Práctica 1
GRETL. Regresión Simple

Estadística-II. INTRODUCCIÓN a la ECONOMETRÍA. UC3M

1. Utilizamos la base de datos ATTEND.RAW de Wooldridge para estudiar la relación entre la asistencia a clase y las notas en los exámenes finales en la Universidad. La asistencia a clase está medida por la proporción de clases a las que ha asistido el estudiante ($atndrte$) y la nota en los exámenes finales está estandarizada ($stndfnl$).

   a) Descargue la base de datos e importe los datos en *gretl*. Haga gráficos para representar la distribución univariante de cada variable y obtener los estadísticos descriptivos.

   b) Realiza un gráfico de dispersión de ambas variables con $atndrte$ en el eje de las $x's$ (abcisas).

   Utilice el modelo lineal

   $$stndfnl = \beta_0 + \beta_1 atndrte + \varepsilon \tag{1}$$

   para explicar la relación entre la proporción de clases atendidas y las notas en el examen final, donde β_0 y β_1 son parámetros desconocidos y ε es un término de error.

   a) ¿Es este modelo razonable? ¿Hay variables en ε que pudieran explicar de forma significativa en media la nota final?

   b) Estime el modelo por mínimos cuadrados ordinarios (MCO) y presente los resultados en el formato habitual.

   c) Compute los valores predichos $\widehat{stndfnl}_i = \hat{\beta}_0 + \hat{\beta}_1 atndrte_i$ donde $\hat{\beta}_0$ y $\hat{\beta}_1$ son los estimadores MCO de β_0 y β_1 y compute los residuos $\hat{\varepsilon}_i = stndfnl_i - \widehat{stndfnl}_i$, $i = 1, \ldots, n$.

   d) Calcule la media de los residuos $\hat{\varepsilon}_i$, y su covarianza con la varible $atndrte$ (x) y con la variable de las predicciones $\left(\widehat{stndfnl}\right)$. Comente sobre los resultados obtenidos.

   ¿Qué relación tienen esas cantidades con estos cálculos:

   $$\sum_{i=1}^{n} \hat{\varepsilon}_i, \quad \sum_{i=1}^{n} \hat{\varepsilon}_i \cdot atndrte_i \quad \sum_{i=1}^{n} \hat{\varepsilon}_i \cdot \widehat{stndfnl}_i?$$

   e) Calcule la media muestral de las variables $stndfnl$ (y) y $\widehat{stndfl}$ $(\hat{y})$. Comente sobre los resultados obtenidos.

   f) Calcule las varianzas muestrales $\widehat{Var}\,(stndfnl)$, $\widehat{Var}\left(\widehat{stndfnl}\right)$, y $\widehat{Var}\,(\hat{\varepsilon})$, y compruebe que

   $$\widehat{Var}\,(stndfnl) = \widehat{Var}\left(\widehat{stndfnl}\right) + \widehat{Var}\,(\hat{\varepsilon})$$

y que

$$\hat{R}^2 = \frac{\widehat{Var}\left(\widehat{stndfnl}\right)}{\widehat{Var}\left(stndfnl\right)} = 1 - \frac{\widehat{Var}\left(\widehat{\varepsilon}\right)}{\widehat{Var}\left(stndfnl\right)}.$$

g) Comprueba el cálculo del error estándar de $\hat{\beta}_1$. ?'Qué información proporciona?

2. Ahora considere el modelo

$$stndfnl = \delta \cdot atndrte + U$$

donde δ es un parámetro, posiblemente diferente de β_1 y U es un término de error, posiblemente diferente de ε.

a) Estime este nuevo modelo por MCO y compare los estimadores de β_1 y de δ.

b) Compute los valores predichos $\widetilde{stndfnl}_i = \tilde{\delta} \cdot atndrte_i$, $i = 1, \ldots, n$; donde $\tilde{\delta}$ es el estimador MCO de δ y compute los residuos $\tilde{U}_i = stndfnl_i - \widetilde{stndfnl}_i$, $i = 1, \ldots, n$.

c) Calcule

$$\sum_{i=1}^{n} \tilde{U}_i, \quad \sum_{i=1}^{n} \tilde{U}_i \cdot atndrte_i \quad \text{y} \quad \sum_{i=1}^{n} \tilde{U}_i \cdot \widetilde{stndfnl}_i.$$

Comente sobre los resultados obtenidos.

d) Calcule las medias muestrales de las predicciones y de la variable dependiente,

$$\bar{\tilde{y}} = \frac{1}{n} \sum_{i=1}^{n} \widetilde{stndfnl}_i \quad \text{y} \quad \bar{y} = \frac{1}{n} \sum_{i=1}^{n} stndfnl_i.$$

Comente sobre los resultados obtenidos.

3. Considere la relación entre medias muestrales

$$\overline{stndfnl} = \beta_0 + \beta_1 \overline{atndrte} + \bar{\varepsilon}$$

y el modelo (1) en diferencias con respecto a la media muestral,

$$y_i' = \beta_1 x_i' + V_i, \quad i = 1, \ldots, n \tag{2}$$

donde

$$\begin{aligned} y_i' &= y_i - \bar{y} = stndfnl_i - \overline{stndfnl} \\ x_i' &= x_i - \bar{x} = atndrte_i - \overline{atndrte} \\ V_i &= \varepsilon_i - \bar{\varepsilon}. \end{aligned}$$

a) Estime β_1 por MCO en el modelo (2), sin constante, donde la variable dependiente es $y_i' = stndfnl_i - \overline{stndfnl}$ y la variable explicativa es $x_i' = atndrte_i - \overline{atndrte}$. Compare este valor estimado con el obtenido en 1(b).

b) ?'Cuál es la relación entre las predicciones $\widetilde{y'}_i = \widetilde{stndfnl_i - \overline{stndfnl}}$, $\widehat{y}_i = \widehat{stndfnl}_i$ y $\widetilde{y}_i = \widetilde{stndfnl}_i$?