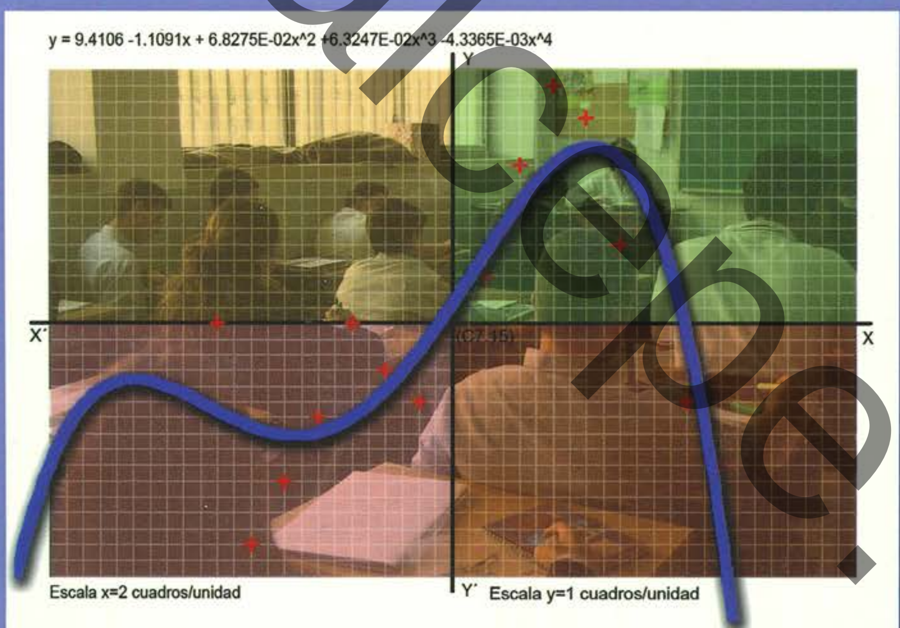
José María Maíllo

ESTADÍSTICA APLICADA A LAS CIENCIAS HUMANAS

(con CD-ROM)



ÍNDICE

1ª Parte. PSICOMETRÍA

Primera Parte

PSICOMETRÍA

Lección 1

METODOLOGÍA CIENTÍFICA
de la Estadística

A. NOCIONES ESTADÍSTICAS

1. INTRODUCCIÓN. MÉTODO EXPERIMENTAL

Marco. Dentro del "carácter *científico* de esta materia".

La **ciencia estadística** trata de aplicar el *método científico experimental* a la conducta y al comportamiento humano. La medida y expresión cuantitativa de variables humanas ha sido una meta en la historia de esta ciencia.

Muchas cualidades humanas, individuales o de grupo, puedan ser objeto de medida (al menos relativamente, por relación a otros criterios o puntos de referencia). Los periódicos nos tiene acostumbrados a una serie de estadísticas sobre el comportamiento humano. Ejemplo: el porcentaje de votante en unas elecciones, el % de consumidores, accidentes, resultados de las encuestas de opinión pública.

A nosotros lo que nos importa es que los futuros profesores que se educan aquí dispongan de una base de conocimientos de *estadística* suficiente para que ellos puedan entender los datos psico-métricos y socio-métricos aplicados por los especialistas en el campo de enseñanza, (por ejemplo a los resultados de tests, encuestas sociogramas), y para que puedan aplicar ellos mismos con rigor científico algunos modelos operativos sencillos. Veamos:

a. Ordenación de datos.

La simple ordenación de datos estadísticos puede decirnos mucho acerca de la variable estudiada, o del grupo de dónde fue obtenida. Ejemplo: La estatura y el peso de los niños *varones* de una determinada edad (11-12 años) es muy diferente de la talla y peso de las *niñas* de la misma edad.

La posición relativa de un alumno con respecto al grupo, cuando se mide una determinada cualidad (por ejemplo la inteligencia) puede ser muy expresiva respecto del éxito futuro en los estudios.

La construcción de gráficos, con aquellos datos ordenados, puede mostrarnos cómo un todo, el grado de homogeneidad del grupo que estudiamos: Necesitamos saber confeccionar **histogramas, polígonos de frecuencias, pictogramas**...

b: Diseños experimentales

Desde Galileo se ha buscado el modo de controlar las variables que intervienen en un proceso. Se busca mantener quieto el cuadro general de variables, modificando una sola de ellas: variable experimental.

Las modificaciones que se presenten habrá que atribuirlas a dicha variable experimental.

B. MEDIDAS DE TENDENCIA CENTRAL:

1. MEDIA ARITMÉTICA. MÉTODOS DE CÁLCULO.

Los datos pueden ser recogidos directamente, a través de un simple recuento.

Esto es lo que los ordenadores hacen, y lo que nosotros nos vemos obligados a hacer cuando queramos hallar la **media aritmética** o la **desviación típica**, por medio de una pequeña calculadoras de bolsillo.

También podemos utilizar un segundo método que consiste en ***agrupar las frecuencias*** en intervalos- (Cf. infra los *métodos abreviados de la Estadística*).

También nosotros podemos ordenar los datos desde el valor más pequeño hasta el más grande. (Este es el *Método Ordinal*, que resulta siempre costoso y es poco práctico).

Empezaremos por un ejemplo. Supongamos que un alumno ha obtenido estas calificaciones:

Lengua	6
Matemáticas	7
Área Social	5
Inglés	6
Deportes	6

$\Sigma x = 30$

SIMBOLOS.- A cada puntuación directa, llamamos "X".
- A la suma total Σx (= Suma de "x")
- Al número de asignaturas: $N = 5$
- A la Media Aritmética $\overline{X}$

Fórmula $$\overline{X} = \frac{\Sigma x}{N} = \frac{30}{5} = 6$$

a) DEFINICIÓN de Media Aritmética. Método 1º, de la "Media simple"

La "media aritmética" es el promedio o valor medio de todas las puntuaciones.

Interpretación: Eso significa que las puntuaciones recontadas oscilan en torno al valor (6), y que unas puntuaciones estarán por encima, pero la cantidad que sobrepasan a la media, se compensa con las otras puntuaciones que están por debajo.

ACTIVIDAD 1.1

Comprobarlo. Aquí hay un número 7 que excede un punto por encima de la Media Aritmética (+1), y hay un número 5 que vale un punto por debajo de la Media (-1).
Total = +1 – 1 = 0

Conclusión: Por eso la Media Aritmética es una "Medida de Tendencia Central".

b) Media Ponderada:

Como no sería justo valorar lo mismo a las calificaciones obtenidas por el alumno de la "Tabla Numero 1" en asignaturas fáciles que en las difíciles, esto es por lo que podemos dar un "peso" o "densidad" diferente a cada asignatura. (Estos pesos también son llamados **frecuencias** (= f)).

$$\overline{X}=(\Sigma x_i)/N$$

TABLA 1,1

	x	"f"="p"	x.f
LENGUA	8	3	24
MATEMÁTICAS	7	3	21
Área SOCIAL	5	2	10
INGLÉS	6	1	6
DEPORTES	4	1	4
N = Σ = 10			**65 = Σ(x.f)**

Ahora N = 10; Esto es como si cada Asignatura que ha sido valorada o "pesada" con un peso 2 valiera el doble, y la que recibió el peso 3, como si fueran tres asignaturas.

La suma de los pesos o frecuencias (=10) pasa a ser el número de asignaturas (= N), y por tanto es el nuevo denominador, por eso la fórmula de la media ahora es:

$$\overline{X}_p = \frac{\Sigma\,(f \cdot x)}{\Sigma f} = \frac{65}{10} = 6{,}5$$

NOTA: Vd. puede observar que esta Media Ponderada $\overline{X}_p = 6{,}5$ es distinta que la Media Simple =5

ACTIVIDAD 1.2

¿Cómo hallarías la media de las calificaciones obtenidas en tres evaluaciones sucesivas, cuyos resultados fueran los siguientes:

TABLA 1,2

	1ª Evaluación X	2ª Evaluación Y	3ª Evaluación Z
LENGUA	8	8	9
MATEMÁTICAS	7	6	7
AREA SOCIAL	5	6	8
INGLÉS	6	7	7
DEPORTES	4	5	5

Nota: Tu puedes seguir varios métodos. Debes justificarlos, y comparar los resultados.

Consecuencias: De aquí ya se pueden deducir algunas conclusiones en cuanto a los criterios de evaluación que se estudian en Pedagogía, así como sobre las leyes de cada estado para valorar a los alumnos.

ACTIVIDAD 1.3

Busca en las leyes de educación de tu país la escala de valoración de los alumnos en los exámenes.

Determina el valor mínimo para aprobar un examen. Explica qué sucede si se coloca en ese valor la media aritmética.

ACTIVIDAD 1.4

Calcula la media aritmética para estos valores:2, 4, 6, 8, 9, 11, 13, 15, 8, 6, 7

c) Métodos abreviados para el cálculo de la media
(Mediante "*acumulación de frecuencias en intervalos*")

• *Segundo método: A través de los* ***puntos medios****:*
Cuando se trata de muchos datos, y con cifras grandes (incluso con decimales) el método anterior se complica, es causa de errores etc.
Se empieza determinando la *amplitud, recorrido o rango*

Fórmula de la amplitud a recorrido:
A = Valor Máximo - Valor mínimo + 1 ·

En ese caso: A = 14 - 2 + 1 = 13.
Luego fijamos el número de intervalos , v.g. en este caso $n_i = 4$
Así determinamos el valor de cada intervalo, porque hay una fórmula que dice:

$$i = \frac{A}{n_i} = \frac{13}{4} = 3{,}25 \cdot \text{ como debe ser nº entero=4}$$

Empezamos a construir la escala de intervalos empezando por el valor más pequeño: en este caso el 2; pero como nos sobraría mucho en el último intervalo, empezamos por una unidad menos para que quede más centrado el espacio muestral de los datos dentro de esta escala artificial que estamos confeccionando.

Extensión de cada intervalo.
Si nos fijamos cada intervalo empieza una unidad por encima del anterior; por eso para no perder una unidad entre intervalo e intervalo, se estima que cada intervalo termina 0'5 punto por encima de su valor máximo y empieza -0'5 puntos por debajo de su valor mínimo.

(**Nota pedagógica:** Esto se puede expresar con el símil de las parcelas de terreno de un pueblo separadas por las piedras miliares que indican los Kms. Entre el km 4 y el km 5 se perdería un km. Por eso el término municipal de dos pueblos próximo separados por los kms 4 y 5 estará en la mitad de ese km. O sea que el término municipal de pueblo "A" empieza en 4,5 km y el término del pueblo "B" sigue a partir de ese km 4,5.

En la **práctica**: Cada intervalo empieza 0,5 puntos por debajo del número menor de ese intervalo. Ej. El intervalo IIº empieza en 5 – 0,5 = 4,5; y el intervalo Iº termina en 4 + 0'5. Luego el ***límite*** entre ambos intervalos es 4,5

Punto medio de cada intervalo = x_m

Como cada intervalo se extiende desde -0'5 puntos por debajo de su límite inferior, y llega a hasta +0'5 puntos por encima de su valor superior, el *punto medio del intervalo* será la "*media aritmética de los extremos*"

$$x_m = \frac{(\text{Valor inferior } -0{,}5) + (\text{Valor superior } + 0{,}5)}{2}$$

Veamos un ejemplo

Supongamos que hemos hecho un examen de 16 preguntas a 10 alumnos (podían ser 100, o bien 1.000 alumnos...): Puntuaciones: 2, 14, 5, 7, 10, 6, 12, 13, 8

Por el método 1º de la **media simple**: $\overline{X} = \frac{\Sigma(x)}{N} = \frac{88}{10} = 8{,}8$

Por el **método 2º, abreviado** de los puntos medios x_m

TABLA 1,3

	x_m	f	$f \cdot x_m$
13-16	14,5	2	29
9-12	10,5	3	31,5
5-8	6,5	4	26
1-4	2,5	1	2'5

Observación: Nótese cómo en el método abreviado se comete un pequeño error, debido a que no anotamos la puntuación directa, sino que asignamos a todas la frecuencias de cada intervalo el valor del **punto medio** de ese intervalo (x_m).

Lo cual no deja de ser arbitrario por nuestra parte; pero este es el secreto de la estadística, "***que juega con el error***": es verdad que probablemente ninguno de los alumnos ha obtenido la puntuación del punto medio del intervalo.

También es cierto que tratándose de muchos alumnos (más de 100) se compensarán por de más o por de menos, los alumnos de cada intervalo a los que hayamos asignado el valor del punto medio del intervalo x_m

ACTIVIDAD 1.5

Calcula la media aritmética por el método abreviado para los valores de la actividad 4

$$\overline{X} = \frac{\Sigma f \cdot x_m}{N} = \frac{89}{10} = 8{,}9$$

Error: Se ve que hay una diferencia de **una décima** con respecto al método de la **media simple**, (esta no tiene error)

INTERPRETACIÓN : El error procede de haber atribuido a todas las frecuencias de un intervalo el valor del punto medio de ese intervalo. Lo cual no es cierto; pero se aproxima tanto más a la realidad cuanto mayor sea el número de la muestra.

Por la LEY DE LOS GRANDES NÚMEROS, sabemos que unos datos se tomarán en exceso con respecto al valor medio del intervalo, otros se tomarán por DEFECTO; pero unos y otros se compensan si la muestra es grande (N > 100)

- ***Tercer método** para calcular la **media** (a través de la media supuesta)*

El **método de la media supuesta** ($\overline{X}$s) o método abreviado a través de la "*x prima*" (: x') consiste en suponer conocida la media, y suponer que ésta está situada en el punto medio de un intervalo (el que queramos). A ese intervalo lo numeramos con el valor ***cero*** (x' = 0).

Formamos así una nueva columna que llamamos x', cuyos valores son: *cero* para el intervalo en que hayamos *supuesto* pueda contener la **media**.

Los intervalos de valores más altos los numeramos con valores correlativos +1, +2, +3, etc.

Los valores menores de la variable x los numeramos correlativamente -1, -2, -3, -4 etc. (Ver la tabla 1,4)

TABLA 1,4

Intervalos		x_m	f	x'	f · x'
13-16			2	+1	2
9-12	**Media supuesta $\overline{X}$s**	10,5	3	0	0
5-8			4	-1	-4
1-4			1	-2	-2
Σ					-4

$$\overline{X} = \overline{X}s + (i \cdot \frac{\Sigma f \cdot x'}{N}) = 10{,}5 + (i \cdot \frac{(-4)}{10} = 10{,}5 + (4 \cdot (-0{,}4)) = 10{,}5 - 1{,}6 = 8{,}9$$

ACTIVIDAD 1.6

Confeccionar un FORMULARIO

C. REPRESENTACIONES GRÁFICAS

1. HISTOGRAMA.

Situamos un eje XX' y marcamos en él los límites de los intervalos de la Tabla IV. Luego elevamos un rectángulo tan alto como sea el número de frecuencias de cada intervalo (para eso asignamos previamente un *valor unidad* a cada frecuencia, (v.g. U = 1 cm = 1 f)

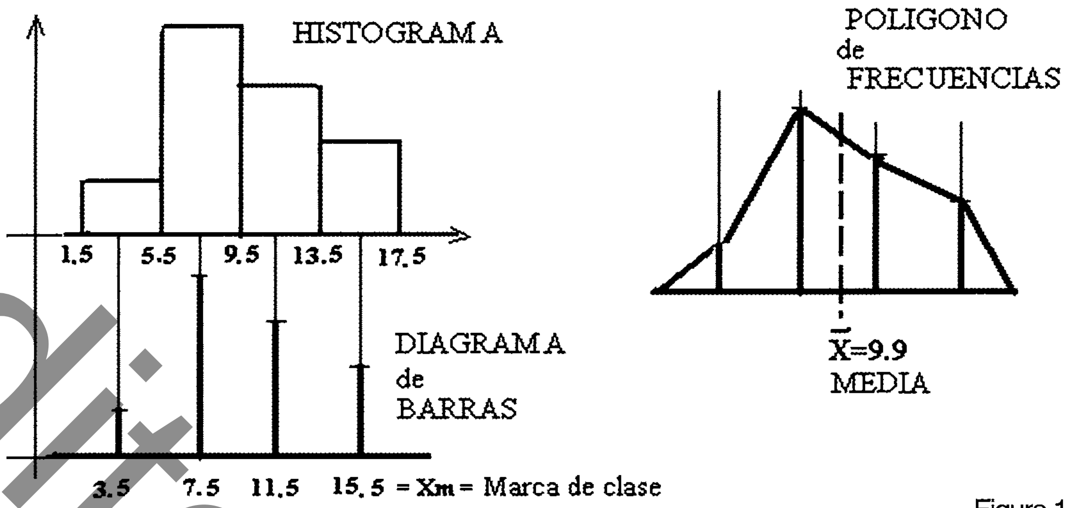


Figura 1,1

2. DIAGRAMA DE BARRAS

Partimos también de la Tabla IV. Dibujamos el eje XX'; pero ahora lo dividimos según los valor de los puntos medios (x_m) y en cada una de esas marcas elevamos una barra que tenga tanta altura como el número de frecuencias de cada intervalo. (También asignamos una *unidad* a cada frecuencia)

3. POLÍGONO DE FRECUENCIAS

Basta unir los extremos de las barras del *"Diagrama de Barras"*.

ACTIVIDAD 1.7

Sean 25 alumnos, cuyo examen de 25 preguntas, una vez corregido, dio estos resultados de preguntas bien respondidas por cada alumno:

TABLA 1,5

Alumno	Punt	Alumno	Punt	Alumno	Punt	Alumno	Punt	Alumno	Punt
A	20	**F**	10	**K**	17	**Ñ**	5	**S**	2
B	4	**G**	10	**L**	11	**O**	16	**T**	14
C	7	**H**	6	**LL**	14	**P**	13	**U**	11
D	12	**I**	13	**M**	12	**Q**	9	**V**	15
E	15	**J**	8	**N**	21	**R**	15	**W**	13

Hallar la **media** agrupando frecuencias.

Lección 2

LA VARIABILIDAD

A. MEDIDA DE LA VARIABILIDAD: LA VARIANZA

La **varianza** es "el momento de 2° orden = "*media cuadrática de las desviaciones*"
La **desviación típica** o "desviación estándar" es la raíz cuadrada de la **varianza.**

TABLA 2,1. FÓRMULAS

PUNTUACIÓN DESVIADA o DESVIACIÓN.

$$dx = Xi - \overline{X}$$

1) **desviación tìpica** simple

$$\sigma = \sqrt{\frac{\Sigma\,(dx)^2}{N}} =$$

NOTA – OTRAS FÓRMULAS de la DESVIACIÓN TÍPICA :

a) Como dx =(Xi - $\overline{X}$)

$$\sigma = \sqrt{\frac{\Sigma\,(X_i - \overline{X})^2}{N}} =$$

b) **desviación típica** cuando hay **frecuencias**

$$\sigma = \sqrt{\frac{\Sigma\ f \cdot (X_i - \overline{X})^2}{N}} =$$

2) **desviación típica** con **puntuaciones agrupadas** (puntos **medios** y **frecuencias** "***f***")

$$\sigma = \sqrt{\frac{\Sigma\, f(x_m - \overline{X})^2}{N}} = \sqrt{\frac{\Sigma\, f{\cdot}(dx_i)^2}{N}}$$

3) **desviación típica** a través de la "***x prima***" (x') (cuando hay ***media supuesta***)

$$\sigma = i \cdot \sqrt{\frac{\Sigma\, f{\cdot}x'^2}{N} - \left[\frac{\Sigma\, f{\cdot}x'}{N}\right]^2}$$

ACTIVIDAD 2.1

Poner un ejemplo de dos hermanos que tengan la misma media; pero que en una de ellos haya mucha variabilidad en las puntuaciones. No se los puede enjuiciar a los dos por solo la **media**. No es verdad que sean iguales

ACTIVIDAD 2.2

Calcular la Desviación Típica con los datos de la Tabla 2.2

TABLA 2,2

	X	**$X - \bar{X} = dx$**	**$(dx)^2$**
LENGUA	8	8-6=+2	4
MATEMATICAS	7	7-6=+1	1
AREA SOCIAL	5	5-6=-1	1
AREA PLASTICA	6	6-6=0	0
AREA DINAMICA	4	4-6=-2	4
Σ		0	10

ACTIVIDAD 2.3

Poner las fórmulas, y calcular:
- La MEDIA:

- Halla la SIGMA o DESVIACIÓN TÍPICA: D.T.= sigma = $\sigma = \sqrt{((10)/5))} = 1{,}4142$

B. CÁLCULO DE LA DESVIACIÓN TÍPICA CON FRECUENCIAS O *PONDERACIONES*

Los valores de cada puntuación vienen afectados por los valores del **peso** atribuido a cada puntuación. En el caso de **notas escolares**, no parece objetivo darle la misma importancia a las asignaturas ***fáciles*** que a las ***difíciles***. Por eso se le da un **peso** cualitativa más alto a las asignaturas difíciles.

(NOTA: Ese **criterio** de valoración, por el que damos más importancia a unas asignaturas que a otras, ya se ve que es ***filosófico*** y por tanto es ***anterior y superior*** al *criterio estadístico*)

Recurrimos a la fórmula (5)

Vemos un ejemplo en la tabla 2,3

TABLA 2,3

	X	F	x.f	X	$dx=X_i-\overline{X}$	$(dx)^2$	$f\cdot dx^2$
LENGUA	8	3	24	6	8-6=2	4	12
MATEMATICAS	7	3	21	6	7-6=1	1	3
A.SOCIAL	5	2	10	6	5-6=-1	1	2
A.PLASTICA	6	1	6	6	6-6= 0	0	0
A.DINAMICA	4	1	4	6	6-4=-2	4	4
Σ	30	10	65		0	10	21

ACTIVIDAD 2.4

Hallar MEDIA y DESVIACIÓN TÌPICA.

–CALCULO de la **desviación típica** con agrupación de frecuencias en intervalos y a través de los **puntos medios** ($=x_m$)

Ej: Tomemos los datos de la tabla 2.4

TABLA 2.4

i=4	x_m	f	$f\cdot x_m$	MEDIA	$x_m-\overline{X}=dx$	$(dx)^2$	$f\cdot(dx)^2$
13-16	14,5	2	29	8,9	5,6	31,36	62,72
9-12	10,5	3	31,5	8,9	1,6	2,56	7,58
5-8	6,5	4	26	8,9	-2,4	5,76	23,04
1-4	2,5	1	2,5	8,9	-6,4	41	41
Σ		N=10	89				134,3

ACTIVIDAD 2.5

Hallar **media** y **desviación tìpica**.

$$\overline{X}=\frac{\Sigma f\cdot x_m}{N}=\frac{89}{10}=8,9$$

$$,\quad \sigma=\sqrt{((\Sigma f(x_m-\overline{X})^2)/N}$$

$$,\quad \sigma=\sqrt{(134,3/10)}=\sqrt{13,43}=3,66$$

C. MÉTODO TERCERO PARA CALCULAR LA DESVIACIÓN TÍPICA CON AGRUPACIÓN EN INTERVALOS NUMERADOS POR LAS X' (=MÉTODO DE LA "MEDIA SUPUESTA" A TRAVÉS DE LA X')

Se llama método de la "**media supuesta**"; porque suponemos resuelto ya el problema y que ha dado, para la media, un valor que nosotros **suponemos ya conocido**, *a priori*. En este caso lo más cómodo es tomar el punto medio de un intervalo; y a ese intervalo lo numeramos con el valor **cero**, luego vamos numerando los demás intervalos con números correlativos, de modo que sean número **positivos** si procedemos ascendentemente, y **negativos** si vamos descendiendo (así creamos la columna x'.

Ejemplo: Supongamos que la **media** está en el *punto medio* (x_m) del intervalo 5 – 8, o sea en 6,5:

ACTIVIDAD 2.6

Calcularemos la media supuesta con los datos de la tabla 2,6

TABLA 2,5

1=4	f	x'	f · x'	f · x'²	x_m
16-13	2	2	4	8	
12-9	3	1	3	3	
5-8	4	0	0	0	6,5= Media Supuesta
4-1	1	-1	-1	1	
Σ	10		6	12	

("*Nos ha parecido bien*" situal la **media supuesta** en el punto medio del intervalo (5 - 8), el cual vale $x_m = 6'5$. Pero podíamos haberla supuesto en otro cualquiera.

$$\overline{X} = 6'5 + 4(6/10)= 6,5 + 2,4 = 8,9$$

ACTIVIDAD 2.7

-Situar la "**media supuesta**" en otro intervalo y comprobar que da el mismo resultado.
-Hallar **media** y **desviaciòn típica** por varios métodos.

$$\sigma = i \cdot \sqrt{((\Sigma f\cdot(x')^2)/N-((\Sigma f\cdot x')/N)^2)}$$

$$\sigma = 4 \cdot \sqrt{((12/10)-((6/10)^2))} = 4 \cdot \sqrt{((1,2)-0,36)}= 4 \cdot 0,916=3,66$$

ACTIVIDAD 2.8

Ejercicio de cálculo de "**media**" y "**desviaciòn típica**" por varios métodos

TABLA 2,6

·i =5	f							
46-50	3							
41-45	5							
36-40	8							
31-35	15							
26-30	25							
21-25	35							
16-20	30							
11-15	15							
6-10	10							
1-5	4							
Σ								

$$\sigma = \sqrt{((\Sigma f\cdot(x_m-\overline{X})^2)/N))}$$

$$\sigma = i\cdot\sqrt{((\Sigma f\cdot(x')^2)/N)-((\Sigma f\cdot x')/N)^2)}$$

D. DEMOSTRACIÓN DE LA TERCERA FÓRMULA DE LA DESVIACIÓN TÍPICA

DEMOSTRACIÓN de la Fórmula de SIGMA por el Método de la Media Supuesta
Partimos del concepto de VARIANZA = Momento de 2° Orden

$$\sigma = \sqrt{\frac{\Sigma\ f\,.\,dx^2}{N}} = \sqrt{\frac{\Sigma\ f\,.\,(x_m - \overline{X})^2}{N}}$$

El paréntesis contiene una DIFERENCIA AL CUADRADO

$$\sigma = \sqrt{\frac{\Sigma\ f\,(x_m^2 + \overline{X}^2 - 2\,x_m\overline{X})}{N}} = \sqrt{\Sigma f\left(\frac{x_m^2}{N} + \frac{\overline{X}^2}{N} - \frac{2.x_m\overline{X}}{N}\right)}$$

Deshacemos N FACTOR COMÚN

Introducimos Σ, que es Factor Común, dentro del paréntesis

$$\sigma = \sqrt{f\,.\left(\frac{\Sigma\,x_m^2}{N} + \frac{\Sigma\,\overline{X}^2}{N} - \frac{\Sigma\,2.x_m\overline{X}}{N}\right)}$$

Aplicamos la propiedad de los SUMATORIOS
Σ f= N

$$\sigma = \sqrt{\frac{\Sigma f\ x_m^2}{N} + \frac{\Sigma f\ \overline{X}^2}{N} - \frac{\Sigma f\ \ 2.x_m\overline{X}}{N}}$$

En el segundo término simplificamos; porque $\Sigma f = N$

$$\sigma = \sqrt{\frac{\Sigma f\ x_m^2}{N} + \frac{\cancel{N}\ \overline{X}^2}{\cancel{N}} - \frac{\Sigma f\ \ 2.x_m\overline{X}}{N}}$$

En el tercer término vemos que $\frac{\Sigma f\ .x_m}{N}$ = ***la MEDIA*** $= \overline{X}$

$$\sigma = \sqrt{\frac{\Sigma f\ x_m^2}{N} + \overline{X}^2 - 2\overline{X}\ \overline{X}} = \sqrt{\frac{\Sigma f\ x_m^2}{N} + \overline{X}^2 - 2\ \overline{X}^2} = \sqrt{\frac{\Sigma f\ x_m^2}{N} - \overline{X}^2}$$

Substituimos la $\overline{X}$ por su valor $= \frac{f\ .\ x_m^2}{N}$

$$\sigma = \sqrt{\frac{\Sigma\ f.\ x_m^2}{N} - \left(\frac{\Sigma\ f\ .\ x_m}{N}\right)^2} =$$

Substituyamos $f x_m$ por $i\ .fx'$
(los x_m se desplazan según múltiplos de i)

$$\sigma = i\ .\sqrt{\frac{\Sigma\ f.\ x'^2}{N} - \left(\frac{\Sigma\ f\ .\ x'}{N}\right)^2}$$

E. OTRAS MEDIDAS DE LA VARIACIÓN. DESVIACIÓN MEDIA

Noción: Es la media de las diferencias (en valor absoluta) respecto de la MEDIA

Fórmula:

$$D.M. = (|X_1-\overline{X}| + |X_2 - \overline{X}\ | + |X_3 - \overline{X}|....) / N$$

O bien:

$$D.M. = (\Sigma\ |Xi - \overline{X}|\ ) / N$$

1. CON DATOS NO AGRUPADOS

Ejemplo: Los tiempos de reacción de las 10 ratas en un “Campo experimental” han sido, expresados en segundos:

25, 1,43, 20, 1, 8,56, 16,5, 34, 4,43, 44, 1º hallamos la **media**:

$$\overline{X} = (\Sigma Xi) / N = 17{,}09$$

2º DM

$$DM = \frac{|25 - 17{,}09| + |1{,}43-17{,}09| + |20-17{,}09|+|1-17{,}09|+|8{,}56-17{,}09|\ldots}{10}$$

DM = 10,926 segs.

2. CON DATOS AGRUPADOS:

Sea la tabla de datos 2,7

TABLA 2,7

Tiempo	f	x_m	$f \cdot x_m$	dx	$f \cdot dx$
31-45	2	38	76	18	36
16-30	4	23	92	3	12
1-15	4	8	32	12	48
TOTAL	10		200		96

1º Hallamos la **media**

$$\overline{X} = \frac{\Sigma f \cdot x_m}{N} = \frac{200}{10} = 20$$

2º Luego aplicamos la definición de “desviación media”

$$D.M. = \frac{\Sigma f \cdot (x_m - \overline{X})}{N} = \frac{\Sigma f \cdot dx}{10} = \frac{96}{10} = 9{,}6 \text{ segs.}$$

$$D.M. = \frac{\Sigma f \cdot |x_m - \overline{X}|}{N} = \frac{96}{10} = 9{,}6$$

3. MÉTODO ABREVIADO A TRAVÉS DE LA *MEDIA SUPUESTA* ($\overline{X}_s$)

$$D.M. = i \cdot \frac{\Sigma f \cdot |x' - \overline{X}|}{N}$$

NB: Donde $\overline{X}$’ es el factor de corrección de la **media**

TABLA 2,8

Tiempo	f	x'	f · x'	|xi-$\overline{X}$|	f · |xi- $\overline{X}$|
31-45	2	+1	2	18	2'4
16-30	4	0	0	3	0'8
1-15	4	-1	-4	12	3'2
TOTAL	10				6'4

La **media** $\overline{X} = \overline{X}s + i \cdot \dfrac{\Sigma fx'}{N} = 23 + 15 \cdot \left(\dfrac{-2}{10}\right) = 23 - 3 = 20$

Factor de corrección de la **media** $\overline{X} = \dfrac{\Sigma f \cdot x'}{N} = \dfrac{-2}{10} = -0{,}2 = \overline{X}$

$$D.M. = i \,.\, \frac{\Sigma f \cdot |xi - \overline{X}|}{N} = 15 \frac{6{,}4}{10} = 9{,}6 \text{ seg}$$

Nota: Los dos métodos abreviados dan igual y cometen el mismo error; lo cual es lógico, porque el origen del error es e lmismo: se han tomado los mismos intervalos y se les ha adjudicado el mismo valor a las frecuencias de cada intervalo; el del punto medio.

Lección 3

MEDIDAS DE TENDENCIA CENTRAL

A. MEDIANA

La **mediana** es aquel valor que deja bajo sí a la mitad de los casos considerados. Es una medida de **tendencia central**

Definición: Es aquel punto que deja por encima o por debajo al 50 % de los datos o casos.
Tarea PREVIA: Ordenar los datos de **menor a mayor** o viceversa

1. CÁLCULO CON DATOS NO AGRUPADOS

• Ejemplo 1): Cuando el número de casos es **impar**:
Sean los datos: 14, 1, 7, 6, 32
Para calcular la mediana
1°. **Ordenamos** los casos de menor a mayor: 1,6,7,14,32
2°. Contamos hasta la mitad de los casos que hay. Cuando llegamos a la mitad nos paramos: ese es la **mediana**
En este Ejemplo es 7; porque a la derecha del 7 hay dos casos y a la izquierda otros dos.... Luego el 7 está en el medio: deja bajó sí al 50 % y le supera el otro 50 %

• Ejemplo 2) Cuando el número de casos es **par**.
Sean los datos: 34, 2, 16, 7
1°. **Ordenamos** los casos: 2, 7, 16, 32
2°. El **espacio muestral** comprendido entre 7 y 16 contiene a la Mediana. Dividimos ese espacio muestral entre dos y la línea divisoria es la Mediana.

$$(7 + 16) : 2 = 11{,}5 = \text{Mediana.}$$

• Ejemplo 3): Cuando se **repite** el valor central.
Sean los números 15, 2, 4, 11, 6, 6, 9
1°. **Ordenamos** los números: 2, 4, 6 , 6, 9, 11, 15

2°. Contamos hasta el 50 %, que es **tres**, el 4ª sería la Mediana, porque luego vienen otro **tres**; pero como el 4° y el 3° es el número 6, (éste se repite, el **espacio muestral** correspondiente al segundo 6 es el que debemos dividir por la mitad: en esa mitad estará la Mediana.

➔ 5,5 ➔ 6 ➔ 6,5

El espacio muestral del segundo n° 6 va desde 6 — 6,5.
Ese espacio dividido por la mitad da: (6 + 6,5):2 = 6,25 = **mediana**.
Nota: Hemos recurrido a la *geometria*, lo cual es perfectamente válido...

2. CÁLCULO CON DATOS AGRUPADOS

1. Formamos intervalos, y agrupamos las frecuencias en ellos.
2. Formamos otra columna con las frecuencias acumuladas ((fa).
3. Aplicamos el concepto de "Mediana": "*aquel valor que deja bajo si al 50 % de los sujetos*", y, en su virtud,
4. Dividimos "N" por la mitad.
5. Buscamos el intervalo donde se pueda estar ese valor mitad de sujetos.
6. Allí estará la mediana. Lo marcamos.
7. Dividimos el espacio muestral de ese intervalo en partes proporcionales a los sujetos que *faltan o exceden* al 50 % de los sujetos. Ahora podemos hacer dos cosas:

8, a) El valor de ese intervalo que "excede" al 50 % se lo quitamos al limite superior del intervalo en cuestión, o bien

8, b) El valor de ese intervalo que "falta" se lo añadimos al límite inferior del intervalo en cuestión.

Ejemplo. DATOS: 1, 2, 2, 3, 3, 4, 4, 5, 5, 5, 6,7, 7, 7, 8, 8, 8, 9, 10,12
N = 20

1) Agrupamos en intervalos. Límite superior: el extremo más medio punto.
Límite inferior: el extremo menos medio punto.

TABLA 3,1

Límites de los intervalos	· i = 3 INTERVALOS	Frecuencias (f)	Frecuencias acumuladas= (fa)	
Límite superior 12'5 *Límite inferior 9'5*	10-12	2	20	
Límite superior 9'5 *Límite inferior 6'5*	7-9	7	18	
Límite superior 6'5 *Límite inferior 3'5*	3-6	6	11	LUGAR de la MEDIANA = 50% de N = 20 / 2 = 10
Límite superior 3'5 *Límite inferior 0'5*	1-3	5	5	
TOTAL		N=20		

2) Hemos formado la columna con las frecuencias acumuladas (fa). Para eso hemos ido sumando, de abajo arriba, la columna de las frecuencias (f), y lo escribimos en la columna (fa).

3) Aplicamos el concepto de "Mediana": "aquel valor que deja bajo si al 50 % de los sujetos". Hallamos el 50 % de N = 20 /2= 10,

4) Dividimos "N" por la mitad. Es 10.

5) Buscamos el intervalo donde se pueda estar ese valor mitad de sujetos. Vemos que 10 se halla en el intervalo "4-6".

6) Allí estará la mediana. Lo marcamos.

7) Dividimos el espacio muestral de ese intervalo en partes proporcionales a los sujetos que faltan o exceden al 50 % de los sujetos. En ese intervalo las frecuencias acumuladas son 11. Como la mediana debe estar en 10, nos sobra una frecuencias (11-10 = 1) . Como en ese intervalo hay 6 frecuencias, y nos sobra una por encima de la mediana, podemos hacer dos cosas:

8, a) El valor de ese intervalo que "excede" al 50 % se lo restamos al limite superior del intervalo en cuestión.

El espacio muestral del intervalo (4-6) va desde 3,5 hasta 6,5 = 3 unidades que debemos repartir proporcionalmente a los 6 frecuencias que hay en ese intervalo: 3 unidades / 6 frecuencias = 0'5 puntos de unidad por frecuencias.

(Imaginemos que la unidades son Hectáreas de un terreno y que las 6 frecuencias son 6 hermanos que deben repartirse el terreno), e imaginemos que solo la propiedad de **uno** de esos hermanos pertenece al termino municipal que separa su pueblo del municipio siguiente, y que por esa linde pasa la **mediana**).

Como en ese intervalo hay 6 frecuencias, nos sobran 5 frecuencias, es decir, 5 están ya por encima de la mediana. Debemos restar el espacio muestral que corresponde a esas 5 frecuencias del ***limite superior***. (Así: El espacio muestral del intervalo (4-6) contiene 3 unidades. Dividimos esas 3 unidades (hectáreas que mide una parcela heredada) entre los 6 frecuencias (supongamos que son hermanos herederos de la parcela). Le toca a cada hermano 3 Ha / 6 hermanos = 0'5 Ha para cada uno (= para cada frecuencia)**]**.

Y como la Mediana es la linde que separa los términos municipales de dos pueblos vecinos, sólo el terreno de un hermano cae por debajo de la linde (= Mediana) mientras los otros 5 terrenitos están ya por encima de la linde (= Mediana). Por tanto debemos restar esos 5 terrenitos (cada uno de 0,5 Ha) (que suman 5 x 0,5 = 2,5 Ha o unidades del espacio muestral) **al límite superior del espacio muestral (la finca del padre). O sea 6,5 – 2,5 = 4.**

Conclusión: La mediana pasa por el número cuatro -"4"- del espacio muestral

8, b) o bien el valor de ese intervalo que nos "falta" para llegar a la mediana se lo SUMAMOS al límite inferior del intervalo en cuestión.

O sea, debemos **sumar** la propiedad de **un** hermano al límite inferior de ese intervalo, para que haya 10 sujetos por debajo de la Mediana. Sólo el terreno de **un** hermano cae entre los 10 propietarios inferiores a la mediana, y esa parcela mide 0'5 unidades que es lo que debemos añadir al limite inferior del intervalo (que medía 3'5). Por tanto la linde que separa los términos Municipales de los pueblos – que aquí llamamos **mediana**- pasa por el final del espacio 3'5 + 0'5 unidades **= 4**, para que haya N= 10 frecuencias por encima y otras 10 frecuencias por debajo (o sea al 50 %, que es la definición de Mediana).

Con ambos cómputos resulta el mismo valor para la mediana = **4.**

NOTA: Los alumnos suelen pararse en la primera parte de la definición de Mediana (50 %) y no reparan que no es el 50% de las frecuencias lo que nos piden, sino el lugar del espacio muestral que separa el 50 % de las frecuencias.

B. MODA.

La moda es el valor que más veces se repite. Puede que se repitan dos número o frecuencias. Entonces se dice que la distribución es "***bi-modal***". O que haya tres, será "***tri-modal***", etc.

C. MEDIAS GENERALIZADAS.

1. MEDIA GEOMÉTRICA

Noción: La media geométrica $\overline{X}g$ es la raíz enésima del ***producto*** de los "n" valores: $X_1, X_2, X_3....$"

Fórmula:

$$\overline{X}g = \sqrt[n]{(X_1) \cdot (X_2) \cdot (X_3)....}$$

Métodos:
1) Cuando la raíz es múltiplo de 2 se puede calcular mediante la propiedad de los radicales: sucesivas raíces de 2
2) Cuando no suceda así se aplican *logaritmos* a los dos miembros de la igualdad:

$$\text{Log. } \overline{X}g = \text{Log } (\sqrt[n]{(X_1) \cdot (X_2) \cdot (X_3)....})$$

$$\text{Log } \overline{X}g = 1/n \text{ (Log } \sqrt{((X_1) \cdot (X_2) \cdot (X_3)....)}$$

Ejemplo: Sean los Tiempos de Reacción en un test: 22, 11, 7, 7, 28
(Nb: Se ve que son números muy poco agrupados, muy variantes..)

$$\overline{X}\,g = \sqrt[5]{(22) \cdot (11).\ (7) \cdot (7) \cdot (28)}$$

$$\text{Log. } \overline{X} = (1/5).\ \text{Log. } ((22) \cdot (11) \cdot (7) \cdot (7) \cdot (28)) =$$
$$= (1/5) \cdot ((\text{Log } 22) + (\text{Log } 11) + (\text{Log } 7) + (\text{Log } 7) + (\text{Log } 28))$$
$$= (1/5) \cdot (1{,}34 + 1{,}04 + 0{,}84 + 0{,}84 + 1{,}45)$$
$$= (1/5) \cdot 5{,}507 = 1{,}10$$

Colección

PropuestasCurriculares15

Este **tratado de Estadística**, para uso de aquellos que se dedican a las "Ciencias Humanas" (como son Psicología, Pedagogía, Sociología, Psicopedagogía, Medicina, etc.), así como para los alumnos que cursan estudios, está concebido con una intención eminentemente práctica, en orden a aligerar los complicados procesos matemáticos que se necesitan en esas materias y facilitar la expresión gráfica de los mismos.

De ahí que hayamos prescindido de prolijas demostraciones y hayamos procurado la mayor claridad y sencillez de las explicaciones, ayudándonos de ejemplos y expresiones gráficas.

Hemos recurrido a la ayuda de la técnica informática para aliviar a profesores y alumnos de enojosos cálculos. Por eso hemos incorporado una serie de programas informáticos que van en un CD-ROM adjunto.

Nuestra intención pedagógica nos ha llevado a confeccionar los programas informáticos con abundantes etapas intermedias, que el alumno puede obtener y comprobar, paso a paso, porque van explicitadas en los títulos de cada resultado y en explicaciones pertinentes.

También las instrucciones para el uso de esos programas informáticos debían ser claras y explícitas. Por eso hemos incluido en el mismo CD-ROM ficheros de «instrucciones» para el uso correcto de los programas. Incluso se ha dedicado la lección 16 del texto a explicar el uso de esos programas. Con ese fin las gráficas de todas las pantallas de «salidas de ordenador» se han presentado en esa lección 16, a efecto de explicitar los resultados ofrecidos en cada pantalla.

Finalmente diremos que cualquier sugerencia de nuestros lectores será bien recibida.

ISBN 84-7869-554-0
9788478695546

CEPE
CIENCIAS DE LA EDUCACIÓN PREESCOLAR Y ESPECIAL
General Pardiñas, 95 - 28006 Madrid
Telf.: 915626524 - Fax: 915640354
venta@editorialcepe
www.editorialcepe.es