

Regresión lineal simple

Simple linear regression

Verónica Martínez Espinosa^a

Abstract:

Linear regression analysis is a statistical process technique that helps predict the behaviour of a dependent (response) variable from another, called an independent (predictive) variable, as long as they are continuous. Simple linear regression refers to a single independent variable used for prediction, and it uses a linear function to predict.

Linear regression models the relationship between two variables, so its usefulness spans different disciplines such as medicine, statistics, economics, sociology, and others. It allows us to model functions, analyse trends, optimize, and interpret data.

Keywords:

Analysis, linear regression, least squares method, variables

Resumen:

El análisis de regresión lineal es una técnica del proceso estadístico que ayuda a predecir el comportamiento de una variable dependiente (de respuesta), a partir de otra llamada independiente (predictiva) siempre y cuando sean continuas. La regresión lineal simple se refiere cuando solo tenemos una sola variable independiente para realizar la predicción, además usa una función lineal para predecir.

La regresión lineal modela la relación entre 2 variables por lo que su utilidad es abarca diferentes disciplinas como la medicina, estadística, economía, sociología entre otros. Así podemos modelar funciones, analizar tendencias, optimizar e interpretar datos.

Palabras Clave:

Análisis, regresión lineal, mínimos cuadrados, variables

Regresión lineal

La relación entre variables está presente en la mayoría de fenómenos, dichas variables son esenciales para definir cualquier objeto de estudio. La regresión lineal es una técnica estadística que además de analizar la relación entre 2 variables, permite modelar para encontrar la mejor función lineal que se ajuste a los datos observados en cuestión y predecir el comportamiento de las variables, esto a su vez resulta importante para la toma de decisiones.

Existen 2 tipos de regresión lineal: simple y múltiple; en la regresión simple solo se analiza la relación entre una variable dependiente y una independiente, mientras que en la regresión múltiple existe una variable dependiente y 2 o más variables independientes, existiendo múltiples factores que influyen. La regresión lineal sobresale por su simplicidad y una interpretación fácil, pero existen otros métodos para situaciones más complejas como la regresión logística o modelos no lineales.

- Las ventajas del uso de la regresión lineal se traducen en: El modelado de funciones;

^a Verónica Martínez Espinosa, Universidad Autónoma del Estado de Hidalgo | Escuela Preparatoria Número Cuatro | Pachuca de Soto, Hidalgo | México, <https://orcid.org/0000-0002-0411-203X>, Email: profe_4506@uaeh.edu.mx

fundamental para campos como la ciencia, ingeniería y economía.

- La predicción; vital para la planificación y toma de decisiones.
- Análisis de tendencias; para comprender el comportamiento de las variables.
- Control y optimización: ayuda a identificar las variables de mayor impacto.

Su aplicación puede ir desde las predicciones o pronósticos para valorar resultados futuros, monitorear y controlar procesos que impactan en el control de calidad, optimización de procesos, investigaciones y modelados de funciones. Los campos de aplicación son la medicina, economía, biología, ingeniería y estadística entre otros.

Método de mínimos cuadrados

Al hablar de regresión lineal se utiliza la técnica de análisis numérico, conocida como mínimos cuadrados, ya que ayuda a determinar la recta de ajuste y obtener la mejor aproximación lineal de un conjunto de datos, es muy recomendable cuando existen cierto tipo de errores en los datos. Este método ayuda a minimizar errores, el desarrollo de éste método se le atribuye a Carl Friedrich Gauss.

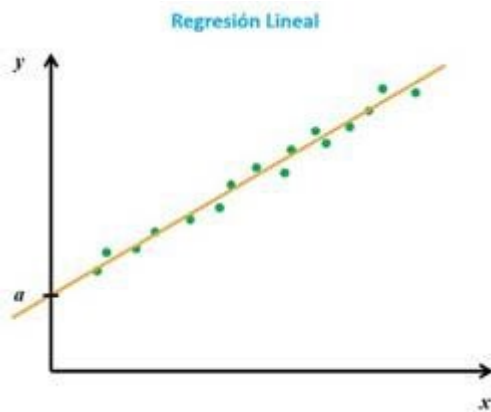


Figura 1. Ajuste lineal de datos con mínimos cuadrados

Este método de regresión por mínimos cuadrados permite ajustar el comportamiento de los datos, a través de una recta, como puede verse en la figura, así se minimiza la suma de los cuadrados de las distancias verticales de los puntos hacia la recta, para buscar un comportamiento normal.

Su expresión general se basa en la ecuación de una recta

$$y = mx + b$$

Donde **m** es la pendiente y **b** la ordenada al origen, que se convierte en el punto de corte y vienen expresadas de la siguiente manera:

$$m = \frac{n \cdot \sum(x \cdot y) - \sum x \cdot \sum y}{n \cdot \sum x^2 - |\sum x|^2}$$

$$b = \frac{\sum y \cdot \sum x^2 - \sum x \cdot \sum(x \cdot y)}{n \cdot \sum x^2 - |\sum x|^2}$$

Así, a partir de un conjunto de datos (variables), se obtiene el valor de la pendiente y la ordenada al origen, para construir la recta que ajusta la dispersión de los datos, como lo muestra la figura 1.

El método de mínimos cuadrados es un método de estimación, de gran ayuda para minimizar errores en una serie de datos, a través de una aproximación lineal. Resulta ideal para situaciones experimentales, proporciona resultados precisos, predice resultados que inciden en la toma de decisiones. Además, pueden implicarse en su estimación grandes conjuntos de datos, esta eficiencia permite su utilidad en campos como ingeniería. Su uso es fácil con paquetes de software como Excel, MATLAB, entre otros más especializados.

Referencias

- [1] Chapra, S, (2006). Métodos numéricos para ingenieros. McGrawHill. México.
- [2] Burden, R, Faires, D, (2011). Numerical Analysis. Books Ed. Boston 77-27.
- [3] Cruces, S, (s/f). El método de mínimos cuadrados. Consultado en: <https://personal.us.es/sergio/PDocente/lectura.pdf>, 1-5.