REGRESIÓN LINEAL Y CAMBIO DE VARIABLE

Fís. Carlos Adrián Jiménez Carballo
Escuela de Física
Instituto Tecnológico de Costa Rica

TEC | Tecnológico de Costa Rica

Regresión lineal

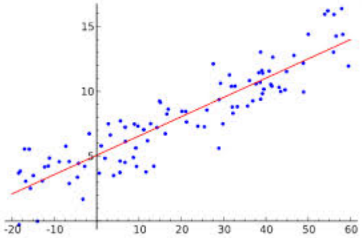
Cambio de variable

Contenido

Regresión Lineal

El análisis de regresión lineal es una técnica estadística la cual tiene como objetivo modelar en forma matemática el comportamiento de una variable de respuesta (o dependiente y) en función de una o más factores (o variables independientes $x_1, x_2, ..., x_n$).

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n + \varepsilon$$



Mínimos Cuadrados

- Es una técnica de análisis numérico de optimización matemática.
- Lo que se busca es encontrar la ecuación que mejore se ajuste entre un conjunto de pares ordenados (*variable dependiente y variable independiente*), y una familia de funcione, esto de acuerdo con el criterio de mínimo error cuadrático.
- En su forma más simple, intenta minimizar la suma de cuadrados de las diferencias en las ordenadas entre los puntos generados por la función elegida y los correspondientes valores en los datos.

Pendiente e intersección

El objetivo del método, para el caso de nosotros es encontrar, entre las variables dependiente e independiente, una relación que sea lineal

$$y = mx + b,$$

donde la pendiente m de dicho gráfico se determina:

$$m = \frac{n\sum_i y_i x_i - \sum_i y_i \sum_i x_i}{n\sum_i x_i^2 - \left(\sum_i x_i\right)^2}$$

y la intersección b

$$b = \frac{\sum_i y_i - m\sum_i x_i}{n}$$

nota: n representa el número de pares ordenados

Coeficiente de correlación lineal “r”

- Indica que tan correcto es el ajuste lineal que se está realizando para el conjunto de pares ordenados.
- Está definido entre los valores 1 y -1. Entre mas se acerque r a 1 ó -1 mejor es la relación lineal. Si r se acerca a 0 ó se aleja mas de 1 ó -1 eso quiere decir que el ajuste lineal no es una buena idea
- Se determina

$$r = \frac{n\sum_i y_i x_i - \sum_i y_i \sum_i x_i}{\sqrt{n\sum_i x_i^2 - \left(\sum_i x_i\right)^2}\sqrt{n\sum_i y_i^2 - \left(\sum_i y_i\right)^2}}$$

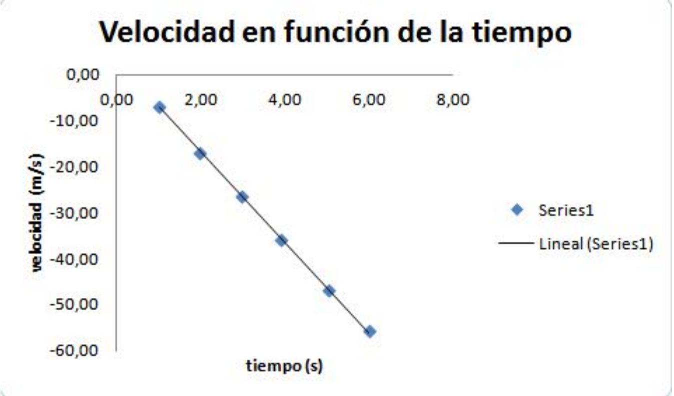
Ejemplo Mínimos Cuadrados

En un laboratorio se desea encontrar la relación entre la velocidad y el tiempo de una manzana en caída libre y se obtienen los siguientes valores

Tiempo (s)	Velocidad (m/s)
1,01	-6,90
1,99	-17,00
2,98	-26,35
3,90	-35,90
5,05	-47,00
6,00	-55,79

Tabla: Velocidad en función del tiempo

Ejemplo Mínimos Cuadrados



Contenido

Linealización de gráficas y cambio de variable

La linealización de gráficas es un método que tiene como objetivo trasformar líneas curvas en líneas rectas. Lo anterior se realiza pues los modelos lineales son más fáciles de analizar

Un cambio de variable es la transformación de una variable en otra empleada para reducir un problema muy complicado a uno más fácil de tratar. Para el caso de la gráficación lo que se busca con un cambio de variables es buscar un conjunto de variables que satisfagan una relación lineal

Cambio de variable:
Si se conocen la relación teórica entre las varibles gráficadas

- El objetivo del método es expresar la relación teórica entre dos o más variables graficadas en la forma
$$y = mx + b.$$
- Como ejemplo considere un experimento en el que se quiere analizar la relación entre el periodo y la longitud de un péndulo simple.

Ejemplo: Péndulo simple

- Si se grafica el periodo del péndulo en función de su longitud se obtiene una curva con la forma:

L (m)	T (s)
1	2,01
2	2,84
4	4,01
8	5,68
11	6,66

Tabla: Periodo péndulo simple en función de su longitud

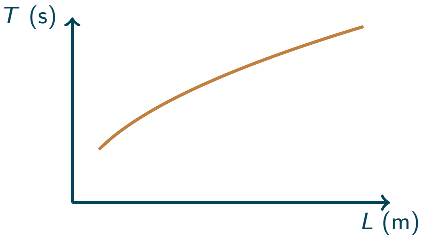


Figura: T vs L

Paso 1

Lo primero que se debe que hacer es escribir la ecuación teórica que relaciona las variables a graficar. Posteriormente se debe hacer la distinción entre variables medidas y constantes presentes en dicha ecuación.
Para este caso se tiene

$$T = 2\pi\sqrt{\frac{L}{g}},$$

donde es tienen

- Variables medidas: T y g,
- Constantes: g.

Pasos 2 y 3

Paso 2

Lo siguiente es separar las constantes de las variables consideradas y encerrarlas entre paréntesis y buscar escribir las ecuación de tal manera que tenga la forma de una ecuación de recta

$$T = \left(\frac{2\pi}{\sqrt{g}}\right)\sqrt{L}.$$

Paso 3

Finalmente comparare la ecuación resultante con la ecuación general de una línea recta y defina quien representa m, b, x (la nueva variable a graficar en el eje x) y y (la nueva variable a graficar en el eje y)

$$T = \left(\tfrac{2\pi}{\sqrt{g}}\right)\sqrt{L} + 0$$

$$\downarrow \quad \downarrow \quad \downarrow \quad \downarrow$$

$$y = m \quad x + b$$

Resultado

De lo anterior se obtiene

L (m)	T (s)
1.00	2,01
1.41	2,84
2.00	4,01
2.83	5,68
3.31	6,66

Tabla: Periodo en función de $\sqrt{L}$

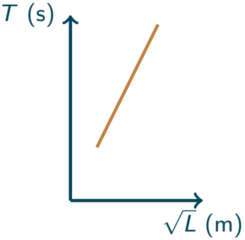


Figura: T vs $\sqrt{L}$

Ejemplo

Considere el periodo del péndulo físico el cual se define como

$$T = 2\pi\sqrt{\frac{I_{cm} + md^2}{mgd}}$$

Suponga que para este caso las variables medidas en el laboratorio son el periodo T y la distancia d que hay desde el centro de masa hasta el punto donde gira el objeto, ademásI_{cm} y g son constantes por determinar y m es una constante conocida. Realice un cambio de variable que linealice la relación entre las variables medidas.

Bibliografía

- Douglas, M. (2002). *Diseño y análisis de experimentos*. Limusa Wiley, Segunda Edición, México.

- Pulido, H. G., De la Vara Salazar, R., González, P. G., Martínez, C. T., & Pérez, M. D. C. T. (2012). *Análisis y diseño de experimentos*. New York, NY, USA:: McGraw-Hill.

- Laguna, C. (2014). Correlación y regresión lineal. *Aragón: Instituto Aragonés de Ciencias de la Salud*.

- Rodríguez, M. E. M. (2005). Errores frecuentes en la interpretación del coeficiente de determinación lineal. *Anuario jurídico y económico escurialense, (38), 315-331.*