

Regresión Lineal Multiple

Restricciones de la Regresión Lineal

1. Linealidad
2. Homocedasticidad
3. Normalidad Multivariable
4. Independencia de los errores
5. Ausencia de multicolinealidad. El modelo es incapaz de distinguir los efectos de una variable dummy.

Variable Dummy

Cuando necesitemos construir un modelo con variables ficticias (variables dummy), hay que omitir uno de los factores, uno de los niveles de la variable ficticia. Es decir, si tenemos 100 países como variables ficticias, solo colocaríamos como dummy 99 en el modelo de regresión.

P-Valor

Que no es

- El p-valor no es la probabilidad de que la afirmación sea cierta.
- El p-valor no es la probabilidad de que la hipótesis nula sea cierta.

Nos indica que tan probable es obtener un resultado con una hipótesis nula verdadera.

[\[\[1\]\]](#)

Paso a paso en la regresión lineal multiple

Añadir todas las variables independientes

No, por dos razones:

1. NO por añadir más variables vamos a tener más información. In Basura = Out Basura.
2. Si el número de variables va creciendo hace difícil la explicación lógica del proceso.

5 métodos obtención de variables relevantes

Tenemos 5 métodos disponibles para obtener las variables importantes y que tendrán relevancia en el algoritmo de regresión lineal multiple.

Exhaustivo (All-in)

Metemos todas las variables en el modelo, razones por las que hacer esto:

- Conocimiento previo de todas las variables. Todas son variables predictoras.
- Por necesidad, nos obligan a utilizar todas las variables.
- Preparación previa para realizar la eliminación hacia atrás.

Eliminación hacia atrás

1. Selección del nivel de significación en el modelo, normalmente $SL=0.05$
2. Se calcula el modelo con todas las variables
3. Se obtiene la variable predictoras con el p-valor más grande. Si $P > SL$, entonces pasamos al paso 4, sino vamos a fin.
4. Se elimina la variable predictora.
5. Reajuste del modelo sin dicha variable.

Con el nuevo modelo creado, las variables de ese tendrán una serie de p-valores y por tanto repetimos el paso 3, c

Selección hacia adelante

1. Seleccionamos un nivel de significación, pero en este caso será para entrar en el modelo.
2. Ajustamos todos los modelos de regresión lineal simple. Elegimos el que tiene el menor p-valor.
3. Conservamos esta variables y ajustamos todos los modelos con la variable extra añadida a la que ya tenga el modelo en ese momento.
4. Consideramos la variable predictora con el menor p-valor. Si $P < SL$ volvemos al paso 3.

Seguiremos sucesivamente añadiendo variables mientras el p-valor sean inferior al nivel de significación, en el ca

Eliminación bidireccional o regresión dual

1. Seleccionamos dos niveles de significación para entrar y salir del modelo.
2. Selección hacía delante $p\text{-valor} < SL_{\text{Enter}}$
3. Selección hacía atrás $p\text{-valor} < SL_{\text{Stay}}$
4. No hay nuevas variables para entrar ni tampoco variables antiguas para salir

Comparaciones de puntos

1. Seleccionar un criterio de la bondad de ajuste. Cuando un modelo será mejor que otro.
2. Construir todos los Modelos: $2^N - 1$
3. Selección del modelo con mejor criterio

OJO! -> 10 columnas de datos = 1023 modelos

Revision #1

Created 29 November 2023 00:18:42 by adminROM

Updated 29 November 2023 00:19:06 by adminROM