Práctica 1
GRETL. Simple Regression

Estadística-II. INTRODUCCIÓN a la ECONOMETRÍA. UC3M

1. We use Wooldridge's dataset ATTEND.RAW to study the relation between class assistance and final exam results at the university. Assistance is measured as the proportion of classes attended by the student (*atndrte*) and exam results are standardized (*stndfnl*). Use the simple linear regression model

   (a) Download the data set and import it to *gretl*. Plot graphs to represent the univariate distribution of each variable and obtain descriptive statistics.

   (b) Plot a dispersion graph for both variables with *atndrte* in the $x's$ axis.

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

   to explain the relation between the rate of class assistance and final marks, where β_0 and β_1 are unknown parameters and ε is an error term.

   (a) Are the model assumptions reasonable? Could there be effects from unobserved variables ε which could partly explain on average the final exam result?

   (b) Estimate the model by OLS and present results in the standard format.

   (c) Compute $\widehat{stndfnl}_i = \hat{\beta}_0 + \hat{\beta}_1 atndrt_i$, $i = 1, ..., n$, where $\hat{\beta}_0$ and $\hat{\beta}_1$ are the OLS estimators for β_0 and β_1. Also, compute residuals $\hat{\varepsilon}_i = stndfnl_i - \widehat{stndfnl}_i$, $i = 1, ..., n$.

   (d) Get the residuals' average and the sample covariance with $atndrt_i$ and $\widehat{stndfnl}_i$.Discuss the results. How these sample moments relate to

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

   (e) Get the sample average for $stndfnl$ (y) and $\widehat{stndfl}$ $(\hat{y})$. Discuss the results.

   (f) Compute the sample variances $\widehat{Var}(stndfnl)$, $\widehat{Var}\left(\widehat{stndfnl}\right)$, $\widehat{Var}(\hat{\varepsilon})$ and check that

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

   and that

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

   (g) Check the value of the standard error for $\hat{\beta}_1$. what information does it entail?

2. Consider now the following model

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

   where δ is a parameter, perhaps different from β_1, and U is an error term, perhaps different from ε.

   (a) Estimate this new model by OLS and compare the estimates for β_1 and from δ.

   (b) Compute $\widetilde{stndfnl}_i = \tilde{\delta} atndrt_i$, $i = 1, ..., n$, where $\tilde{\delta}$ is the OLS estimator for δ and also compute residuals $\tilde{U}_i = stndfnl_i - \widetilde{stndfnl}_i$, $i = 1, ..., n$.

(c) Get $\sum_{i=1}^{n} \tilde{U}_i$, $\sum_{i=1}^{n} \tilde{U}_i \cdot atndrt_i$ and $\sum_{i=1}^{n} \tilde{U}_i \cdot \widetilde{stndfnl}_i$. Discuss the results.

(d) Get the sample averages for the model predictions and the dependent variable,

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

Discuss the results.

3. Consider the equality

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

and model (1) in differences with respect to the sample average

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

where

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

(a) Estimate β_1 by OLS in model (2) without a constant, where the dependent variable is $y_i' = stndfnl_i - \overline{stndfnl}$ and the explanatory variable is $x_i' = atndrte_i - \overline{atndrte}$. Compare this value with that obtained in 1.b.

(b) What is the relation between the predictions $\widetilde{y'}_i = \widetilde{stndfnl_i - \overline{stndfnl}}$, $\widehat{y}_i = \widehat{stndfnl}_i$ and $\widetilde{y}_i = \widetilde{stndfnl}_i$?