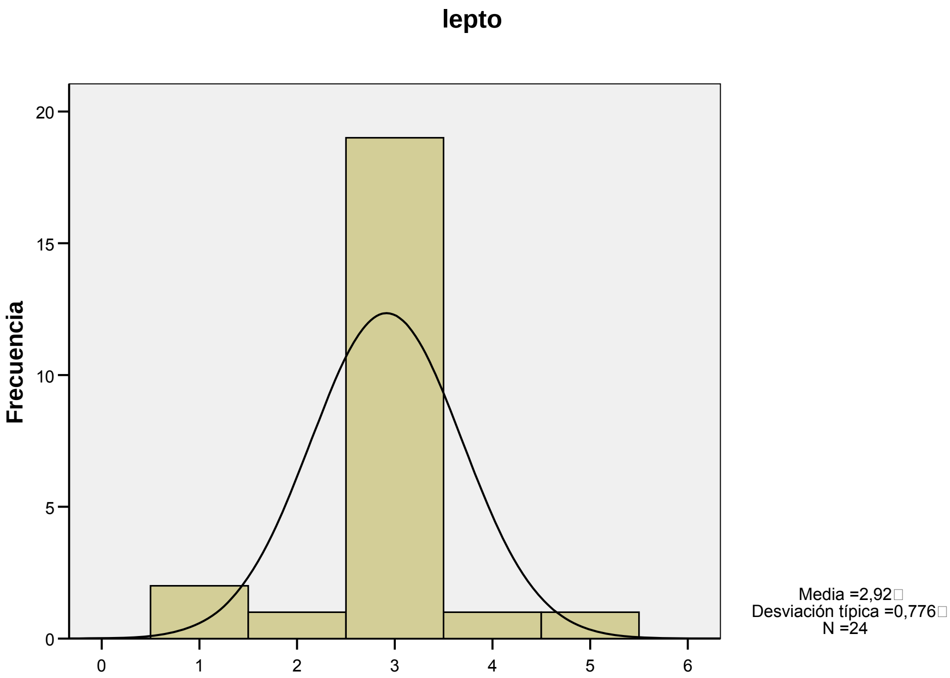
***Plati* (-1,299)**

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	1	6	25,0	25,0	25,0
	2	4	16,7	16,7	41,7
	3	5	20,8	20,8	62,5
	4	6	25,0	25,0	87,5
	5	3	12,5	12,5	100,0
	Total	24	100,0	100,0	



Lepto (4,044)

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	1	2	8,3	8,3	8,3
	2	1	4,2	4,2	12,5
	3	19	79,2	79,2	91,7
	4	1	4,2	4,2	95,8
	5	1	4,2	4,2	100,0
	Total	24	100,0	100,0	



Índice en base a cuartiles y percentiles:

$$K_1 = \frac{Q}{P_{90} - P_{10}} = \frac{Q_3 - Q_1}{2(P_{90} - P_{10})}$$

Interpretación: en base a la curtosis de la curva normal. 1) K>0,263 la curva es platicúrtica (K>0,4 es anormal). 2) K=0,263 es normal. 3) K<0,263 es leptocúrtica (sobre todo, <0,12).

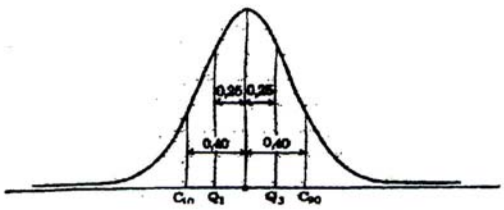
Demostración de que la curtosis de la curva normal es 0,263:

Q_3 en z = $0{,}6745\sigma$

Q_1 en z = $-0{,}6745\sigma$

P_{90} en z = $1{,}28\sigma$

P_{10} en z = $-1{,}28\sigma$



Coeficiente de exceso(o de Fisher):

Podemos utilizar la siguiente formula que es la que usan el SPSS, Excel... y que es la que utilizaremos habitualmente.

$$K_2 = \left\{ \frac{N(N+1)}{(N-1)(N-2)(N-3)} \sum \left(\frac{x}{\sigma} \right)^4 \right\} - \frac{3(N-1)^2}{(N-2)(N-3)}$$

Interpretación: 0 = mesocúrtica. >0 = leptocúrtica y <0 = platicúrtica.

Las tablas 5a y 5b nos presentan los valores críticos de la curtosis y nos permiten hacer la estimación de si la población es mesocúrtica, platicúrtica o leptocúrtica.

Este coeficiente, cuando el N>150, se puede estandarizar y sería normal con una media de 0 y una varianza de 24/N. Por lo que el **coeficiente de curtosis estandarizado** se calcularía con la siguiente expresión. Este coeficiente es asintóticamente normal (0,1).

$$K_{2S} = \frac{K_2}{\sqrt{\frac{24}{N}}}$$

Estadísticos obtenidos mediante el SPSS

		ASIMN	SIME	ASIMP	MESO	LEPTO	PLATI
N	**Válidos**	**24**	**24**	**24**	**24**	**24**	**24**
	Perdidos	**0**	**0**	**0**	**0**	**0**	**0**
Media		**3,46**	**3,00**	**2,67**	**2,96**	**3,38**	**2,67**
Mediana		**3,00**	**3,00**	**2,00**	**3,00**	**3,50**	**3,00**
Moda		**5**	**3**	**2**	**3**	**4**	**2(a)**
Desv. típ.		**1,35**	**1,22**	**1,31**	**1,20**	**1,01**	**1,13**
Varianza		**1,82**	**1,48**	**1,71**	**1,43**	**1,03**	**1,28**
Asimetría		**-,250**	**,000**	**,302**	**,086**	**-,859**	**,139**
Error típ. de asimetría		**,472**	**,472**	**,472**	**,472**	**,472**	**,472**
Curtosis		**-1,112**	**-,696**	**-1,148**	**-,543**	**,999**	**-,735**
Error típ. de curtosis		**,918**	**,918**	**,918**	**,918**	**,918**	**,918**
Rango		**4**	**4**	**4**	**4**	**4**	**4**

a Existen varias modas. Se mostrará el menor de los valores.

Si consultamos las citadas tablas 5a y 5b para un N=24 nos encontramos con unos valores críticos para el α = 5% (-1,194 <-----> 1,620) y para el 1% (-1,348 <-----> 2,895). Dado que todos nuestros valores se encuentran dentro de estos valores, las distribuciones de las muestras pueden considerarse normales y el signo nos indica el apuntamiento (+) o achatamiento (-) en cada caso. Ver ARDANUY,R./ SOLDEVILLA, Mª M. (1992, p. 37). Cuando el valor que obtenemos se encuentra dentro de los límites encontrados en las tablas, podemos decir "que la población de la que proceden los datos puede considerarse mesocúrtica y que el exceso observado en la muestra es debido al azar".

EJEMPLO del cálculo de la asimetría y la curtosis:

asim/curtosis	x	x2	x3	x4
19	-2	4	-8	16
19	-2	4	-8	16
20	-1	1	-1	1
20	-1	1	-1	1
21	0	0	0	0
21	0	0	0	0
22	1	1	1	1
23	2	4	8	16
24	3	9	27	81

Excel

Suma x	0,000	24,000	18,000	132,000	
SUMA	189,000	desv3	5,1962	desv4	9,0000
N	9				
MEDIA	21,000	**Calculo Manual**			
MEDIANA	21,000	As1	0,000		
MODA	19,000	As2	0,048		
DESV	1,732	As3	0,000		
asimetria	0,557	As4	0,385	As4s	0,4714
curtosis	-0,643	As5	0,557		
Q1	20,00	K1	0,238		
Q2	21,00	K2	-0,643	K2s	-0,39367
Q3	22,00				
P10	19,00				
P25	20,00				
P50	21,00				
P75	22,00				
P90	23,20				

Asimetría:

$$As_1 = \frac{Q_1 + Q_3 - 2Mdn}{Q_3 - Q_1} = \frac{20 + 22 - 42}{22 - 40} = \frac{0}{2} = 0$$

$$As_2 = \frac{(P_{90} - P_{50}) - (P_{50} - P_{10})}{(P_{90} - P_{10})} = \frac{(23{,}20 - 21) - (21 - 19)}{(23{,}20 - 19)} = \frac{2{,}20 - 2}{4{,}2} = \frac{0{,}20}{4{,}2} = 0{,}0476$$

$$As_3 = \frac{3 \cdot (\overline{X} - Mdn)}{\sigma} = \frac{3 \cdot (21 - 21)}{1{,}732} = \frac{0}{1{,}732} = 0$$

$$As_4 = \frac{m_3}{m_2\sqrt{m_2}} = \frac{\dfrac{\sum (X - \overline{X})^3}{N}}{\sigma^3} = \frac{\dfrac{18}{9}}{5{,}196} = \frac{2}{5{,}196} = 0{,}385$$

$$As_{4s} = \frac{As_4}{\sqrt{\frac{6}{N}}} = \frac{0,385}{\sqrt{0,666}} = \frac{0,385}{0,8164} = 0,4714$$

$$As_5 = \frac{N}{(N-1)\cdot(N-2)} \bullet \sum\left(\frac{X_i - \overline{X}}{\sigma}\right)^3 = \frac{9}{8 \bullet 7} \bullet \frac{18}{5,196} = 0,1607 \bullet 3,464 = 0,557$$

Curtosis:

$$K_1 = \frac{Q_3 - Q_1}{2(P_{90} - P_{10})} = \frac{22-20}{2(23,2-19)} = \frac{2}{8,4} = 0,238$$

$$K_2 = \left\{\frac{N(N+1)}{(N-1)(N-2)(N-3)} \sum\left(\frac{x}{\sigma}\right)^4\right\} - \frac{3(N-1)^2}{(N-2)(N-3)}$$

$$K_2 = \left\{\frac{9(10)}{(8)(7)(6)} \bullet \frac{132}{9}\right\} - \frac{3(64)}{(7)(6)} = \left\{\frac{90}{336} \bullet \frac{132}{9}\right\} - \frac{192}{42} =$$

$$K_2 = (0,2678 \bullet 14,667) - 4,571 = 3,932 - 4,571 = -0,6434$$

$$K_{2S} = \frac{K_2}{\sqrt{\frac{24}{N}}} = \frac{-0,6434}{\sqrt{2,6666}} = -0,3937$$

Finalmente, comprobaremos si los índices obtenidos, especialmente As_5 y K_2, se encuentran entre los límites establecidos en las tablas 5a y 5b.

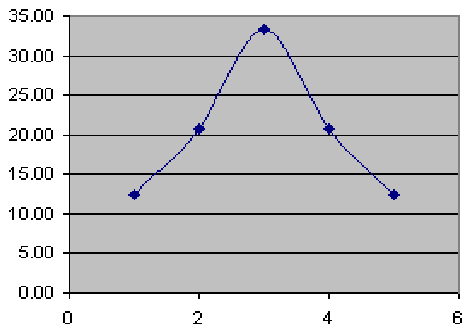
EJEMPLOS DE PROCEDIMIENTOS DE CÁLCULO MEDIANTE EXCEL (ver ejercicios apuntes.xls hoja asim_curto)

Ejemplo 1:								
SIME	x	x2	x3	x4	zx	zx3	zx4	
1	-2.000	4.000	-8.000	16.000	-1.645	-4.451	7.322	
1	-2.000	4.000	-8.000	16.000	-1.645	-4.451	7.322	
1	-2.000	4.000	-8.000	16.000	-1.645	-4.451	7.322	
2	-1.000	1.000	-1.000	1.000	-0.822	-0.556	0.458	
2	-1.000	1.000	-1.000	1.000	-0.822	-0.556	0.458	
2	-1.000	1.000	-1.000	1.000	-0.822	-0.556	0.458	
2	-1.000	1.000	-1.000	1.000	-0.822	-0.556	0.458	
2	-1.000	1.000	-1.000	1.000	-0.822	-0.556	0.458	
3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
4	1.000	1.000	1.000	1.000	0.822	0.556	0.458	
4	1.000	1.000	1.000	1.000	0.822	0.556	0.458	
4	1.000	1.000	1.000	1.000	0.822	0.556	0.458	
4	1.000	1.000	1.000	1.000	0.822	0.556	0.458	
4	1.000	1.000	1.000	1.000	0.822	0.556	0.458	
5	2.000	4.000	8.000	16.000	1.645	4.451	7.322	
5	2.000	4.000	8.000	16.000	1.645	4.451	7.322	
5	2.000	4.000	8.000	16.000	1.645	4.451	7.322	
Excel	0.000	34.000	0.000	106.000	0.000	0.000	48.507	

SUMA	72.000		
N	24		
MEDIA	3.000	**Calculo Manual**	
MEDIANA	3.000	As1	0.000
MODA	3.000	As2	0.000
DESV	1.216	As3	0.000
asimetría	0.000	As4	0.000
curtosis	-0.696	As5	0.000
Q1	2.00	K1	0.294
Q2	3.00	K2	-0.696
Q3	4.00		
P10	1.30		
P25	2.00		
P50	3.00		
P75	4.00		
P90	4.70		

Tabla de frecuencias y porcentajes

Valores	f	P	Pe	Pa
1	3	12.50	12.50	12.50
2	5	20.83	20.83	33.33
3	8	33.33	33.33	66.67
4	5	20.83	20.83	87.50
5	3	12.50	12.50	100.00
N	24	100.00		
Ne	24			



Simétrica

Ejemplo 2:

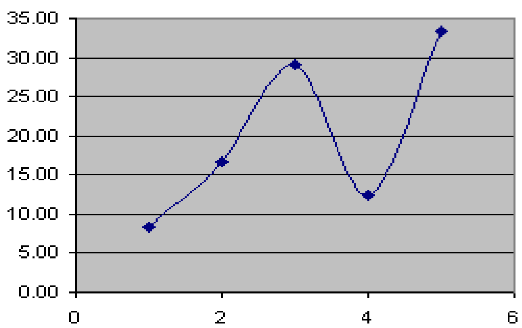
ASIMN	x	x2	x3	x4	zx	zx3	zx4
1	-2.458	6.043	-14.857	36.523	-1.820	-6.030	10.974
1	-2.458	6.043	-14.857	36.523	-1.820	-6.030	10.974
2	-1.458	2.127	-3.101	4.523	-1.080	-1.259	1.359
2	-1.458	2.127	-3.101	4.523	-1.080	-1.259	1.359
2	-1.458	2.127	-3.101	4.523	-1.080	-1.259	1.359
2	-1.458	2.127	-3.101	4.523	-1.080	-1.259	1.359
3	-0.458	0.210	-0.096	0.044	-0.339	-0.039	0.013
3	-0.458	0.210	-0.096	0.044	-0.339	-0.039	0.013
3	-0.458	0.210	-0.096	0.044	-0.339	-0.039	0.013
3	-0.458	0.210	-0.096	0.044	-0.339	-0.039	0.013
3	-0.458	0.210	-0.096	0.044	-0.339	-0.039	0.013
3	-0.458	0.210	-0.096	0.044	-0.339	-0.039	0.013
3	-0.458	0.210	-0.096	0.044	-0.339	-0.039	0.013
4	0.542	0.293	0.159	0.086	0.401	0.065	0.026
4	0.542	0.293	0.159	0.086	0.401	0.065	0.026
5	1.542	2.377	3.664	5.649	1.141	1.487	1.697
5	1.542	2.377	3.664	5.649	1.141	1.487	1.697
4	0.542	0.293	0.159	0.086	0.401	0.065	0.026
5	1.542	2.377	3.664	5.649	1.141	1.487	1.697
5	1.542	2.377	3.664	5.649	1.141	1.487	1.697
5	1.542	2.377	3.664	5.649	1.141	1.487	1.697
5	1.542	2.377	3.664	5.649	1.141	1.487	1.697
5	1.542	2.377	3.664	5.649	1.141	1.487	1.697
5	1.542	2.377	3.664	5.649	1.141	1.487	1.697
Excel	0.000	41.958	-13.003	136.896	0.000	-5.277	41.135
SUMA	83.000						

N	24
MEDIA	3.458
MEDIANA	3.000
MODA	5.000
DESV	1.351
asimetría	-0.250
curtosis	-1.112
Q1	2.75
Q2	3.00
Q3	5.00
P10	2.00
P25	2.75
P50	3.00
P75	5.00
P90	5.00

Calculo Manual	
As1	0.778
As2	0.333
As3	1.018
As4	-0.220
As5	-0.250
K1	0.375
K2	-1.112

Tabla de frecuencias y porcentajes

Valores	f	P	Pe	Pa
1	2	8.33	8.33	8.33
2	4	16.67	16.67	25.00
3	7	29.17	29.17	54.17
4	3	12.50	12.50	66.67
5	8	33.33	33.33	100.00
N	24	100.00		
Ne	24			



Asimetría negativa

Ejemplo 3

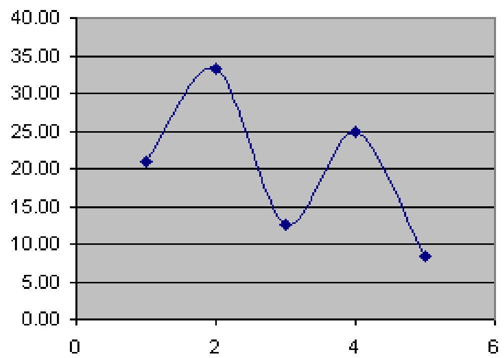
ASIMP	x	x2	x3	x4	zx	zx3	zx4
1	-1.667	2.778	-4.630	7.716	-1.274	-2.070	2.638
1	-1.667	2.778	-4.630	7.716	-1.274	-2.070	2.638
1	-1.667	2.778	-4.630	7.716	-1.274	-2.070	2.638
1	-1.667	2.778	-4.630	7.716	-1.274	-2.070	2.638
1	-1.667	2.778	-4.630	7.716	-1.274	-2.070	2.638
2	-0.667	0.444	-0.296	0.198	-0.510	-0.132	0.068
2	-0.667	0.444	-0.296	0.198	-0.510	-0.132	0.068
2	-0.667	0.444	-0.296	0.198	-0.510	-0.132	0.068
2	-0.667	0.444	-0.296	0.198	-0.510	-0.132	0.068
2	-0.667	0.444	-0.296	0.198	-0.510	-0.132	0.068
2	-0.667	0.444	-0.296	0.198	-0.510	-0.132	0.068
2	-0.667	0.444	-0.296	0.198	-0.510	-0.132	0.068
2	-0.667	0.444	-0.296	0.198	-0.510	-0.132	0.068
3	0.333	0.111	0.037	0.012	0.255	0.017	0.004
3	0.333	0.111	0.037	0.012	0.255	0.017	0.004
3	0.333	0.111	0.037	0.012	0.255	0.017	0.004
4	1.333	1.778	2.370	3.160	1.020	1.060	1.081
4	1.333	1.778	2.370	3.160	1.020	1.060	1.081
4	1.333	1.778	2.370	3.160	1.020	1.060	1.081
4	1.333	1.778	2.370	3.160	1.020	1.060	1.081
4	1.333	1.778	2.370	3.160	1.020	1.060	1.081
4	1.333	1.778	2.370	3.160	1.020	1.060	1.081
5	2.333	5.444	12.704	29.642	1.784	5.680	10.135
5	2.333	5.444	12.704	29.642	1.784	5.680	10.135
Excel	0.000	39.333	14.222	118.444	0.000	6.359	40.499

SUMA	64.000
N	24
MEDIA	2.667
MEDIANA	2.000
MODA	2.000
DESV	1.308
asimetría	0.302
curtosis	-1.148
Q1	2.00
Q2	2.00
Q3	4.00
P10	1.00
P25	2.00
P50	2.00
P75	4.00
P90	4.00

Calculo Manual	
As1	1.000
As2	0.333
As3	1.529
As4	0.265
As5	0.302
K1	0.333
K2	-1.148

Tabla de frecuencias y porcentajes

Valores	f	P	Pe	Pa
1	5	20.83	20.83	20.83
2	8	33.33	33.33	54.17
3	3	12.50	12.50	66.67
4	6	25.00	25.00	91.67
5	2	8.33	8.33	100.00
N	24	100.00		
Ne	24			



Asimetría positiva

Ejemplo 4

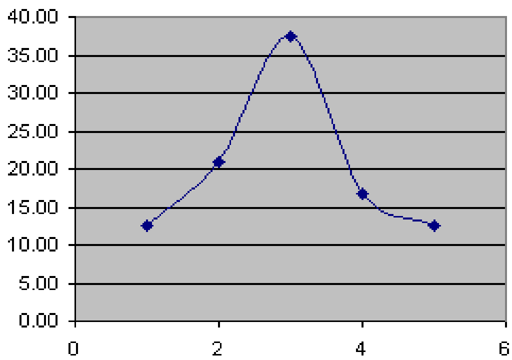
MESO	x	x2	x3	x4	zx	zx3	zx4
1	-1.958	3.835	-7.510	14.708	-1.636	-4.378	7.163
1	-1.958	3.835	-7.510	14.708	-1.636	-4.378	7.163
1	-1.958	3.835	-7.510	14.708	-1.636	-4.378	7.163
2	-0.958	0.918	-0.880	0.843	-0.801	-0.513	0.411
2	-0.958	0.918	-0.880	0.843	-0.801	-0.513	0.411
2	-0.958	0.918	-0.880	0.843	-0.801	-0.513	0.411
2	-0.958	0.918	-0.880	0.843	-0.801	-0.513	0.411
2	-0.958	0.918	-0.880	0.843	-0.801	-0.513	0.411
3	0.042	0.002	0.000	0.000	0.035	0.000	0.000
3	0.042	0.002	0.000	0.000	0.035	0.000	0.000
3	0.042	0.002	0.000	0.000	0.035	0.000	0.000
3	0.042	0.002	0.000	0.000	0.035	0.000	0.000
3	0.042	0.002	0.000	0.000	0.035	0.000	0.000
3	0.042	0.002	0.000	0.000	0.035	0.000	0.000
3	0.042	0.002	0.000	0.000	0.035	0.000	0.000
3	0.042	0.002	0.000	0.000	0.035	0.000	0.000
3	0.042	0.002	0.000	0.000	0.035	0.000	0.000
4	1.042	1.085	1.130	1.177	0.870	0.659	0.573
4	1.042	1.085	1.130	1.177	0.870	0.659	0.573
4	1.042	1.085	1.130	1.177	0.870	0.659	0.573
4	1.042	1.085	1.130	1.177	0.870	0.659	0.573
5	2.042	4.168	8.510	17.376	1.706	4.961	8.462
5	2.042	4.168	8.510	17.376	1.706	4.961	8.462
5	2.042	4.168	8.510	17.376	1.706	4.961	8.462
Excel	0.000	32.958	3.122	105.177	0.000	1.820	51.221

SUMA	71.000
N	24
MEDIA	2.958
MEDIANA	3.000
MODA	3.000
DESV	1.197
asimetría	0.086
curtosis	-0.543
Q1	2.00
Q2	3.00
Q3	4.00
P10	1.30
P25	2.00
P50	3.00
P75	4.00
P90	4.70

Calculo Manual	
As1	0.000
As2	0.000
As3	-0.104
As4	0.076
As5	0.086
K1	0.294
K2	-0.543

Tabla de frecuencias y porcentajes

Valores	f	P	Pe	Pa
1	3	12.50	12.50	12.50
2	5	20.83	20.83	33.33
3	9	37.50	37.50	70.83
4	4	16.67	16.67	87.50
5	3	12.50	12.50	100.00
N	24	100.00		
Ne	24			



Mesocúrtica

Ejemplo 5								
PLATI	**x**	**x2**	**x3**	**x4**	**zx**	**zx3**	**zx4**	
1	-1.667	2.778	-4.630	7.716	-1.476	-3.214	4.744	
1	-1.667	2.778	-4.630	7.716	-1.476	-3.214	4.744	
1	-1.667	2.778	-4.630	7.716	-1.476	-3.214	4.744	
1	-1.667	2.778	-4.630	7.716	-1.476	-3.214	4.744	
2	-0.667	0.444	-0.296	0.198	-0.590	-0.206	0.121	
2	-0.667	0.444	-0.296	0.198	-0.590	-0.206	0.121	
2	-0.667	0.444	-0.296	0.198	-0.590	-0.206	0.121	
2	-0.667	0.444	-0.296	0.198	-0.590	-0.206	0.121	
2	-0.667	0.444	-0.296	0.198	-0.590	-0.206	0.121	
2	-0.667	0.444	-0.296	0.198	-0.590	-0.206	0.121	
2	-0.667	0.444	-0.296	0.198	-0.590	-0.206	0.121	
3	0.333	0.111	0.037	0.012	0.295	0.026	0.008	
3	0.333	0.111	0.037	0.012	0.295	0.026	0.008	
3	0.333	0.111	0.037	0.012	0.295	0.026	0.008	
3	0.333	0.111	0.037	0.012	0.295	0.026	0.008	
3	0.333	0.111	0.037	0.012	0.295	0.026	0.008	
3	0.333	0.111	0.037	0.012	0.295	0.026	0.008	
3	0.333	0.111	0.037	0.012	0.295	0.026	0.008	
4	1.333	1.778	2.370	3.160	1.181	1.646	1.943	
4	1.333	1.778	2.370	3.160	1.181	1.646	1.943	
4	1.333	1.778	2.370	3.160	1.181	1.646	1.943	
4	1.333	1.778	2.370	3.160	1.181	1.646	1.943	
4	1.333	1.778	2.370	3.160	1.181	1.646	1.943	
5	2.333	5.444	12.704	29.642	2.066	8.820	18.224	
Excel	0.000	29.333	4.222	77.778	0.000	2.932	47.818	

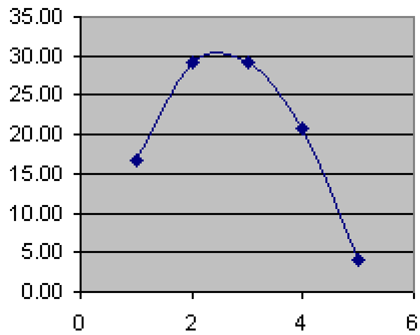
SUMA	64.000
N	24
MEDIA	2.667
MEDIANA	3.000
MODA	2.000
DESV	1.129
asimetría	0.139
curtosis	-0.735
Q1	2.00
Q2	3.00
Q3	3.25
P10	1.00
P25	2.00
P50	3.00
P75	3.25
P90	4.00

Calculo Manual	
As1	-0.600
As2	-0.333
As3	-0.885
As4	0.122
As5	0.139
K1	0.208
K2	-0.735

Tabla de frecuencias y porcentajes

Valores	f	P	Pe	Pa
1	4	16.67	16.67	16.67
2	7	29.17	29.17	45.83
3	7	29.17	29.17	75.00
4	5	20.83	20.83	95.83
5	1	4.17	4.17	100.00
N	24	100.00		
Ne	24			

Platicúrtica



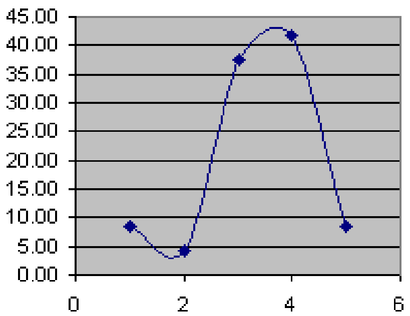
Ejemplo 6								
LEPTO	**x**	**x2**	**x3**	**x4**	**zx**	**zx3**	**zx4**	
1	-2.375	5.641	-13.396	31.817	-2.343	-12.868	30.155	
1	-2.375	5.641	-13.396	31.817	-2.343	-12.868	30.155	
2	-1.375	1.891	-2.600	3.574	-1.357	-2.497	3.388	
3	-0.375	0.141	-0.053	0.020	-0.370	-0.051	0.019	
3	-0.375	0.141	-0.053	0.020	-0.370	-0.051	0.019	
3	-0.375	0.141	-0.053	0.020	-0.370	-0.051	0.019	
3	-0.375	0.141	-0.053	0.020	-0.370	-0.051	0.019	
3	-0.375	0.141	-0.053	0.020	-0.370	-0.051	0.019	
3	-0.375	0.141	-0.053	0.020	-0.370	-0.051	0.019	
3	-0.375	0.141	-0.053	0.020	-0.370	-0.051	0.019	
3	-0.375	0.141	-0.053	0.020	-0.370	-0.051	0.019	
3	-0.375	0.141	-0.053	0.020	-0.370	-0.051	0.019	
4	0.625	0.391	0.244	0.153	0.617	0.235	0.145	
4	0.625	0.391	0.244	0.153	0.617	0.235	0.145	
4	0.625	0.391	0.244	0.153	0.617	0.235	0.145	
4	0.625	0.391	0.244	0.153	0.617	0.235	0.145	
4	0.625	0.391	0.244	0.153	0.617	0.235	0.145	
4	0.625	0.391	0.244	0.153	0.617	0.235	0.145	
4	0.625	0.391	0.244	0.153	0.617	0.235	0.145	
4	0.625	0.391	0.244	0.153	0.617	0.235	0.145	
4	0.625	0.391	0.244	0.153	0.617	0.235	0.145	
4	0.625	0.391	0.244	0.153	0.617	0.235	0.145	
5	1.625	2.641	4.291	6.973	1.603	4.122	6.609	
5	1.625	2.641	4.291	6.973	1.603	4.122	6.609	
Excel	0.000	23.625	-18.844	82.857	0.000	-18.101	78.531	

SUMA	81.000
N	24
MEDIA	3.375
MEDIANA	3.500
MODA	4.000
DESV	1.013
asimetría	-0.859
curtosis	0.999
Q1	3.00
Q2	3.50
Q3	4.00
P10	2.30
P25	3.00
P50	3.50
P75	4.00
P90	4.00

Calculo Manual	
As1	0.000
As2	-0.412
As3	-0.370
As4	-0.754
As5	-0.859
K1	0.294
K2	0.999

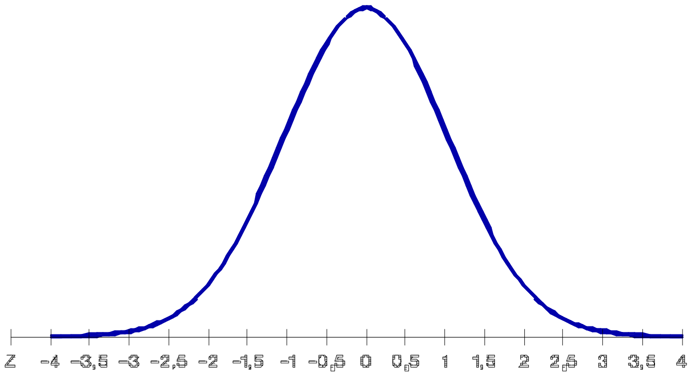
Tabla de frecuencias y porcentajes

Valores	f	P	Pe	Pa
1	2	8.33	8.33	8.33
2	1	4.17	4.17	12.50
3	9	37.50	37.50	50.00
4	10	41.67	41.67	91.67
5	2	8.33	8.33	100.00
N	24	100.00		
Ne	24			



Leptocúrtica

5.- LA CURVA NORMAL Y SUS APLICACIONES



La curva normal es la representación gráfica de una ecuación matemática como expresión de la relación operativa entre variables. En Psicología, los datos empíricos suelen representarse en una curva que tiene forma aproximada de campana (de Gauss). Cualquier fenómeno aleatorio al repetirse un número infinito de veces viene representado por una curva que se denomina "curva de probabilidad" (teoría central del límite). Dicha curva se denomina "normal" y es la representación de la ecuación de normalidad. Su altura máxima ($Y_{máx}$ = 0,3989) se encuentra en la media y técnicamente se expresa diciendo que la ordenada máxima corresponde a una abscisa igual a la media.

La curva normal es asintótica (en teoría las ramas son tangentes al eje horizontal de abscisas, en la práctica casi todos los casos quedan comprendidos entre las tres desviaciones típicas a cada lado de la media), la asimetría es nula y la curtosis es mesocúrtica.

La ecuación matemática de la curva de una distribución normal es

$$Y = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)}$$

x = cualquier valor de X   e = 2,7182
σ = desviación típica   μ = media
π = 3,1416

Tipificando una curva normal, es decir, convirtiendo sus valores directos (X) en puntuaciones típicas Z, podemos aplicar la fórmula siguiente y trazar la curva normal tipificada en base a las ordenadas correspondientes (y). La escala de puntuaciones típicas tiene una media=0 y una desviación típica de 1, su amplitud se extiende entre ± 5, aunque entre ± 3 se incluiría el 99,73%. De esta manera obtenemos una *curva normal unitaria* que no depende de la $\overline{X}$ ni de la σ de la distribución concreta.

$$Y = \frac{1}{\sqrt{2\pi}} e^{-\left(\frac{Z^2}{2}\right)}$$

Para una Z=0, obtenemos un valor de y=0,3989. Comprobar en Tabla 2.

$$Y = \frac{1}{\sqrt{2 \bullet 3{,}1416}} 2{,}7183^{-\left(\frac{0^2}{2}\right)} = 0{,}3989$$

Para realizar el cálculo de $2{,}7183^{-0{,}5}$ utilizamos la calculadora escribiendo -0,5 y pulsamos a continuación la tecla e^x. En este caso nos dará 0,6065 que multiplicado por 0,3989 es igual a 0,2419.

$$Y = \frac{1}{\sqrt{2 \bullet 3{,}1416}} 2{,}7183^{-\left(\frac{1^2}{2}\right)} = 0{,}3989 \bullet 2{,}7183^{-0{,}5} = 0{,}2420$$

En esta curva, puede calcularse el área que existe entre dos ordenadas (entre dos valores de Z), área por debajo o por encima de un valor Z determinado. Podemos comprobar en la Tabla de curva normal el área que existe entre la ordenada máxima (media) y un valor concreto de Z (Tabla 1). Dichas áreas se interpretan como probabilidad de que un valor elegido al azar esté comprendido entre Z valores típicos dados o como porcentaje de casos con respecto al total que hay entre Z valores dados.

Equivalencia entre índices en una curva normal:

* $\overline{X}$, Mdn y Mo coinciden.
* $\overline{X}$ ± 1DM = 58% de los casos
* $\overline{X}$ ± Q = 50% de los casos
* Q = 2/3 σ
* DM = 4/5 σ
* Q < DM < σ

Propiedades:

* La curva normal es simétrica con respecto a la ordenada correspondiente a la media.
* Coinciden media, mediana y moda.
* Asintótica respecto al eje de abscisas, se acerca al eje pero nunca coincide.
* Entre dos valores determinados de su abscisa (escala de puntuaciones de la variable), hay siempre el mismo área relativa con respecto al total.

Aplicaciones:

A) Dado uno o más puntos de la distribución se pueden averiguar las frecuencias (%) que corresponden a esos puntos.

1 Dado un punto averiguar la frecuencia (%) bajo o sobre ese punto.

2 Dados dos puntos averiguar la frecuencia (%) entre ambos.

5 Dado un punto de la escala averiguar la probabilidad (%) de escoger al azar un valor que esté sobre o debajo de ese punto (igual que 1)

6 Dados dos puntos, averiguar la probabilidad de escoger al azar un valor que esté entre ellos (igual que 2).

7 Dada una probabilidad (%), averiguar el punto con respecto al cual existe esa probabilidad (%) de que un valor escogido al azar esté por encima o por debajo de él (igual que 3).

8 Dada una probabilidad (%), averiguar los puntos a ambos lados de la media con respecto a los cuales existe esa probabilidad de que un caso elegido al azar esté entre ellos (igual que 4).

9 Dado un punto, averiguar la probabilidad (%) de que un valor elegido al azar se separe de la media más o menos que ese punto (inverso de 8).

B) Dada una o más frecuencias, averiguar los puntos.

3 Dada una frecuencia, averiguar el punto bajo o sobre el cual está dicha frecuencia (%) (inverso de 1)

4 Dada una frecuencia (%) central averiguar los puntos a ambos lados de la media que comprenden esa frecuencia (%) (inverso del caso 2).

Siempre que queramos emplear las tablas para resolver las aplicaciones anteriores, tendremos que convertir las puntuaciones directas a puntuaciones típicas dado que la escala de puntuaciones típicas (Z) es el patrón de referencia en la distribución de normalidad. Para ello, a todas las puntuaciones directas se les aplica la siguiente fórmula:

$$Z = \frac{X - \overline{X}}{\sigma}$$

La puntuación Z es, por tanto, igual a la puntuación diferencial dividida por la variabilidad de las puntuaciones.

Todos los casos presentados pueden resolverse conociendo la $\overline{X}$, la σ de la muestra, pasando las puntuaciones a Z y manejando la tabla de áreas bajo la curva normal entre la media y un valor típico dado. Igualmente es necesario despejar adecuadamente los elementos de la fórmula para la obtención de las puntuaciones Z.

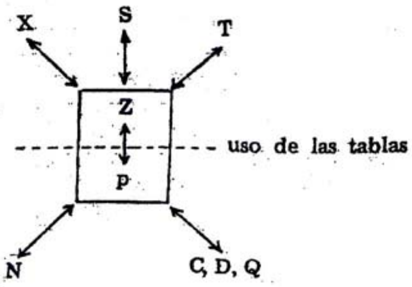
$$X = z \bullet \sigma + \overline{X}$$

$$\overline{X} = X - z \bullet \sigma$$

$$\sigma = \frac{X - \overline{X}}{z}$$

Las dos utilidades básicas son las siguientes:

* Dado un punto (X, S, T --->Z), averiguar p --->%
* Dado un porcentaje (% ---> p), averiguar Z --->X, S, T



Ejemplo:

Base: $\overline{X}$ = 50, σ= 6, distribución normal. N = 1000.

1 Averiguar la frecuencia por encima y por debajo de 60.

z=(60-50)/6=1,67. Vamos a las tablas y nos encontramos con que entre la media y 1,67 está el 45,25%, luego habrá un 95,25% por debajo y un 4,75% por encima.

2 Averiguar el porcentaje de individuos que hay entre las puntuaciones 45 y 35.

z1= (45-50)/6= -0,83

z2= (35-50)/6= -2,5

Entre z2 y la media hay un 49,38%, entre z1 y la media hay un 29,67%. Por tanto, entre ambos habrá 49,38-29,67=19,71%

5 Dado un punto de la escala averiguar la probabilidad (%) de escoger al azar un valor que esté sobre o debajo de ese punto (igual que 1)

6 Dados dos puntos, averiguar la probabilidad de escoger al azar un valor que esté entre ellos (igual que 2).

7 Dada una probabilidad (%), averiguar el punto con respecto al cual existe esa probabilidad (%) de que un valor escogido al azar esté por encima o por debajo de él (igual que 3).

8 Dada una probabilidad (%), averiguar los puntos a ambos lados de la media con respecto a los cuales existe esa probabilidad de que un caso elegido al azar esté entre ellos (igual que 4).

9 Dado un punto, averiguar la probabilidad (%) de que un valor elegido al azar se separe de la media más o menos que ese punto (inverso de 8).

3 Dada una frecuencia, averiguar el punto bajo o sobre el cual está dicha frecuencia (%) (inverso de 1). ¿Cuál es el punto que deja bajo si al 75% de los casos (Q_3)?. Buscamos el área más próxima a 25%, en la tabla es 24,86 al que corresponde una z de 0,67.

x = z·σ= 0,67·6= 4,02

X = x + $\overline{X}$ = 4,02+50=54,02 será el punto bajo el que está el 75% y sobre el cual está el 25% restante.

4 Dada una frecuencia (%) central averiguar los puntos a ambos lados de la media que comprenden esa frecuencia (%) (inverso del caso 2). Ej.: averiguar la distancia que hay que tomar a ambos lados de la media para que el área comprendida entre ambos puntos contenga el 50% medio de los casos. A ambos lados habrá un 25% por lo que buscamos en las tablas, lo más próximo es 2486 que corresponde a 0,67z.

x = z·σ = 0,67·6 = 4,02

X1 = x+ $\overline{X}$ = 54,02

X2 = x- $\overline{X}$ = 45,98 Entre estos valores estará el 50% medio de los casos (esto es Q o amplitud semiintercuartil).

6.- ESTABLECIMIENTO DE PUNTUACIONES INDIVIDUALES

Hasta el momento hemos descrito índices o valores que nos daban información sobre la muestra con la que estábamos trabajando en conjunto. Sin embargo, en Psicología, Pedagogía y Psicopedagogía es muy importante la interpretación y valoración de las puntuaciones individuales. Por ejemplo, si decimos que una persona ha obtenido una puntuación de 10 en un test, no sabemos si eso es mucho, poco, alto, bajo...

El objetivo de este apartado es describir los diversos tipos de puntuaciones que pueden utilizarse para caracterizar a un sujeto y cómo realizar su interpretación en relación con el grupo de individuos al que pertenece (muestra) o a otro (normalización, tipificación, baremación...).

6.1.- Puntuaciones directas

O "brutas". Son las que se obtienen directamente mediante los instrumentos de medida. Se denotan por letras latinas mayúsculas (ej. X_i, como la puntuación del sujeto i en la variable X). Vienen expresadas en las mismas unidades de medida de la variable.

No aportan información para su interpretación por lo que, en general, hay que traspasarlas a ordinales o a típicas.

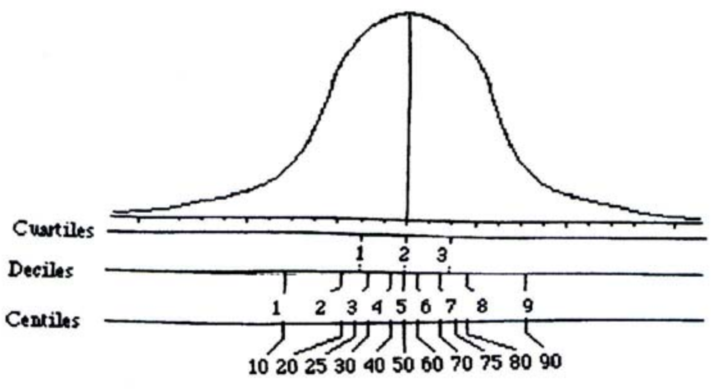
6.2.- Puntuaciones ordinales

Nos permiten indicar la posición que ocupa una cierta puntuación en el conjunto de la muestra, de manera que podemos interpretar la puntuación de un individuo en base a esas referencias. Se denominan ordinales porque la posición se determina mediante una escala ordenada que no tiene unidades constantes. (Ver referencia a las escalas ordinales). Esto quiere decir que no podemos comparar percentiles similares de distintas muestras. Cuando queramos hacer esto, habrá que normalizar las puntuaciones.

Generalmente, se distinguen tres tipos de puntuaciones que dividen a la muestra en un número de partes iguales:

* Los cuartiles (Q_1, Q_2 y Q_3) como tres puntos que dividen a la muestra en cuatro zonas al 25%.

* Los quintiles como cuatro puntos que dividen a la muestra en cinco partes (el 1º sería el C_{20}, el 2º el C_{40}, el tercero el C_{60} y cuarto el C_{80}).
* Los deciles (D_1 ... D_9) como nueve puntos que dividen a la muestra en 10 partes iguales.
* Los centiles (C_1 ... C_{99}) o 99 puntos que dividen a la muestra en 100 zonas iguales.



Correspondencia entre cuartiles, deciles y centiles.

El procedimiento de cálculo es el que hemos utilizado para el apartado de la Mdn (C_{50}) que se correspondería con Q_2, D_5 y C_{50}. Lo único que habría que variar es el procedimiento de cálculo del lugar o posición. $Q_q=(q \cdot N)/4$; $D_d=(d \cdot N)/10$; $P_p=(p \cdot N)/100$. Lo más sencillo esa adoptar el procedimiento de cálculo de los centiles.

Convendría hacer notar que, a veces, de se manejan estos mismos términos cuando se trabaja con la curva normal sobre todo en el caso de los centiles y percentiles. Por lo que, habrá que especificar si estamos trabajando con la curva normal donde las unidades (distancias) son constantes (escala Z) o no. Desde nuestro punto de vista, debería hablarse de centiles cuando trabajemos con puntuaciones ordinales y de percentiles cuando trabajemos con curva normal. Sin embargo, no hay un acuerdo al respecto.

En general, parece conveniente concentrarse en el cálculo de los centiles, dado que se puede buscar su correspondencia con el resto de los índices.

Hay que insistir en que las valoraciones individuales no tienen propiedades aritméticas plenas. Por ello, no es legítimo efectuar operaciones con estos valores; por ejemplo, si un sujeto ha obtenido en la asignatura A una puntuación igual al centil 55, en la B el centil 80 y en la C el 48 y se desea obtener el promedio de su rendimiento, no se deberán sumar los valores y dividir por 3. Habría que transformar los centiles en valores y, si las tres puntuaciones están en la misma escala, promediarlas pero nunca desde el centil que sólo nos indica posición en una serie ordenada.

6.3.- Puntuaciones cuantitativas

6.3.1.- Puntuaciones diferenciales: $x = X - \overline{X}$

Presentan la ventaja, respecto a las puntuaciones directas, de que el valor se interpreta en base a la media del grupo y mantiene las mismas unidades de medida que la variable a la que representa. Indica el número de unidades de medida que un valor se aparta de la media. Y, como ya sabemos, la suma de todas las puntuaciones diferenciales de los sujetos de una muestra es igual a cero.

+ > 0: superior a la media

- < 0: inferior a la media

0: igual a la media

Sin embargo, este tipo de puntuaciones -pese a servir de base para un gran número de cálculos estadísticos- no permiten la comparación intergrupos.

6.3.2.- Puntuaciones típicas o estándar: Z

$$Z = \frac{X - \overline{X}}{\sigma}$$

Las puntuaciones típicas son una transformación lineal de las puntuaciones directas a una escala con $\overline{X}$ =0 y σ=1 por lo que es independiente de la escala original. A continuación ofrecemos la demostración de esta afirmación.

$$\overline{Z} = \frac{\sum Z}{N} = \frac{\sum \frac{X - \overline{X}}{\sigma}}{N} = \frac{\sum X - \overline{X}}{\sigma N} = \frac{0}{\sigma N} = 0$$

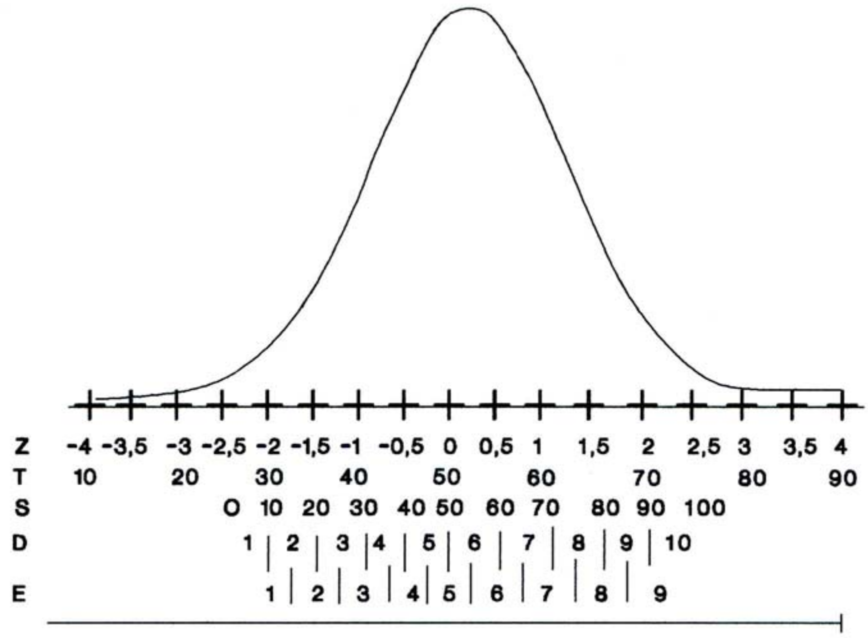
$$\sigma_Z = \sqrt{\frac{\sum (Z-\overline{Z})^2}{N}} = \sqrt{\frac{\sum Z^2}{N}} = \sqrt{\frac{\sum (X-\overline{X})^2/\sigma^2}{N}} = \sqrt{\frac{\sum (X-\overline{X})^2/N}{\sigma^2}} = \sqrt{\frac{\sigma^2}{\sigma^2}} = 1$$

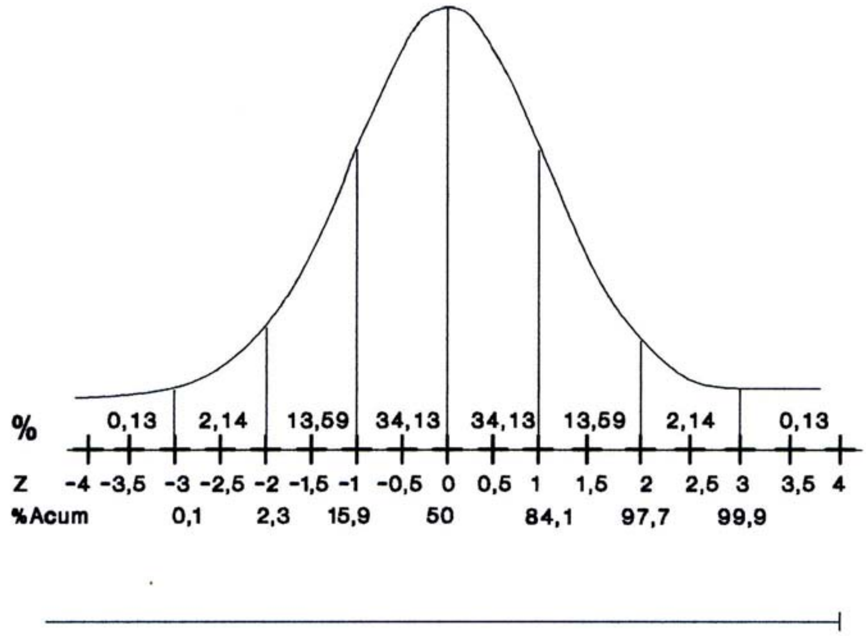
Las puntuaciones típicas permiten la comparación entre varios sujetos (intersujetos) o entre distintas variables del mismo sujeto (intrasujeto).

6.3.3.- Puntuaciones típicas derivadas de Z:

Son cualquier puntuación derivada de Z en base al establecimiento de una media y una desviación típica arbitraria. Frente a las puntuaciones Z, ofrecen las ventajas de dar mayor amplitud a los valores, evitándose los números negativos y con menor dependencia de los decimales.

ESCALA	MEDIA	DESVIACIÓN	FORMULA
T Abarca ± 5 σ (100%)	50	10	T = 50 + 10z
S Abarca ± 2,5 σ (99,38%)	50	20	S = 50 + 20z
CIs Abarca ± 5 σ (100%)	100	15	CIs = 100 + 15z
ENEATIPOS Abarca ± 2,5 σ (99,38%)	5	2	E = 5 + 2z
DECATIPOS Abarca ± 2,5 σ (99,38%)	5,5	2	D = 5,5 + 2z
PERCENTILES	Puntuación <------> Z <-----> % curva normal		$P_n = \sigma \cdot Z_n + \overline{X}$


Z -4 -3,5 -3 -2,5 -2 -1,5 -1 -0,5 0 0,5 1 1,5 2 2,5 3 3,5 4
T 10 20 30 40 50 60 70 80 90
S 0 10 20 30 40 50 60 70 80 90 100
D 1 2 3 4 5 6 7 8 9 10
E 1 2 3 4 5 6 7 8 9


% 0,13 2,14 13,59 34,13 34,13 13,59 2,14 0,13
Z -4 -3,5 -3 -2,5 -2 -1,5 -1 -0,5 0 0,5 1 1,5 2 2,5 3 3,5 4
%Acum 0,1 2,3 15,9 50 84,1 97,7 99,9

BAREMACIÓN

```
DATA LIST FILE='PATRON.DAT'
 /SEXO 1 VAR1 2_4.
VARIABLE LABELS SEXO "SEXO"
 /VAR1 "PUNT. EN UN TEST".
VALUE LABELS SEXO 1 "VARON" 2 "MUJER".
MISSING VALUES ALL (0).
SET MORE OFF/LENGTH=60/EJECT=ON.
compute Z=(var1_52.325)/21.736.
compute S=RND(20*Z+50).
compute T=RND(10*Z+50).
list var1 Z S T.
The raw data or transformation pass is proceeding
    199 cases are written to the compressed active file.
Number of cases read =     199    Number of cases listed =     199
```

Lo primero que hay que hacer es pasar las puntuaciones directas a típicas. Cada sujeto tendrá los tipos de puntuaciones que hemos pedido.

En base la salida de resultados, podríamos establecer la correspondencia entre las distintas escalas (baremo) de manera que cualquier puntuación directa de otra muestra de sujetos con los que se ha utilizado el mismo test pudiera ser interpretada en base a estos datos. El baremo se iría perfeccionando por acumulación de datos (repetición del proceso presentado) hasta el momento en que tuviéramos cubiertas todas las puntuaciones directas de la variable y, como el N de la muestra sería muy grande, se podría considerar como el patrón de referencia.

VAR1(X)	Z	S	T	VAR1(X)	Z	S	T
1	-2.36	3.00	26.00	51	-.06	49.00	49.00
2	-2.32	4.00	27.00	52	-.01	50.00	50.00
3	-2.27	5.00	27.00	53	.03	51.00	50.00
4	-2.22	6.00	28.00	54	.08	52.00	51.00
9	-1.99	10.00	30.00	55	.12	52.00	51.00
12	-1.86	13.00	31.00	56	.17	53.00	52.00
13	-1.81	14.00	32.00	57	.22	54.00	52.00
15	-1.72	16.00	33.00	58	.26	55.00	53.00
18	-1.58	18.00	34.00	59	.31	56.00	53.00
21	-1.44	21.00	36.00	60	.35	57.00	54.00
23	-1.35	23.00	37.00	61	.40	58.00	54.00
24	-1.30	24.00	37.00	62	.45	59.00	54.00
25	-1.26	25.00	37.00	63	.49	60.00	55.00
26	-1.21	26.00	38.00	64	.54	61.00	55.00
29	-1.07	29.00	39.00	65	.58	62.00	56.00
30	-1.03	29.00	40.00	66	.63	63.00	56.00
31	-.98	30.00	40.00	67	.68	64.00	57.00
32	-.94	31.00	41.00	68	.72	64.00	57.00
33	-.89	32.00	41.00	69	.77	65.00	58.00
34	-.84	33.00	42.00	70	.81	66.00	58.00
35	-.80	34.00	42.00	71	.86	67.00	59.00
36	-.75	35.00	42.00	72	.91	68.00	59.00
37	-.71	36.00	43.00	73	.95	69.00	60.00
38	-.66	37.00	43.00	75	1.04	71.00	60.00
39	-.61	38.00	44.00	76	1.09	72.00	61.00
40	-.57	39.00	44.00	77	1.14	73.00	61.00
41	-.52	40.00	45.00	79	1.23	75.00	62.00
42	-.48	40.00	45.00	80	1.27	75.00	63.00
43	-.43	41.00	46.00	81	1.32	76.00	63.00
43	-.43	41.00	46.00	82	1.37	77.00	64.00
44	-.38	42.00	46.00	83	1.41	78.00	64.00
45	-.34	43.00	47.00	84	1.46	79.00	65.00
46	-.29	44.00	47.00	86	1.55	81.00	65.00
47	-.24	45.00	48.00	87	1.60	82.00	66.00
48	-.20	46.00	48.00	88	1.64	83.00	66.00
49	-.15	47.00	48.00	93	1.87	87.00	69.00
50	-.11	48.00	49.00	95	1.96	89.00	70.00
				97	2.06	91.00	71.00
				99	2.15	93.00	71.00
				100	2.19	94.00	72.00

Media = 52,325
σ = 21,736 y N = 197

Deberíamos calcular las puntuaciones X que faltan -por ejemplo X = 8- mediante la fórmula ya expresada. Por el mismo procedimiento se podrían establecer baremos por sexos, edades, cursos...

EJERCICIO

Calcular todos los estadísticos que se indican para MAT

Valorar la asimetría y curtosis (Tablas al 1% y 5%)

Calcular las PSMAT del sujeto 21 e interpretarlas en %

Que porcentaje deja por debajo una z=1,25

Qué z deja por debajo un 77%

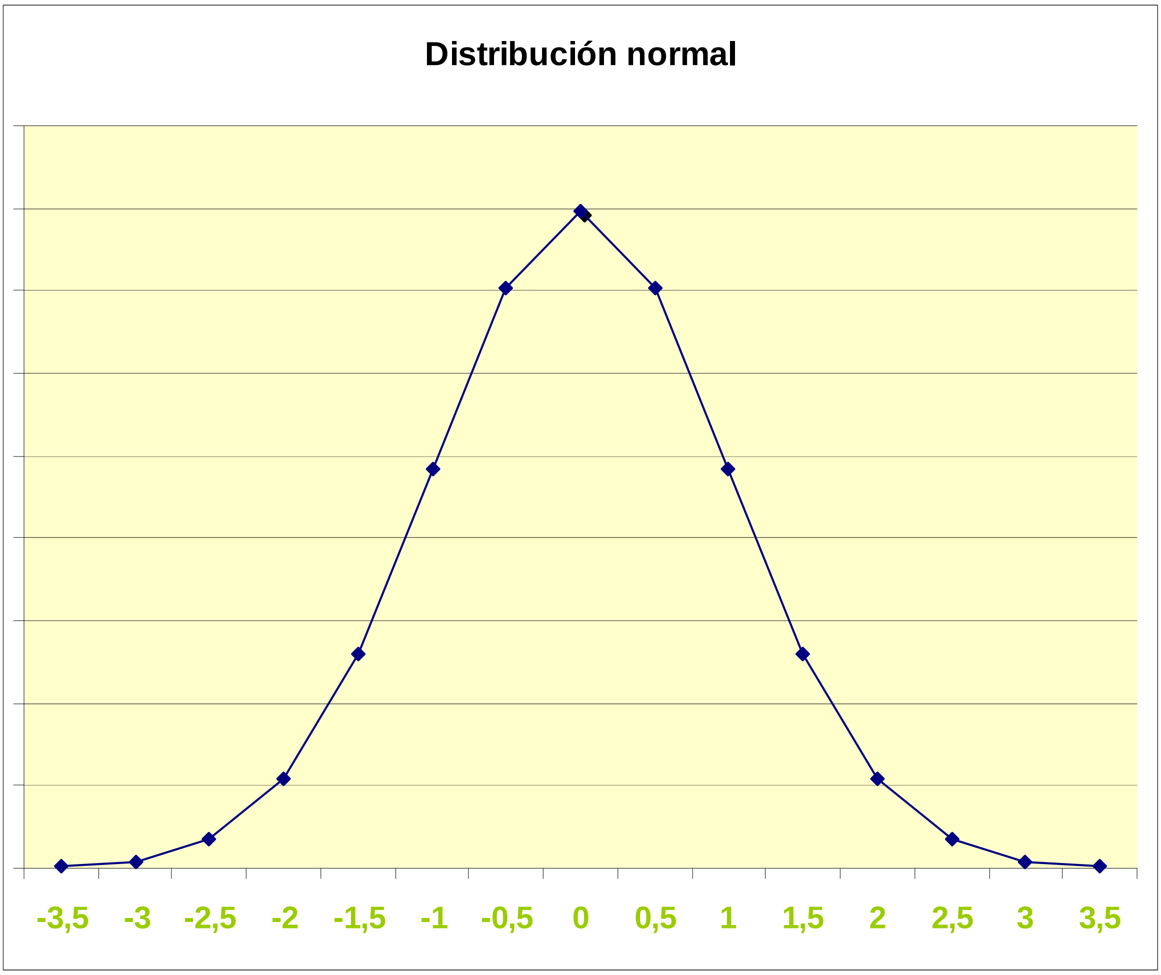
Este ejercicio se puede localizar en ejercicios apuntes.xls, hoja ejercicio datos.

Nº	EDA	SEX	MAT	LEN		psMAT						
1	19	2	8	6								
2	21	2	7	9								
3	20	2	5	6								
4	19	2	5	5								
5	24	2	7	6								
6	21	2	6	7								
7	23	2	5	8								
10	22	2	7	5								
11	20	2	8	8								
12	21	2	6	5								
13	27	1	7	7								
14	20	1	6	9								
15	20	2	5	8								
16	21	2	6	5								
22	19	2	6	7								
23	19	2	6	8								
19	22	1	5	5								
20	21	1	5	7								
21	20	1	6	7								

suma datos	**399**		**116**	**127**
N	**19**	**19**	**19**	**19**
asimetria (As5)	**1.676**		**0.528**	**0.155**
curtosis (K2)	**3.528**		**-0.610**	**-0.922**
Q1	**20.000**		**5.000**	**5.500**
Q2	**21.000**		**6.000**	**7.000**
Q3	**21.500**		**7.000**	**7.500**
Mdn	**21.000**		**6.000**	**7.000**
Promedio	**21.000**		**6.105**	**6.684**
desv.(n-1)	**2.000**		**0.994**	**1.325**
Moda	**21.000**		**6.000**	**5.000**
P10	**19.000**		**5.000**	**5.000**
P90	**23.200**		**7.200**	**8.200**

Correspondencia entre punt. Típicas y percentiles

Z	PS	Perc	Z	PS	Perc
-2.50	0.00	1	0.00	50.00	50
-2.25	5.00	1	0.25	55.00	60
-2.00	10.00	2	0.50	60.00	69
-1.75	15.00	4	0.75	65.00	77
-1.50	20.00	7	1.00	70.00	84
-1.25	25.00	11	1.25	75.00	89
-1.00	30.00	16	1.50	80.00	93
-0.75	35.00	23	1.75	85.00	96
-0.50	40.00	31	2.00	90.00	98
-0.25	45.00	40	2.25	95.00	99
			2.50	100.00	99



PT	10	20	25	30	35	40	45	50	55	60	65	70	75	80	85
PS			**0**	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**		
%Ac		**0,1**	**1**	**2,3**	**7**	**15,9**	**31**	**50**	**69**	**84,1**	**93**	**97,7**	**99**	**99,99**	