

MAPAS CONCEPTUAL

WILLIAMS ERNESTO JIMENEZ AGUILAR

INGENIERIA EN SISTEMAS COMPUTACIONALES

1° GRADO

UDS COMITAN

GEOMETRIA ANALITICA



MODELOS DE REGRESION LINEAL MULTIPLE

REGRESION

Una técnica estadística que consiste en calcular dicha similitud en forma de función matemática.

Existen muchas técnicas de regresión en función del tipo de variables. Las más elementales son las lineales.

Los modelos de regresión múltiple suponen que más de una variable tiene influencia o está correlacionada con el valor de una tercera variable.

$$Y_j = \beta_0 + \beta_1 X_{1j} + \beta_2 X_{2j} + \dots + \beta_k X_{kj} + u_j$$

Y: variable endógena

X: variable exógena

u: residuos

β: efecto marginal entre cada valor de X sobre Y

TIPOS DE VARIABLE

Dichas variables tanto endógenas como exógenas pueden adoptar dos formas generales:

DEPENDIENTE

CONTINUAS

Aquellas que llenan el espacio, números reales con rango $-\infty$ hasta $+\infty$. Suelen ser variables

DISCRETAS

Aquellas que se mueven en un salto. Estas variables suelen ser factores cualitativos que indican una característica de un individuo.

Censurada

Faltan datos en la muestra.

Truncada

Doble Valla

Surge como evolución de un modelo truncado.

Libre

Se tiene datos de una muestra que abarca toda la posible medida de la variable que esta pueda tener la población.

HIPOTESIS Y ESTIMACION

Para que los resultados de la regresión sean "confiables" (insesgados y de varianza mínima) es necesario que:

la
Relación entre las variables sea lineal.

las
Perturbaciones deben resumirse bajo la denominación de "esfericidad".

el
Software tiene dos formas de estimación de una regresión lineal

Mínimos cuadrados ordinarios

Maxima verisimilitud

Consiste en resolver:

$$\hat{b} = (X'X)^{-1}X'Y$$

\hat{b} : vector de estimación de coeficientes.

X' : transpuesta de la matriz X

$$\ln L(Y) = -\frac{n}{2} \ln(2\pi^2) - \frac{n}{2} \ln(2\sigma^2) - \frac{1}{2\sigma^2} Y'$$

;
Derivando respecto de β y de σ e igualando a 0.

la
Ventaja de resolver por MCO es que obtenemos medidas de ajuste confiables (R^2 y \hat{R}).

MODELANDO

en
Regresión múltiple necesitaremos 30 casos generales y se necesitan 10 casos por variable adicionales. Un exceso de variable independientes puede hacer subir artificialmente el R^2 .

DATOS AUSENTES

el
Programa desprecia cualquier individuo a quien le falte uno de los valores de la variable incluidas en el análisis.

CREACION DE VARIABLES DUMMY: es una variable dicotómica (0,1). El propósito es obtener información sobre la influencia de cada escalón de una variable ordinal o factor puro.

REGRESION DE POLINOMIOS FRACCIONALES: esta opción consiste en que STATA calcula automáticamente los polinomios que mejor ajuste consiguen.

CONFIGURAR EL INDIVIDUO DE REFERENCIA: no mejora el ajuste, pero es muy útil para que los resultados sean más inteligibles y la potencia de su interpretación mayor.

VALORES IMPROBABLES

un
Outlayer (valor improbable o imposible) con una base de datos con pocos grados de libertad puede confundir mucho las estimaciones.

LA LINEALIZACION: en factores ordinales es adecuado la creación de dummies, en el caso de variables cuantitativas se suele tomar logaritmos o polinomios.

REGRESION COX-BOX: se utiliza cuando la relación no es lineal.

$$Z(\lambda) = \frac{Y^\lambda - 1}{\lambda} \text{ si } \lambda \neq 0; \ln(Y) \text{ si } \lambda = 0$$



