Sintaxis

Sintaxis	Modelo	Observaciones
`y ~ x`	$y = \beta_0 + \beta_1 x$	Línea recta con intercepto implícita
`y ~ -1 + x`	$y = \beta_1 x$	Línea recta sin intercepción; es decir, un ajuste forzado (0,0)
`y ~ x + I(x^2)`	$y = \beta_0 + \beta_1 x + \beta_2 x^2$	Modelo Polinomial; `I()` permite símbolos matemáticos
`y ~ x + z`	$y = \beta_0 + \beta_1 x + \beta_2 z$	Regresión Lineal Múltiple
`y ~ x:z`	$y = \beta_0 + \beta_1 xz$	Modelo con interacción entre x y z
`y ~ x*z`	$y = \beta_0 + \beta_1 x + \beta_2 z + \beta_3 xz$	Equivalente a `y~x+z+x:z`

Formato stargazer para summary

library(stargazer)

stargazer(ols, type = "text")

Selección del mejor modelo de regresión lineal

Después de hacer comprobado el mejor modelo, se hace summary(mod) y se quitan las variables mayores de 0.05 en Pr(>|t|). Después comparamos R2, AIC y BIC con el anterior modelo.

> Opción 1

step(mod,direction="backward")

> Opción 2

mod_cte= lm(wage~1,data=wage1)

step(mod_cte,direction="forward",scope=list(lower=mod_cte, upper=mod))

> Opción 3

incluir no-linealidad en un modelo de regresión

model1 <- lm(wage ~ educ, data = wdata)

> lineal

model2 <- lm(wage ~ educ+I(educ^2), data = wdata)

> cuadrático

model3 <- lm(log(wage) ~ educ, data = wdata)

> exponencial

10. Extracción del estadístico F y su p-valor

fstat.all <- output$fstatistic[1]

Diagnosis regresión lineal múltiple (1 extra)

Relativos a los residuos

> Los errores del modelo han de seguir una distribución normal (1er gráfico) con media 0 y varianza σ2 constante (2do gráfico)

Relativos al modelo

> Los puntos han de ajustarse a la estructura lineal considerada (3er gráfico)

Relativos a las observaciones anómalas

> Puede que algunas de las observaciones no se ajusten al modelo, comprometiendo su validez general (4to gráfico)

Relativos a las variables independientes

> Ninguna de las variables explicativas es combinación lineal de las demás. En el caso de regresión múltiple es de especial interés el fenómeno de la colinealidad (o multicolinealidad). Cuando algunas variables explicativas estén altamente correlacionadas entre sí, tendremos una situación de alta colinealidad. En este caso las estimaciones de los parámetros pueden verse seriamente afectadas:

Problemas de colinealidad

> Tendrán varianzas muy altas (serán poco eficientes)
> Habrá mucha dependencia entre ellas (al modificar ligeramente el modelo, añadiendo o eliminando una variable o una observación, se producirán grandes cambios en las estimaciones de los efectos).
> Incluso puede ocurrir que el contraste de regresión sea significativo (alto coeficiente de determinación), pero los contrastes individuales sean no significativos.

Diagnosis regresión lineal múltiple (1 extra) (cont)

Por último

> confirméis con evidencias numéricas y contrastadas las hipótesis de homocedasticidad de los residuos y la existencia de una relación lineal entre los valores ajustados y los residuos.

Sponsored by **Readable.com**
Measure your website readability!
https://readable.com