**Análisis de regresión simple**

El modelo lineal en una única variable independiente es

es una variable aleatoria que satisface los siguientes supuestos:

: coeficiente de posición y representa el valor de la ordenada cuando x=0 (no siempre es interpretable)

: pendiente del modelo, es decir, cambio en la variable dependiente cuando se incrementa en una unidad

y se llaman parámetros del modelo

¿cómo estimamos y ?

Para estimar y empleamos el método de los Mínimos Cuadrados Ordinarios (MCO) que consiste en minimizar la suma de los errores al cuadrado, dado por

pero

Si el modelo subestima

Si el modelo sobre estima

Por lo tanto,

Del cálculo para encontrar los valores de y debemos resolver el sistema de ecuaciones

Ecuaciones normales

Además, sabemos que

Resolviendo para y se tiene

**Teorema de Gauss-Markov**

Bajos los supuestos del modelo y son estimadores insesgados y de mínima varianza

En efecto

Siendo

Así

Con esto se tiene

Retomando

**Inferencia sobre**

Una de las inferencias más importantes respecto de es:

Recordemos que es estimador de es . Para determinar la distribución muestral de haremos el supuesto distribucional sobre la variable aleatoria la cual nos dice que es Normal e independiente (NID)

Por teorema

Por otra parte como

Entonces  *con*

* Si es conocido entonces el estadístico de prueba es

Así, será rechazada si p-value< para algún nivel de significancia dado.

* Si es desconocido entonces el estadístico de prueba es

Así, será rechazada si p-value< para algún nivel de significancia dado.

**Análisis de varianza, para el análisis de regresión**

Gráfico, Gráfico de dispersión

Descripción generada automáticamente

Por lo tanto

es la suma de cuadrados totales = SCT

es la suma de cuadrados del error = SCE

es la suma de cuadrados de la regresión = SCR

Así,

Tabla ANOVA

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Fuente de variación** | **Grados de libertad** | **Suma de cuadrados** | **Cuadrados medios** | **F** | **p-value** |
| Regresión o modelo |  |  |  |  |  |
| Error o residuo |  |  |  |  |  |
| Total |  |  |  |  |  |

Esta tabla ANOVA sirve para contrastar la hipótesis

Cuando se vea regresión multivariada, tomará más relevancia esta tabla.

recibe el nombre de coeficiente de determinación y su interpretación es dada en término del porcentaje de variabilidad total que queda explicada por las variables independientes en el modelo. es considerado también como la capacidad predictiva que tiene el modelo.

Mientras más cercano a uno sea este valor, el modelo de considera que representa de mejor forma los datos y por lo tanto, puede ser usado con mayor confianza para predecir los valores de la variable respuesta.