

Temas de Estadística Práctica

Antonio Roldán Martínez

Proyecto <http://www.hojamat.es/>

Tema 2: medidas de tipo paramétrico

Documento

Datos de tipo cuantitativo

Son aquellos que están representados por números. Si entre cada dos datos puede haber una infinidad de ellos, se llaman **continuos**, y si entre un dato y otro siempre hay un hueco o salto, se llaman **discretos**.

Son datos discretos: Número de hijos, nota de un examen, goles a favor o en contra, etc.

Son continuos: El peso, la estatura, la densidad de un líquido, la fuerza de un muelle, etc.

En muchos estudios los datos aparecen aislados, sin frecuencias. Por ejemplo, las notas de un examen, como se puede ver en la siguiente lista:

Notas de una prueba de Inglés:

0 , 0 , 1 , 2 , 3 , 3 , 4 , 4 , 4 , 5 , 5 , 5 , 6 , 6 , 7 , 8 , 8 , 9 , 4 ,
6 , 2 , 2 , 3 , 4

En la Hoja de Cálculo puedes escribir los datos en columna, para que ocupen menos espacio.

Inicia **OpenOffice** y escribe los datos. Debes guardarlo con otro nombre por si lo estropeas. En ese modelo están contenidos todos los datos que usarás en esta práctica.

Si deseas obtener las frecuencias deberás ir las contando tú y escribir aparte la tabla de frecuencias. Como curiosidad puedes ver cómo la Hoja de Cálculo calcularía la frecuencia de un dato, por ejemplo el 5. Junto a la columna de datos aislados, y a la derecha del 5 escribe la fórmula

=CONTAR.SI(RANGO;5)

donde la palabra **RANGO** la debes sustituir por las celdas ocupadas por los datos, por ejemplo B3:B19. Comprueba que resulta la frecuencia.

Lo normal es dar los datos acompañados por las frecuencias. Busca en la Hoja la tabla correspondiente a este ejemplo.

x	f
0	2
1	1
2	3
3	3
4	5
5	3
6	3
7	1
8	2
9	1

Frecuencias acumuladas

En el caso de las variables cuantitativas podemos calcular las frecuencias relativas y los porcentajes, tal como lo

hicimos con los datos cualitativos. Si los datos están ordenados, también se acostumbra a *acumular frecuencias*. Esto significa que a cada frecuencia se le pueden sumar todas las anteriores a ella para ver cómo se acumulan a lo largo de la tabla. Evidentemente, la última frecuencia acumulada coincidirá con el número total de datos. En el caso de la relativa, la última será 1 y en los porcentajes 100.

La tabla total quedará así:

x	f	h	p	f acum.	h acum.	p acum.
0	2	0,0833	8,33	2	0,0833	8,33
1	1	0,0417	4,17	3	0,1250	12,50
2	3	0,1250	12,50	6	0,2500	25,00
3	3	0,1250	12,50	9	0,3750	37,50
4	5	0,2083	20,83	14	0,5833	58,33
5	3	0,1250	12,50	17	0,7083	70,83
6	3	0,1250	12,50	20	0,8333	83,33
7	1	0,0417	4,17	21	0,8750	87,50
8	2	0,0833	8,33	23	0,9583	95,83
9	1	0,0417	4,17	24	1,0000	100,00

Ahora debes reproducir con **OpenOffice** la tabla de la página anterior. Te conviene dejar una fila en blanco sobre los datos. ¿Qué fórmula debemos usar para acumular frecuencias?

En todo este trabajo te ayudarán tus profesores.

Cuando tengas construida la tabla procura guardarla en disco cuanto antes.

Gráficos

Los gráficos más adecuados para los datos discretos son: el de **barras**, que ya hemos visto en los cualitativos, y el **gráfico lineal**, que consiste en unir los valores de las frecuencias por una línea continua. Este último es apropiado para datos que representan el paso del tiempo: temperatura a lo largo del día, cotización de unas acciones, índice de precios, etc.

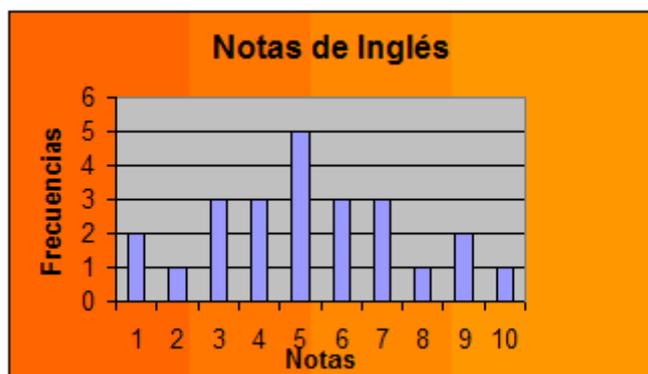
En el caso de las notas de Inglés es preferible el de barras. pues entre los amigos no hay relación de tipo temporal.

x	f
0	2
1	1
2	3
3	3
4	5
5	3
6	3
7	1
8	2
9	1

Adjunta un gráfico de ese tipo en tu Hoja de Cálculo.

Selecciona toda la tabla y la usas para confeccionar el

gráfico de barras. Te quedará algo parecido a esto:



Media y desviación típica

La media

La idea de **media aritmética** es casi idéntica a la de promedio: intentar repartir por igual la suma de los datos entre todos ellos. No obstante, tiene una definición matemática precisa:

Media en datos aislados

Si los datos están aislados, no agrupados en frecuencias, la definición de media es la siguiente:

$$m = \frac{\sum x}{N}$$

donde $\sum x$ significa la suma de todos los datos y n el número de ellos.

Por ejemplo, si las notas de diez amigos en un examen son

2.3 4.5 8 10 5.6 7 3.2 4.8 6 y 6

la media será la suma de 2,3+4,5+8+... dividida entre 10

En la Hoja de Cálculo **LibreOffice** es muy fácil sumar una columna. Copia estos datos en una misma columna de la Hoja de Cálculo. Señala la celda inferior a la tabla y pulsa la tecla **S** de la barra de herramientas o escribe **=SUMA(celda1:celda2)** Obtendrás el resultado de 57,4. Divides después entre 10 y ya tendrás calculada la media.

En general, si escribes

=SUMA(celda1:celda2)/CONTAR(celda1:celda2) siempre obtendrás la media, pues estás pidiendo el resultado de S_x / n

No obstante, el mejor método para Calcular la media es escribir **=PROMEDIO(celda1:celda2)**

Intenta ambos métodos y te saldrá 5,74 de media.

EJERCICIO: Halla (en LibreOffice) la media de la serie 1 1 2 2 2 3 3 4 4 4 4 5 5 6 6 8 9

mediante los dos tipos de fórmula explicados:

Solución: _____

Media en datos agrupados en frecuencias

Edad	Frec.
14	1
15	12
16	23
17	8
18	1

Cuando existen frecuencias, como en la tabla adjunta, la definición de media es:

$$m = \frac{\sum xf}{\sum f}$$

"Suma de los productos de cada dato por su frecuencia dividido entre la suma de las frecuencias"

Busca esa tabla en la Hoja de Cálculo. Junto a la columna de frecuencias debes construir otra columna en la que figure cada dato multiplicado por su frecuencia. Después, con la tecla **S** obtienes las sumas $\sum xf$ y $\sum f$. Comprueba que resulta una tabla como la siguiente:

Edad Frec. Edad*frecuencia

14 1 14

15 12 180

16 23 368

17 8 136

18 1 18

45 716

Divide ahora en otra celda las sumas 716 y 45, con lo que obtendrás 15,9111 como media de la tabla.

EJERCICIO: Halla la media de los datos de la tabla

x	2	4	6	8	10	12	14
f	1	3	6	14	11	4	3

Solución: _____

La desviación típica

Supongamos que los diez amigos de antes consiguen que la profesora repita la prueba y se obtengan estas notas:

2.8 4.5 9 10 7 7.5 3.2 5 9 y 7

en lugar de las anteriores

2.3 4.5 8 10 5.6 7 3.2 4.8 6 y 6

¿Ha subido el nivel?

Para responder escribe las notas nuevas en la Hoja de Cálculo y compara sus media con la de las anteriores. Usa la función PROMEDIO. Si no sabes cómo, pide ayuda. Verás que la media es ahora superior (¿Cuánto ? _____)

, luego ha subido el nivel entre un examen y otro, pues antes era de 5.74.

Los de nivel bajo se quejan. Dicen que se ha ayudado más a los que más saben, que hay ahora más distancia entre las buenas notas y las malas. ¿Tienen razón?

Sí la tienen, porque la desviación típica es ahora de 2.39 en lugar de 2.14. (Más adelante aprenderás a calcularlas)

Efectivamente, **la desviación típica** mide el grado de dispersión que tienen unos datos.

Si es pequeña, los datos son homogéneos, es decir, muy juntos o agrupados alrededor de la media. Si es mayor, los datos están más separados, más lejanos de la media.

Como en este ejemplo ha subido la desviación típica, quiere decir que los alumnos tienen las notas más separadas, que hay más diferencia entre notas altas y bajas.

Desviación típica en datos aislados

La fórmula de la desviación típica para datos aislados (ya simplificada) es la siguiente:

$$\sigma = \sqrt{\frac{\sum x^2}{n} - \bar{x}^2}$$

Es decir “la suma de los cuadrados de **x** dividida entre el número total y restada con el cuadrado de la media. Y a todo eso le extraemos la raíz cuadrada”

Puedes calcular la desviación típica con la fórmula **=DESVESTP(Celda1:celda2)** Inténtalo con la ayuda de tus profesores.

Comprueba que esta fórmula funciona. Construye una tabla con los cuadrados de las notas de Inglés, la sumas y aplicas la fórmula para comprobar que los resultados son 2,39 y 2,14 respectivamente. Hazlo todo con la Hoja de Cálculo y guarda los resultados en disco.

Desviación típica en datos agrupados

La fórmula de la desviación típica para datos agrupados (ya simplificada) es la siguiente:

$$\sigma = \sqrt{\frac{\sum(x^2 \cdot f)}{\sum f} - \bar{x}^2}$$

Es decir: “Multiplicar cada x al cuadrado por su frecuencia, sumarlo todo, dividir entre la suma de frecuencias y restar la media al cuadrado. Al resultado se le extrae la raíz cuadrada”

Prueba la definición con estos datos:

x f

2 8

3 4

4 2

5 1

7 2

Escribe el resultado: Media: _____ Desviación típica:

Coefficiente de variación

A veces, para comparar dos tablas y averiguar cuál de las dos es más homogénea, se usa el coeficiente de variación, que es el cociente de la desviación típica y la media

$$CV = \frac{\sigma}{\bar{x}}$$

EJERCICIO 1: ¿Cuál de estas dos tablas es más homogénea?

x	4	5	6	7	8	9
f	2	3	3	8	8	2

x	4	5	6	7	8	9
f	4	5	12	5	3	1

Respuesta: Es más homogénea la _____

Escribe los dos coeficientes de variación _____

EJERCICIO 2: Encuentra por tanteo el número que falta en la tabla si sabemos que la desviación típica es de 7

x	4	5.5	7	10	12	
f	1	3	3	2	1	1

Repaso de datos discretos con Hoja de Cálculo

Escritura de los datos

Estudia con Hoja de Cálculo los siguientes datos discretos, tomados de los resultados de una prueba de ortografía:

<i>Faltas de ortografía</i>	0	1	2	3	4	5	6	7	8
Frecuencia	5	7	8	5	4	2	0	1	1

Para escribir los datos es preferible hacerlo en columna. Reserva una columna para las faltas y otra consecutiva para las frecuencias y escribe una cabecera para cada una.

Faltas	Frecuencia
0	5
1	7
.....

Debajo de la columna de frecuencias usa la función +SUMA(...) para sumar toda la columna en alguna celda debajo de ella.

Crea otra columna paralela con las frecuencias relativas (Frecuencia partido por la suma) y otra de porcentajes (multiplicando por 100)

Con esto llenas la primera página. Cambia algún dato para ver cómo influye en el resto.

Media y desviación típica

En la página inferior calcularemos la media y la desviación típica. Para ello necesitamos columnas de cálculo. En

primer lugar, debes copiar la tabla en la nueva página. **No lo hagas con Editar Copiar**. Debes mejor copiar el primer dato con una fórmula del tipo =B2 y después usa **Llenar hacia abajo**. Haz lo mismo con las frecuencias. Observa si las dos tablas son idénticas. Cambia algo en la primera página para ver si también cambia en la segunda.

Una vez reproducida la tabla en la segunda página, debes añadir otras dos columnas paralelas: La XF, que debe contener los productos de X por F y la X²F para los productos de X² por F. Suma todas las columnas con la función +SUMA(...) en la parte baja de cada una. Con esas sumas calcula la media y la desviación típica. Te deberá dar Media= 2,39 y Desviación Típica=1,92

Puedes añadir el coeficiente de variación, que es el cociente de dividir la desviación entre la media. Es útil para comparar dos tablas, como verás más adelante.

Gráfico

Selecciona la tabla de frecuencias y pide crear gráfico nuevo. Con **Formato** añade Títulos o etiquetas. Guarda todo en disco.

Ejercicios

1) Usa el coeficiente de variación para comparar estas tablas y decidir cuál de ellas es más dispersa:

x	2	3	4	5
f	3	6	10	2

x	32	35	38	41
f	10	15	20	8

Como escribirás sobre la tabla anterior, no olvides **rellenar hacia abajo con espacios en blanco** las celdas que sobren.

2) ¿Qué frecuencia falta sabiendo que la media aproximada de esta tabla es 4?

x	2	3	4	5	6
f	2	10	20		5

3) Se han estudiado los sueldos de 50 hombres y 50 mujeres entre los directivos de una gran empresa, con el resultado de estas dos tablas, que están expresadas en **miles de euros (redondeados)**:

Hombres

x	4	5	6	7	8
f	8	12	18	8	4

Mujeres

x	4	5	6	7	8
f	8	16	16	6	4

Estudia las dos tablas con una Hoja de Cálculo y crea después un documento con el Procesador de Textos con tablas, gráficos y comentarios. Compara el nivel de sueldo y las medias y desviaciones típicas de ambos colectivos.