

Sesión 4: Datos Agrupados

Módulo 1: Estadística Descriptiva

Magdalena Cornejo

Relaciones entre variables

Hasta ahora hemos discutido cómo utilizar las medidas de tendencia central y variabilidad para resumir un conjunto de datos.

- También estamos interesados en medir la fuerza de la relación entre dos conjuntos de datos.

Relaciones entre variables

Hasta ahora hemos discutido cómo utilizar las medidas de tendencia central y variabilidad para resumir un conjunto de datos.

- También estamos interesados en medir la fuerza de la relación entre dos conjuntos de datos.
- Por ejemplo, ¿cómo se relaciona el gasto de publicidad con las ventas de una compañía? ¿Cómo se relaciona el precio de una bebida con la cantidad consumida?

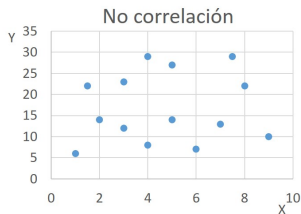
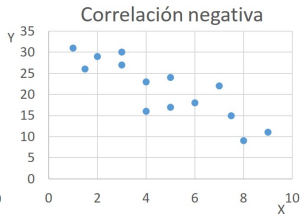
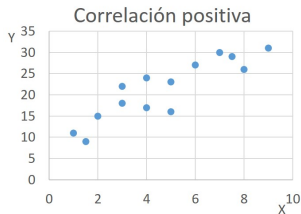
Relaciones entre variables

Hasta ahora hemos discutido cómo utilizar las medidas de tendencia central y variabilidad para resumir un conjunto de datos.

- También estamos interesados en medir la fuerza de la relación entre dos conjuntos de datos.
- Por ejemplo, ¿cómo se relaciona el gasto de publicidad con las ventas de una compañía? ¿Cómo se relaciona el precio de una bebida con la cantidad consumida?
- La relación (lineal) entre dos conjuntos de datos se puede medir a través de la **correlación**.

Diagramas de dispersión (Scatter Plot)

En cada eje (X e Y) se pone una variable y cada punto corresponde a una observación.



Coeficiente de correlación de Pearson

Intenta medir cuán asociadas (linealmente) están dos variables.

$$r_{XY} = \frac{S_{XY}}{S_X S_Y}$$

donde,

- S_{XY} es la covarianza entre X e Y
- S_X es el desvío estándar de X
- S_Y es el desvío estándar de Y

En Excel: =COEF.DE.CORREL()

Covarianza

- Para calcular el coeficiente de correlación (r_{XY}) necesitamos calcular la **covarianza** (S_{XY}).
- La covarianza indica el co-movimiento de dos variables respecto a su media.
- Formalmente:

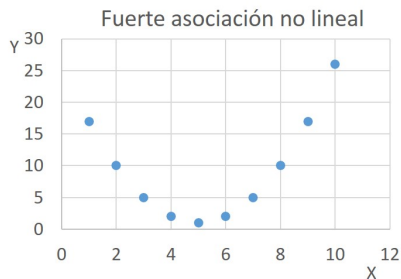
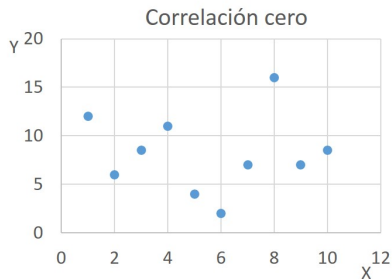
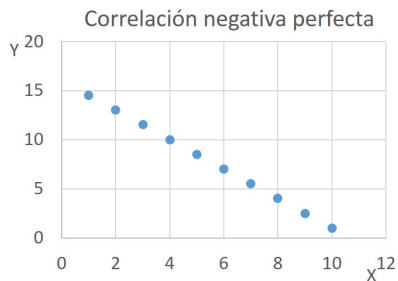
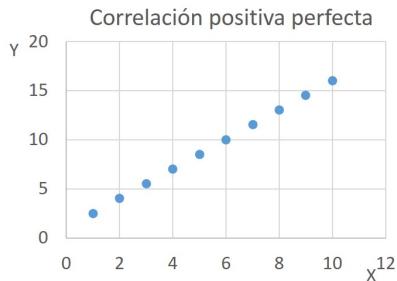
$$S_{XY} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n - 1}$$

- En Excel: =COVAR()
- $S_{XY} > 0$ indica dependencia lineal directa entre las dos variables
- $S_{XY} < 0$ indica dependencia lineal inversa entre las dos variables
- Al igual que la varianza, no es un estadístico libre de unidades de medida.

Propiedades del coeficiente de correlación

- Libre de unidades de medida.
- Invariante al cambio de unidades (p. ej., si paso se medir X en miles de USD a millones de USD, no cambia el coeficiente).
- $-1 \leq r \leq 1$, ya que está normalizado por los desvíos.
- Su signo está dado por el signo de la covarianza.
- Mide dependencia lineal, si los datos (X_i, Y_i) tienden a caer sobre una recta.
- Cuantifica la fuerza de la relación, pero no la forma de la recta (su pendiente y ordenada al origen).

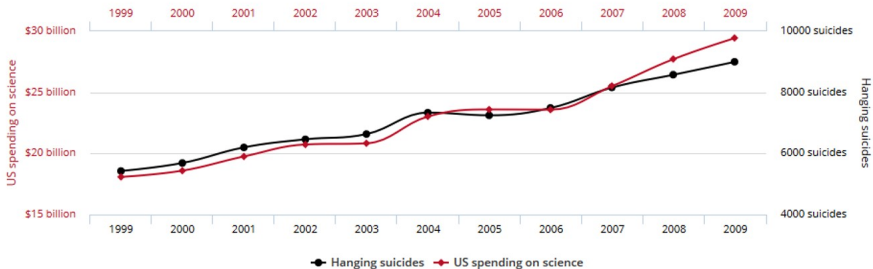
Coeficiente de correlación



¡Correlación **no** implica causalidad!

US spending on science, space, and technology correlates with Suicides by hanging, strangulation and suffocation

Correlation: 99.79% ($r=0.99789126$)

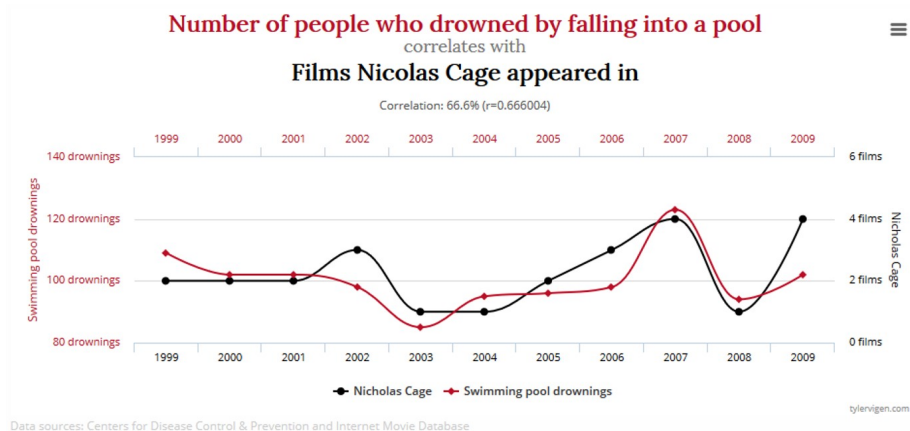


Data sources: U.S. Office of Management and Budget and Centers for Disease Control & Prevention

tylervigen.com

Fuente: tylervigen.com

¡Correlación **no** implica causalidad!



Fuente: tylervigen.com

Aplicación en Excel

- Bajar de **Yahoo Finance** (finance.yahoo.com) el precio de cierre ajustado de Intel, Cisco, General Motors, Microsoft e IBM entre el 4 de enero de 2016 y el 22 de julio de 2016.
- Compute el retorno diario de cada una de estas acciones a través de la siguiente fórmula:

$$R_t = \frac{(P_t - P_{t-1})}{P_{t-1}} \times 100 = \left(\frac{P_t}{P_{t-1}} - 1 \right) \times 100$$

- Utilice el complemento de Excel “Herramienta para análisis” para obtener la matriz de correlaciones.
- ¿Qué observa?