

REGRESIÓN CON MS EXCEL

1. UN POCO DE HISTORIA

La teoría de la correlación y la regresión son muy recientes y su descubrimiento se debe al médico inglés *Sir Francis Galton*.

Galton nació el año 1822 en Birmingham en el seno de una familia acomodada. Estudió en Hospital General de Birmingham, en el King's College de Londres y en el Trinity de Cambridge.

Sus trabajos se desarrollaron entorno al estudio de la herencia y la expresión matemática de los fenómenos vinculados a ella.

El contexto histórico en el que vivió favoreció su interés por la herencia genética: nació el mismo año que *George Mendel* con el que mantenía una gran afinidad y era primo de *Charles Darwin*.



En 1869 publicó el libro "Hereditary Genius", y través del estudio de problemas de la herencia, llegó al concepto de correlación, siendo el primero en asignar a un conjunto de variables un número que permitía obtener una medida del grado de relación existente entre ellas.

Llegó a inferir que las personas excepcionalmente altas solían tener hijos de estatura menor que sus progenitores, mientras que las personas muy bajas solían tener hijos más altos que sus padres. Esta observación llevó a *Galton* a enunciar su "principio de la mediocridad", aplicable a las tallas de una generación respecto de las siguientes. Éste fue el origen del actual análisis de la regresión.

La observación de *Galton* es sin duda cierta, pero el supuesto de la regresión de la mediocridad es totalmente falso y se considera actualmente como una de las falacias de la regresión.

La justificación que se da hoy a este hecho es que los valores extremos de una distribución se deben en gran parte al azar, de ahí que los factores genéticos que producen una talla excepcional por exceso o por defecto no pasan a los hijos.

Su obra "Meteorográfica" fue el primer intento de previsión del tiempo y por otra parte puede ser considerado como el padre de la eugenesia.

Los trabajos de *Galton* fueron continuados y mejorados, entre otros, por *Karl Pearson*.



Pearson nació en Londres en 1857 y comenzó estudiando derecho. Posteriormente ejerció la abogacía al tiempo que simultaneaba sus actividades políticas y literarias. A los 27 años comenzó a impartir clases de matemáticas aplicadas en la universidad de Londres.

En 1901 fundó la revista "Biométrica", en la que publicó una biografía monumental de *Galton*.

A *Pearson* se deben aportaciones tan importantes como la distribución ji-cuadrado o el test de Pearson para el estudio de la bondad del ajuste de una distribución empírica a otra teórica.

2. CONCEPTOS TEÓRICOS

Cuando en una población estudiamos dos características comunes, decimos que estamos analizando una variable estadística bidimensional. Se plantean dos problemas diferentes aunque relacionados entre sí:

Correlación: que estudia el grado de relación o dependencia que existe entre las dos variables que intervienen en una distribución bidimensional.

Esta puede ser:

- **Lineal** o **curvilínea** según la nube de puntos se condense entorno a una línea recta o a una curva.
- **Positiva** o **directa** cuando al aumentar una variable aumenta la otra y viceversa.
- **Negativa** o **inversa** cuando al crecer una variable la otra decrece y viceversa.
- **Nula** cuando no existe ninguna relación y la nube de puntos están distribuidos al azar. Se dice que están *in correladas*.
- **Funcional** si existe una función tal que todos los valores de la nube de puntos la satisfacen.

Nos centramos en el estudio de la correlación lineal. El grado de dependencia lo mide el coeficiente de correlación lineal de Pearson $r = \frac{s_{xy}}{s_x s_y}$ siendo s_{xy} la covarianza y s_x y s_y las desviaciones típicas de las variables. Se pueden dar los siguientes casos:

$r=1$	$0 < r < 1$	0	$-1 < r < 0$	$r=-1$
funcional directa	estadística directa	correlación nula	estadística inversa	funcional inversa

Rectas de regresión:

Buscamos la recta que mejor se ajusta a la nube de puntos y utilizarla para predecir el valor de una variable conocido el de la otra. Será aquella que haga que la suma de las desviaciones de los puntos de la nube respecto de las correspondientes de la recta sea la menor posible.

Para ello se utiliza el "método de los mínimos cuadrados" para obtener la pendiente y la ordenada en el origen de la recta buscada.

Si la recta es para predecir un valor de **y** conocido el de **x**, se utiliza la **recta de regresión de y sobre x**, en la que debe ser mínima la suma de las d_i^2 (siendo d_i las distancias entre las ordenadas de los puntos de la nube y los correspondientes de la recta).

Su ecuación es: $y - m_y = \frac{s_{xy}}{s_x} (x - m_x)$ siendo m_x y m_y las medias de las variables.

Para obtener el valor esperado de **x** conocido el de **y**, se utiliza la **recta de regresión de x sobre y**, en la que debe ser mínima la suma de las d_i^2 (siendo d_i las distancias entre las abscisas de los puntos de la nube y los correspondientes de la recta).

Su ecuación es: $x - m_x = \frac{s_{xy}}{s_y} (y - m_y)$ siendo m_x y m_y las medias de las variables.

3. ENVEJECIMIENTO

Un estudio afirma que las orejas crecen 0,22 milímetros al año

No por tener las orejas más grandes escuchamos mejor, como pueden atestiguar muchos ancianos que han percibido cómo sus pabellones auditivos se engrandecen con el paso de los años en tanto que su capacidad de oír disminuye.



Esta evolución del tamaño de las orejas en correlación con el progresivo envejecimiento ha dejado de ser una mera percepción subjetiva. Unos médicos británicos, decididos a tomar el toro por las astas – en este caso por las orejas –, han medido los pabellones de todos sus pacientes mayores de 30 años y han llegado a la conclusión de que aumentan de longitud a una media de 0,22 milímetros por año.

Un total de 206 pacientes con una edad media de 53 años (abarcando desde los 30 a los 93) participó de muy buen grado en el

estudio, cuyos resultados fueron publicados en el *British Medical Journal*. Con este objetivo, los cuatro autores de la investigación midieron de arriba abajo la longitud del pabellón externo, obteniendo una longitud media de 675 milímetros (entre 520 y 840 milímetros). A partir de los datos recogidos, elaboraron una ecuación lineal de regresión, de la cual resulta que el tamaño de las orejas de cualquier persona se obtiene con la siguiente fórmula:

$$55,9 + (0,72 \times \text{la edad del individuo})$$

Comprobando los resultados, los médicos, todos ellos miembros del *Royal College of General Practitioners* del Reino Unido, encontraron que, en efecto, cuanto más se envejece más aumenta el tamaño de las orejas. De ser válidos los números calculados, significaría que una persona de 30 años puede prever que, cuando cumpla los 80, sus orejas le habrán crecido algo más de un centímetro.

P.F. Madrid. El País