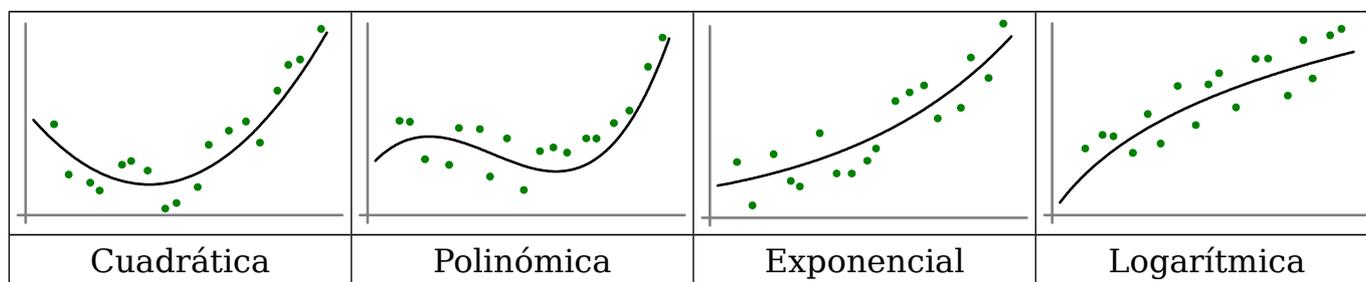


Significado de la correlación

Cuando estudiamos una distribución estadística bidimensional y nos encontramos con un coeficiente de correlación muy próximo a 1 o a -1 , sabemos que los valores de las dos variables están muy relacionados, pero aún no sabemos por qué. Puede ocurrir que no haya ningún motivo en especial, no tiene por qué haber una relación causa-efecto que explique la correlación; en otras palabras: alta correlación no implica causalidad. Pero también podría haber una causa subyacente que explique la correlación, luego será tarea de los investigadores que trabajan con esas variables ahondar más en el estudio. Es decir: alta correlación sugiere más estudio.

Tipos de correlación

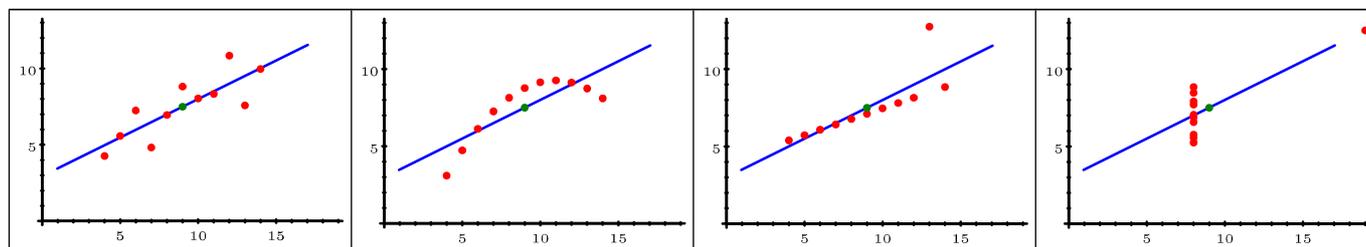
En enseñanza secundaria solo se estudia la correlación lineal, pero la correlación entre dos variables estadísticas puede ser de más tipos: cuadrática, polinómica, exponencial, logarítmica, etcétera, según sea la función que mejor se ajuste a los valores observados.



En estos casos, hay que pensar que es muy probable que realmente exista una relación funcional, aún desconocida, entre las variables y que hay algo que impide obtener valores exactos. En la naturaleza existen multitud de pequeños factores que, al acumularse, desvían los datos de los que tendría una función matemática.

Importancia de la visualización

Los cálculos del coeficiente de correlación y la recta de regresión no lo son todo para estudiar un conjunto de datos estadísticos bidimensional. Para ilustrarlo, en 1973 el estadístico británico Frank Anscombe (1910-2001) construyó cuatro conjuntos de datos bidimensionales que presentan los mismos valores para los parámetros estadísticos y sin embargo al visualizarlos vemos que son completamente diferentes. Presentamos los puntos en rojo, el centro de gravedad en verde y la recta de regresión en azul.



Importancia de la cantidad de datos disponibles

Cuando hay muchos datos disponibles para estudiar un problema, es más sencillo hacerlo. Esto explica los avances que se consiguen usando técnicas estadísticas en el llamado *big data* y también lo corrobora lo difícil que es descifrar un problema cuando es único, como ejemplifica el disco de Festo (a la derecha), de propósito y significado aún desconocidos.

