

Análisis de Regresión.

Seguramente este tema lo cubrirán en otras clases como métodos estadísticos, incluso es fundamental e en econometría, por tanto no se pretende tocarlo a profundidad. El qué y el por qué de este análisis lo cubrirán en dichas clases; por los momentos aquí nos centraremos en el cómo realizarlo, sobre todo con la ayuda de una hoja de cálculo. Pero para poder ayudarles a estimar los coeficientes de regresión no está de más introducir la parte teórica del tema, para lo cuál presento lo que dice wikipedia al respecto:

En estadística, el análisis de la regresión es un proceso estadístico para estimar las relaciones entre variables. Incluye muchas técnicas para el modelado y análisis de diversas variables, cuando la atención se centra en la relación entre una variable dependiente y una o más variables independientes (o predictoras). Más específicamente, el análisis de regresión ayuda a entender cómo el valor de la variable dependiente varía al cambiar el valor de una de las variables independientes, manteniendo el valor de las otras variables independientes fijas.

Damodar Gujarati, comparte algo muy similar:

El análisis de regresión trata del estudio de la dependencia de una variable (variable dependiente) respecto de una o más variables (variables explicativas) con el objetivo de estimar o predecir la media o valor promedio poblacional de la primera en términos de los valores conocidos o fijos (en muestras repetidas) de las segundas.

En todo caso, el objetivo es la estimación de una función, relacionando las variables independientes con la dependiente, llamada la función de regresión. Sin embargo, ambas fuentes advierten de las limitaciones de este análisis, ya que la correlación no necesariamente implica causalidad.

En la presente práctica se busca obtener una función lineal de regresión en dos variables, una variable dependiente y una independiente o explicativa; cuya forma general es: $y = ax + b$ o de una forma un poco más técnica: $Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + u_i$

Por lo tanto, el trabajo consiste en obtener los coeficientes, llámense “a” y “b” ó $\hat{\beta}_0$ y $\hat{\beta}_1$

Estos coeficientes (que son el objetivo de la presente práctica) se obtienen de varias maneras, una de la cuales es mediante las siguientes formulas:

$$\hat{\beta}_1 = \frac{\sum x_i y_i}{\sum x_i^2} \quad \text{y} \quad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

Donde:

$$x_i = X_i - \bar{X} \quad \text{y} \quad y_i = Y_i - \bar{Y}$$

Por si no han cubierto o si no lo recuerdan, cuando una variable tiene una línea sobre ella, indica el promedio de dicha variable; el uso de mayúsculas y minúsculas para los nombres de variables, normalmente indica que la mayúscula es la variable observada y la minúscula la variable ajustada, el “sub i” indica que eso aplica a la i-esima observación.

Si todavía no esta muy claro como se relaciona este tema con las hojas de calculo, con el siguiente ejemplo se aclarará.

Para la siguiente muestra de datos, que corresponden al valor de un grupo de casas (en miles de dólares) y a su área (en miles de pies cuadrados), realizar un análisis de regresión lineal en dos variables. Identificando las variables dependiente e independiente (¿Cuál es X? ¿Cuál Y?). Es decir, encontrar los coeficientes de regresión que nos permitirán estimar el valor de una casa en base a su tamaño. (De manera muy similar a lo desarrollado en clase).

El primer paso, es identificar cuál es la variable dependiente y cuál la independiente. En ese sentido la pregunta que conviene hacerse es que es más lógico ¿Qué el precio de una casa dependa del tamaño de la misma o que el tamaño de la casa dependa de su precio? (aunque hay más variables independientes).

Una vez se ha contestado esto esto podemos empezar a crear la tabla, organizando los datos en columnas, etiquetando con la letra Y mayúscula, la columna de la variable dependiente y con la X mayúscula la variable independiente. A estas dos columnas se les debe totalizar y obtener su promedio. La tabla, también contendrá cuatro columnas más, una para los valores ajustados de “y”, otra para los de “x”, una para el producto de estas dos columnas y la última para el cuadrado de los valores ajustados de “x”.

Con esta tabla prácticamente ya está todo el trabajo realizado, solo falta obtener los coeficientes; tal como lo indican las formulas previamente compartidas, consisten en totalizar las columnas del producto de las “x” y “y” ajustadas, así como la del cuadrado de las “x” ajustadas y el $\hat{\beta}_1$ es el cociente del primer total, sobre el segundo. En tanto, que el $\hat{\beta}_0$ también conocido como intercepto, se obtiene de la resta del promedio de Y menos el producto del $\hat{\beta}_1$ por el promedio de X.

Los datos son los siguientes:

Observaciones	1	2	3	4	5	6	7	8	9	10	11	12	13
Valor (miles de dolares)	75	95	110	102	95	87	82	115	122	98	90	72	117
Área (miles de pies cuadrados)	1.1	1.5	1.6	1.6	1.4	1.3	1.1	1.7	1.9	1.5	1.3	1	1.8

Finalmente, como se dijo previamente con estos Betas se puede formar la ecuación de regresión, que en este caso nos permite predecir el valor de una casa si conocemos su tamaño. Digamos para una casa de 1,200 pies cuadrados ¿Cuál sería su valor?

Enlaces de interés:

https://es.wikipedia.org/wiki/An%C3%A1lisis_de_la_regresi%C3%B3n

<https://economipedia.com/definiciones/analisis-de-regresion.html>