Estadística

Población: conjunto de todos los elementos a los que se somete a un estudio estadístico.
Individuo: cada uno de los elementos que componen la población.
Muestra: subconjunto representativo de la población.
Variable estadística: cada una de las características o cualidades estudio de la población. Puede ser:

- *Cualitativa nominal: casado, mujer*
- *Cualitativa ordinal: primero, notable*
- *Cuantitativa discreta: Valores concretos: 2, 3*
- *Cuantitativa continua: Valores intermedios: entre 4 y 5*

Frecuencia absoluta: Número de veces que se repite un valor
Frecuencia relativa: Cociente entre la frecuencia absoluta y el número total de datos.

xi	fi
0	1
1	1
2	3
3	2
4	6
5	11
6	4
7	2
8	3
9	7
Σ	**40**

Parámetros estadísticos:

De centralización: Media aritmética, mediana y moda
De posición: Cuartiles y percentiles
De dispersión: Desviac. Media, varianza, desv, típica, Coef. de Varciación (CV)

Medidas de posición central:

- *Media aritmética (Mitjana): Suma de los valores dividido por el total de valores.*
  Ejemplo: Los pesos de seis amigos son: 84, 91, 72, 68, 87 y 78 kg. Hallar el peso medio:
  X= 84+91+72+68+87+78 / 6 = 80 Kg

- *Mediana (Mediana): Valor central de una lista ordenada*
  Ejemplo: 1,5, 9 Mediana(Me)=5
  Si la serie tiene un número par de puntuaciones la mediana es la media entre las dos puntuaciones centrales.
  Ejemplo: 7, 8, 9, 10, 11, 12 Me= 9.5

- *Moda: El valor más repetido.*
  Ejemplo: Hallar la moda de la distribución:2, 3, 3, 4, 4, 4, 5, 5 Mo= 4
  Si en un grupo hay varias modas se llama es bimodal o multimodal.
  1, 1, 1, 4, 4, 5, 5, 5, 7, 8, 9, 9, 9 Mo= 1, 5, 9
  Si hay 2 modas adyacentes, la moda es la media de esas dos modas.
  0, 1, 3, 3, 5, 5, 7, 8 Mo = 4

Ejercicio 1:

Dadas las notas: 2, 4, 4, 5, 5, 5, 6, 7, 9 . Halla la media, la mediana y la moda.
Di la frecuencia absoluta de la nota 4

3. Distribuciones de frecuencia

Frecuencia absoluta, ni: Es el número de veces que se repite el valor de xi.
Por ejemplo si estamos estudiando las notas de Matemáticas de una clase y hay 8 alumnos que tienen un 7, 8 es la frecuencia absoluta de la modalidad 7.
N: suma de todas las frecuencias absolutas, o sea es el número total de individuos de la muestra: $\boldsymbol{N = \sum ni}$

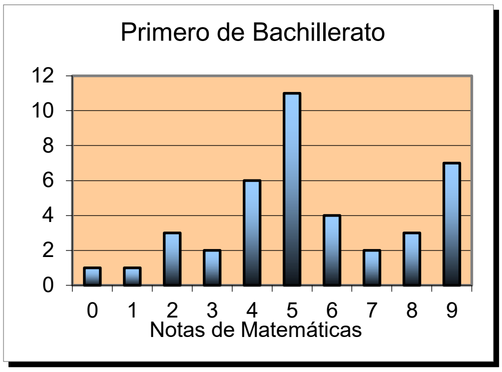
Frecuencia relativa, fi: Cociente entre la frecuencia absoluta y el número total de individuos: $fi = \frac{ni}{N}$
Frecuencia porcentual, pi Es la frecuencia relativa en tantos por ciento. $pi = 100 * fi$
Frecuencias acumuladas, Ni, Fi, Pi Suma de las frecuencias de los valores anteriores al valor considerado.
Puede ser absoluta, Ni, relativa, Fi o porcentual Pi. Se calculan sumando las frecuencias anteriores a cada una.
$N3 = n1 + n2 + n3$

4. Tablas de frecuencia y gráfico de barras:

Elabora la tabla de frecuencias de las notas de un grupo de 40 alumnos:

Valores	0	1	2	3	4	5	6	7	8	9
Frec. Absoluta	1	1	3	2	6	11	4	2	3	7

5. Parámetros estadísticos



Parámetros de centralización

1. **Media aritmética** Es el resultado que obtenemos al sumar todos los datos y dividir entre el número total de ellos. Sólo se puede hallar si los datos son cuantitativos: $\overline{x} = \dfrac{\sum xi\ fi}{N}$
2. **Mediana:** La mediana es el valor que ocupa el valor central al ordenar los datos. Sólo se puede hallar si los datos son cualitativos ordenables o cuantitativos.
3. **Moda:** Es la modalidad que tiene mayor frecuencia.

Parámetros de dispersión Nos indican si los valores están concentrados.

1. **Recorrido:** Es la diferencia entre el valor mayor y el valor menor de la variable.
2. **Varianza:** Es la media aritmética de los cuadrados de las desviaciones: $V = \dfrac{\sum(xi-\overline{x})^2 fi}{N}$
   Para calcularla añadimos otra columna: $(xi - \overline{x})^2 \cdot fi$
3. **Desviación típica**: Es la raíz cuadrada positiva de la Varianza: $\sigma = \sqrt{V}$

xi	fi	hi	Hi	% hi	% Hi	xi · fi	f ·| xi - x̄ |	f ·(xi - x̄)²
0	1	0,025	0,025	2,5	2,5	-	5	29
1	1	0,025	0,050	2,5	5,0	1	4	20
2	3	0,075	0,125	7,5	12,5	6	10	35
3	2	0,050	0,175	5,0	17,5	6	5	12
4	6	0,150	0,325	15,0	32,5	24	9	12
5	11	0,275	0,600	27,5	60,0	55	5	2
6	4	0,100	0,700	10,0	70,0	24	2	1
7	2	0,050	0,750	5,0	75,0	14	3	5
8	3	0,075	0,825	7,5	82,5	24	8	20
9	7	0,175	1,000	17,5	100,0	63	25	89
Σ	**40**	1		100		**217**	76,4	226

parámetros de centalización o concentración	parámetros de desviación o dispersión
Media artimética x̄: $\bar{x} = \dfrac{\sum x_i \cdot f_i}{n} =$ 5,425	Desviación media: $dm = \dfrac{\sum f_i \cdot \lvert x_i - \bar{x} \rvert}{n} =$ 2
Moda (Mo): fi max. = 5	Varianza: $Var = \dfrac{\sum f_i \cdot (x_i - \bar{x})^2}{n} =$ 5,6
Mediana (me): 50% de Hi = 5	Desviación típica: $s = \sqrt{\dfrac{\sum f_i \cdot (x_i - \bar{x})^2}{n}} =$ 2,4

6. Estadística con Hoja de Cálculo Excel

FUNCIONES ESTADÍSTICA DE EXCEL	
Contar datos	=Contar(Rango)
Media aritmética	=PROMEDIO(Rango)
Mediana	=MEDIANA(Rango)
Moda	=MODA(Rango)
Recorrido	=MAX(Rango)-MIN(Rango)
Varianza	=VARP(Rango)
Desviación típica	=DESVESTP(Rango)

DISTRIBUCIONES BIDIMENSIONALES - REGRESIÓN Y CORRELACIÓN

Ahora estudiaremos **dos** características de una población y su relación entre ellas.

Una variable estadística así definida se llama **bidimensional**. Estudiaremos la posible relación entre las dos características de la variable. Por ejemplo, si anotamos el peso y la estatura de una muestra de 50 personas, se observa una relación entre ambas características. La relación no será estrictamente funcional (a tal peso, tal talla), pero su estudio tiene gran interés, debido a lo frecuente de tales relaciones.

1. Dependencia funcional y dependencia estadística

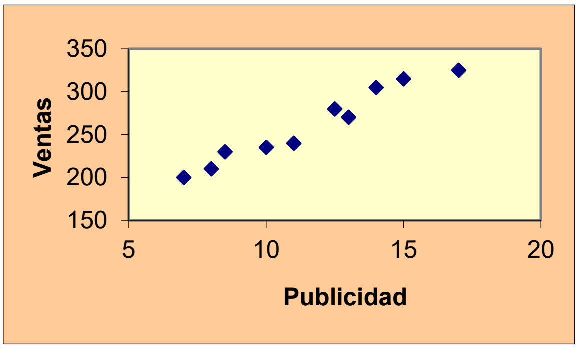
Al considerar los dos caracteres de una variable bidimensional, puede ocurrir:

- Que haya una **dependencia funcional** entre ellos, de tal modo que a cada valor de una le corresponda un único valor del otro. Por ejemplo, la altura desde la que cae un cuerpo y el tiempo que tarda en llegar al suelo.
- Que haya una **dependencia estadística o correlativa**, de tal manera que los valores sigan unas pautas similares. Por ejemplo, la estatura y el peso de un grupo de personas; la edad del marido y de la esposa; la importancia de las cosechas y el nivel de precipitaciones; etc.
- Que se dé una **independencia** entre los caracteres. Por ejemplo, la estatura de los alumnos y su calificación en Matemáticas.

2. Tablas y gráficos de una variable bidimensional

De modo similar al caso unidimensional, podemos recoger la información relativa a una variable bidimensional en una tabla. Por ejemplo, los gastos en publicidad y las ventas de una compañía (ambas cosas en millones de pesetas) fueron, en diez años consecutivos, los siguientes:

Publicidad	7	8	8,5	10	11	12,5	13	14	15	17
Ventas	200	210	230	235	240	280	270	305	315	325



Una representación gráfica muy ilustrada de estos datos viene dada por los conocidos *diagramas de dispersión*. Se representan sobre el plano cartesiano los puntos ***(x,y)***, siendo ***x*** el primer valor de la variable bidimensional e ***y*** el segundo. El conjunto de puntos resultantes (**nube de puntos**) nos da una primera idea de la relación existente entre los datos. En este caso siguen aproximadamente, una disposición rectilínea, que nos hace pensar en una dependencia entre los gastos de publicidad y las ventas de la compañía.

3. Distribuciones marginales

Una tabla de distribución de variable bidimensional tiene dos tipos de variables unidimensionales: la **distribición marginal de la variable X** y la **distribución marginal de la variable Y**.

Sus respectivos parámetros van a ser importantes en el estudio de regresión: $\bar{x}$ y s_x, media y desviación típica de la primera distribución marginal; $\bar{y}$ y s_y, parámetros de la segunda distribución marginal.

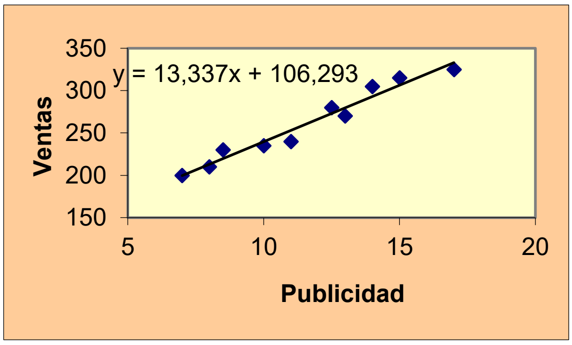
El punto cuyas coordenadas son las medias aritméticas marginales, G($\bar{x}$, $\bar{y}$) se llama **centro de gravedad** de la distribución.

4. Recta de regresión lineal

Cuando la nube de puntos deldiagrama de dispersión nos indica que hay alguna dependencia entre las dos variables X y Y, condensándose los puntos alrededor de una cierta línea **(línea de regresión)**, se plantean dos cuestiones: a) definir la línea; b) medir el nivel de aproximación de dicha línea.

Si la línea en cuestión es una línea recta, el problema es un caso típico de **regresión lineal**, que es el tipo de regresión más utilizado.

Una vez representada la nube de puntos, el averiguar cuál es la recta que más se aproxima a los puntos requiere unos cálculos muy laboriosos, pues consisten en hacer que las distancias de los puntos empíricos a los puntos de la recta sean mínimas.



Pero haciendo uso de la hoja de cálculo de Excel, tanto la representación gráfica como la obtención de la ecuación de dicha recta es inmediata.

En el ejemplo anterior la recta de regresión y su representación gráfica son las mostradas en la figura.

La ecuación de la recta de regresión nos puede servir para hacer *predicciones de resultados*.

Por ejemplo en el caso de la publicidad/ventas podemos averiguar qué venta originará un gasto de publicidad determinado o al revés, cuánto se debe gastar la compañia en publicidad para conseguir un nivel de ventas determinado.

Así si la compañía se gasta 9 millones en publicidad y queremos averiguar qué venta se originará, sustituimos **x=9** en la ecuación de la recta de regresión

y=13,3x+106,3

y=13,3*9+106,3=**226 millones**

Y viceversa, si queremos que se produzca una venta de 250 millones y queremos averiguar el gasto en publicidad necesario, sustituimos **y=250** en la ecuación de la recta de regresión, y posteriormente despejamos la x que nos da el gasto en publicidad.

250=13,3*x+106,3 $\longrightarrow$ **x=10,8 millones**

5. Covarianza: *Es la media aritmética de los productos de las desviaciones de los valores de cada variable por su respectiva frecuencia.* $s_{xy} = \frac{\sum_{i=1}^{n} ni(xi-\overline{x})(yi-\overline{y})}{N}$

Para calcularla con Excel se usa la siguiente fórmula: =COVAR(Rango)

La covarianza según sea positiva o negativa nos indica:

Covarianza positiva. Las variables se relacionan de forma directa. Es decir que al aumentar los valores de la variable X, aumentan los valores de la variable Y. Como el ejmplo anterior, al aumentar la publicidad aumentan las ventas.

Covarianza negativa.- Las variables se relacionan de forma inversa. Es decir, que al aumentar los valores de la variable X, disminuyen los valores de la variable Y. Por ejemplo el número de horas que se dedica a ver la TV y el número de aprobados. Mientras más horas de TV, menos aprobados.

6. Coeficiente de correlación r

Nos indica el grado de aproximación de los puntos de la nube a la recta de regresión: $r = \frac{s_{xy}}{s_x \cdot s_y}$

Si se calcula con Excel la fórmula es: **=COEF.DE.CORREL(Rango)**

Propiedades: Si r=-1 o r=1, la relación es perfecta. Existe correlación **funcional**. Si r está próximo a 1 o a -1 la correlación **es fuerte**. Si r está próximo a cero la correlación **es débil**. Si r>0 la correlación es positiva o **directa**. Si r<0 la correlación es negativa o **inversa**.

Ejercicios:

1. Un profesor ha registrado las notas de un examen corto (puntuado del 1 al 5) de un grupo de 15 estudiantes. Los resultados obtenidos son los siguientes:
   3, 2, 4, 5, 2, 1, 3, 3, 4, 3, 2, 5, 3, 4, 1.

Tu tarea:

1. Organiza estos datos en una **tabla de frecuencias**.
2. Calcula la **media** (el promedio de las notas).
3. Encuentra la **mediana** (el valor central).
4. Identifica la **moda** (la nota que más se repite).