Universidad del País Vasco / Euskal Herriko Unibertsitatea

CAMPUS OF INTERNATIONAL EXCELLENCE

Exercise E5

The Multiple Regression Model. Estimation

Pilar González and Susan Orbe

Dpt. Applied Economics III (Econometrics and Statistics)

Contents

Contents

E5.1. Beach umbrella rental.

Part I. Simple Linear Regression Model.

Estimate a simple linear regression model to determine the number of rented umbrellas as a function of the temperature:

$$U_t = \alpha + \beta T_t + u_t \quad t = 1, \ldots, 22 \tag{1}$$

a. Estimate the coefficients of the model by OLS. Save the fitted values and the residuals.

b. Write down the Sample Regression Function.

c. Interpret the slope coefficient and discuss its sign.

d. Give an interpretation to the coefficient of determination.

e. Plot the actual and the fitted values and comment on the results.

E5.1. Beach umbrella rental.

g. How many beach umbrellas have been rented in this period?

h. What is the sample mean of the number of rented umbrellas? What is the sample mean of the number of rented umbrellas estimated using the regression model? Is the sample mean of the actual values, U, different from the sample mean of the fitted values, $\hat{U}$? Why?

i. Estimate the number of rented umbrellas in the first week of August.

j. Compute the estimation error made when estimating the number of rented umbrellas in the last week of September. How is this estimation error called? What are the sources of this estimation error?

k. What is the estimated number of rented umbrellas when the average temperature is 26 degrees Celsius?

l. What is the estimated change in the number of rented umbrellas when the average temperature increases by 2 degrees Celsius?

E5.1. Beach umbrella rental.

Part II. General Linear Regression Model.

Estimate a regression model to determine the number of rented umbrellas as a linear function of the temperature and the price:

$$U_t = \beta_1 + \beta_2 \, T_t + \beta_3 \, P_t + v_t \quad t = 1, \ldots, 22 \tag{2}$$

a. Estimate the coefficients of the model by OLS. Save the fitted values and the residuals.

b. Write down the Sample Regression Function.

c. Interpret the estimated coefficients for the variables temperature and price. Comment on their signs.

d. Compare models (1) and (2). What are the differences between them?

e. Do the estimates $\hat{\beta}$ and $\hat{\beta}_2$ in models (1) and (2) match? Why?

E5.1. Beach umbrella rental.

f. Give an interpretation to the coefficient of determination of model (2) and compare its value to the one obtained in model (1).

g. Plot the actual and fitted values of U. Comment on this graph comparing it with the one obtained for model (1).

h. Consider all the variables included in model (2) (the dependent variable and the regressors), the fitted values and the residuals. Compute the main descriptive statistics for all of them and comment on the results.

i. What is the estimated number of rented umbrellas when the average temperature is 39 degrees Celsius? And when the price is €13?

j. If the family firm decided to charge the same amount throughout the season, what would be the consequences in the estimation results?

E5.1. Beach umbrella rental.

Part III. General Linear Regression Model.

Estimate a linear regression model to determine the number of rented umbrellas as a function of temperature, price and whether it is a windy week or not:

$$U_t = \gamma_1 + \gamma_2 T_t + \gamma_3 P_t + \gamma_4 WW_t + w_t \quad t = 1, \ldots, 22 \tag{3}$$

a. Estimate the model by OLS and write down the Sample Regression Function.

b. Compare models (2) and (3). What are the differences between them?

c. What is the estimated number of rented umbrellas in a windy week? And in a non windy week?

d. Give and interpretation to the coefficient of the dummy variable WW. Comment on its sign.

e. Estimate the change in the number of rented umbrellas when the price increases by €1, holding the rest of the regressors constant.

f. Estimate the number of rented umbrellas when the average temperature is 30 degrees Celsius and the average price is €7. What is this estimate for a windy week? And for a non windy week?

Contents

E5.2. Holiday cottages.

Model A.

Consider a regression model that specifies a linear relationship among room prices and the number of rooms and the price of breakfast:

$$RP_i = \alpha_1 + \alpha_2\, NR_i + \alpha_3\, BP_i + u_i \tag{4}$$

a. Estimate the coefficients of model (4) by OLS and write down the Sample Regression Function.

b. Interpret the estimated coefficients of NR and BP.

c. What is the estimated price of a room when the holiday cottage has 10 bedrooms and breakfast is included in the price?

d. Assuming that breakfast is included in the price of a room, estimate the difference in the price between a holiday cottage with 15 bedrooms and one with 10 bedrooms.

e. Plot the actual and fitted values for room prices. Comment on the result.

f. Do you think that model (4) is appropriate for this data, that is, do you think this model is able to explain the variation in the prices observed in the sample?

E5.2. Holiday cottages.

Model B.

Include two explanatory variables more in model (4): access to WiFi and location.

$$RP_i = \lambda_1 + \lambda_2\, NR_i + \lambda_3\, BP_i + \lambda_4\, WIFIF_i + \lambda_5\, WIFIP_i + \lambda_6\, LOCC_i + u_i \qquad (5)$$

where $WIFIF$ takes the value 1 if the holiday cottage offers free WiFi access and 0 otherwise; $WIFIP$ takes the value 1 if the holiday cottage offers WiFi access for an additional fee and 0 otherwise; and, $LOCC$ takes the value 1 if the holiday cottage is in the town center and 0 otherwise.

a. Estimate the coefficients of model (5) by OLS and write down the Sample Regression Function.
b. Interpret the estimated coefficients of the dummy variables $WIFIF$ and $WIFIP$. Do they have the expected signs?
c. Estimate the room price when the holiday cottage has 6 bedrooms, offers access to WiFi and the breakfast price is €3. What is the estimated room price when the access to WiFi is free? And if you have to pay €2?
d. Estimate the room price for a cottage with the same characteristics as the first one in the sample. Is this estimate different from the actual price? Why?

E5.2. Holiday cottages.

Model C.

Consider the linear regression model:

$$\begin{aligned} RP_i &= \beta_1 + \beta_2\, NR_i + \beta_3\, BP_i + \beta_4\, WIFIF_i + \beta_5\, NPR_i + \\ &+ \beta_6\, BER_i + \beta_7\, LKR_i + u_i \end{aligned} \tag{6}$$

where NPR, BER and LKR take the value 1 if the holiday cottage is less than 1 km from a natural park, a beach or a lake, respectively; and 0 otherwise.

a. How many explanatory variables are there in the model? What are the differences between models (6) and (5)? Estimate the coefficients of model (6) by OLS and write down the Sample Regression Function.
b. Consider the whole data set. What dummy variables have been included in model (6) and what dummy variables have been excluded? Give an interpretation to the estimated coefficients of NPR, BER and LKR. Do they have the expected signs?
c. Do you think that the fit of model (6) is better than the fit of the previous models? Why?

Contents

E5.3. Soy milk.

Part I. Data file organization.

A firm that produces soy milk wants to analyse its sales (S, in thousands of euros) as a function of the price per litter (P, in euro cents) and advertising expenditures (AE, in hundreds of euros). The available monthly data from January of 1990 until June of 2012 may be found in the data file `soymilk.gdt`.

a. Open the file `soymilk.gdt` and give a time series structure to the data set.

b. Y stands for sales, $X1$ for price and $X2$ for advertising expenditures in the data file `soymilk.gdt`. Edit the attributes of these three variables, changing their names and including the units of measurement.

c. Save all the changes in the data file `soymilk-sales.gdt`.

E5.3. Soy milk.

Part II. $S = f(P)$.

Consider the simple linear regression model relating sales and prices:

$$S_t = \gamma_1 + \gamma_2 P_t + u_t \quad t = 1990:1, \ldots, 2012:6. \tag{7}$$

a. Analyse the information in the data using the main descriptive statistics. What is the range of sales in the sample? What is the sample mean of sales in this period?
b. Compute the correlation coefficient between sales and price.
c. Estimate the coefficients of the regression model by OLS and write down the Sample Regression Function.
d. Interpret the estimated coefficient of price. Comment on its sign.
e. Estimate the variation in sales when the price increases by 30 euro cents.
f. Give an interpretation to the coefficient of determination.
g. Estimate the variance of the error term (σ^2).
h. Estimate the variance of the OLS estimator $\hat{\gamma}_2$?
i. Plot the actual and fitted values for sales. Comment on the result.

E5.3. Soy milk.

Part III. $S = f(P, AE)$.

Consider the regression model where sales depend on prices and advertising expenditures as follows:

$$S_t = \beta_1 + \beta_2 \, P_t + \beta_3 \, AE_t + \beta_4 \, AE_t^2 + u_t \quad t = 1990:1, \ldots, 2012:6. \tag{8}$$

a. How many explanatory variables are there in the model? What are the differences between models (7) and (8)?

b. Does this model satisfy the linearity assumption?

c. Compute the correlation matrix of variables sales, price and advertising expenditures. Do the simple correlation coefficients have the expected signs?

d. Estimate the model by OLS and write down the Sample Regression Function.

e. What is the estimated change in sales when the expenditure on advertising increases by €100 and the price remains constant? And if the firm's advertising expenditures are €1500 at that moment? And if they are 15000 euros?

f. What is the estimated variation in sales when the price increases by half a euro, holding advertising expenditures constant? And if the firm's price is 125 cents at that moment? And if it is 80 cents?

g. What are the estimated sales for December 1990? What is the difference between this estimate and the actual value?

h. Estimate the sales in July of 2012 knowing that the expenditure on advertising was €14600 and the price per litter €1.23.

i. Estimate the covariance matrix of the OLS estimators.

j. Plot the actual and fitted sales. Comment on the result.

E5.3. Soy milk.

Part IV. Trend.

Consider the regression model in Part III and include a time trend as regressor:

$$S_t = \alpha_1 + \alpha_2\, P_t + \alpha_3\, AE_t + \alpha_4\, AE_t^2 + \alpha_5\, time_t + u_t \quad t = 1990:1, \ldots, 2012:6. \tag{9}$$

a. Add a trend variable to the data set and estimate model (9) by OLS.

b. Plot the actual and fitted values for sales. Comment on the result.

c. Give an interpretation to the estimated coefficient of the trend variable.

d. Given all the results obtained so far, do you think that the trend variable might be relevant to explain the evolution of sales? Why?