

Problem Set VII

Content: Simple Linear Regression II

Question 1

Discuss which of the following population regression functions are linear in parameters.

- a. $Y_i = \beta_0 + \beta_1 X_i + u_i$
- b. $Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + u_i$
- c. $Y_i = \beta_0 X_{1i}^{\beta_1} X_{2i}^{\beta_2} e^{u_i}$
- d. $Y_i = \beta_0 + \beta_1 10^{X_i} + u_i$
- e. $Y_i = \beta_0 + \ln(\beta_1 + X_i) + u_i$
- f. $Y_i = \beta_0 X_i^{\beta_1} + u_i$

Solution:

- a) and b) are linear in parameters.
- c) $\ln Y_i = \ln \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$ after taking logarithm of both sides, become linear in parameters.
- d) linear in parameters.
- e) and f) are nonlinear in parameters.

Question 2

Consider a simple linear model,

$$Y_t = \beta_0 + \beta_1 X_t + u_t$$

where $t = 1, 2, \dots, T$ and the model satisfies the assumptions. Show that OLS estimator $\widehat{\beta}_1$ is the unbiased estimator for β_1 and find the variance of $\widehat{\beta}_1$.

Solution: We know that

$$\widehat{\beta}_1 = \frac{\sum x_t y_t}{\sum x_t^2} = \frac{\sum x_t Y_t}{\sum x_t^2}$$

and let, $a_t = \frac{x_t}{\Sigma x_t^2}$. Thus,

$$\begin{aligned}\hat{\beta}_1 &= \Sigma a_t Y_t \\ \hat{\beta}_1 &= \Sigma a_t (\beta_0 + \beta_1 X_t + u_t) \\ \hat{\beta}_1 &= \beta_0 \Sigma a_t + \beta_1 \Sigma a_t X_t + \Sigma a_t u_t \\ \hat{\beta}_1 &= \beta_1 + \Sigma a_t u_t\end{aligned}$$

Thus, $\hat{\beta}_1$ is a linear estimator. Taking the expected value of both sides,

$$E(\hat{\beta}_1) = E(\beta_1) + \Sigma a_t E(u_t)$$

Since $E(u_t) = 0$, we get

$$E(\hat{\beta}_1) = \beta_1$$

Thus, $\hat{\beta}_1$ is an unbiased estimator.

$$\begin{aligned}Var(\hat{\beta}_1) &= E[\hat{\beta}_1 - E(\hat{\beta}_1)]^2 \\ Var(\hat{\beta}_1) &= E[\hat{\beta}_1 - \beta_1]^2 \\ Var(\hat{\beta}_1) &= E[\Sigma a_t u_t]^2 \\ Var(\hat{\beta}_1) &= E[a_1 u_1 + a_2 u_2 + \dots + a_T u_T]^2 \\ Var(\hat{\beta}_1) &= E[a_1^2 u_1^2 + a_2^2 u_2^2 + \dots + a_T^2 u_T^2 + 2crossterms] \\ Var(\hat{\beta}_1) &= a_1^2 E(u_1^2) + a_2^2 E(u_2^2) + \dots + a_T^2 E(u_T^2) + 2E(crossterms) \\ Var(\hat{\beta}_1) &= \Sigma a_t^2 E(u_t^2) \\ Var(\hat{\beta}_1) &= \frac{\sigma^2}{\Sigma x_t^2}\end{aligned}$$

Question 3

Auditors are often required to compare the audited (or current) value of an inventory item with the book (or listed) value. If a company is keeping its inventory and books up to date, there should be a strong linear relationship between the audited and book values. A company sampled ten inventory items and obtained the audited and book values given in the accompanying table.

Item	Audit Value (y_i)	Book Value (x_i)
1	9	10
2	14	12
3	7	9
4	29	27
5	45	47
6	109	112
7	40	36
8	238	241
9	60	59
10	170	167

- Fit the model $Y = \beta_0 + \beta_1 x + \varepsilon$ to these data.
- If the book value is $x = 100$, what would you use to estimate the audited value?
- Write 95% confidence interval for β_1 .

Solution: Summary statistics are:

$$\bar{X} = 72, \quad \bar{Y} = 72.1, \quad S_{xy} = \sum_i (X_i - \bar{X})(Y_i - \bar{Y}) = 54.243, \quad S_{xx} = \sum_i (X_i - \bar{X})^2 = 54.714$$

$$\begin{aligned} \text{a. } \hat{\beta}_1 &= \frac{S_{xy}}{S_{xx}} = \frac{54.234}{54.714} = 0.991392 \quad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} = 72.1 - 0.991392 * 72 = 0.72 \\ \hat{Y} &= 0.72 + 0.99X \end{aligned}$$

$$\text{b. When } x = 100, \text{ the best estimate of } Y \text{ is } \hat{Y} = 0.72 + 0.99 * 100 = 99.72$$

$$\text{c. } 100(1 - \alpha)\% \text{ confidence interval for } \beta_1 :$$

$$\hat{\beta}_1 \pm t_{\frac{\alpha}{2}, n-k} se(\hat{\beta}_1)$$

$$t_{0.025, 8} = 2.306$$

$$se(\hat{\beta}_1) = \sqrt{Var(\hat{\beta}_1)} = \sqrt{\frac{\hat{\sigma}^2}{\sum (X_i - \bar{X})^2}} = \sqrt{\frac{SSR/(10-2)}{54714}} = \sqrt{\frac{56.8454/8}{54714}} = \sqrt{\frac{7.1056}{54714}} = 0.011396$$

$$0.99 \pm 2.306 * 0.011396 = (0.9637, 1.0162)$$

Question 4

Assume that simple linear regression model $Y_i = \beta_1 X_i + u_i$ is appropriate for the data consisted of driving a car at constant speed 50 km/h and observing with 5 minute intervals the distance you have gone. Assume that the errors are iid normal with 0-mean and constant variance σ^2 .

Time	Distance
5	4.4
10	14.5
15	14.1
20	18
25	19.6
30	26.7
35	28.3
40	32.4
45	38
50	43.2
55	45.5
60	49.7

- Write the normal equation(s).
- Fit a simple linear regression model relating Y to X.
- Estimate σ^2 .
- Write 95% confidence interval for β_1 .

Solution:

Data Summary:

$$n = 12, \sum_{i=1}^n X_i = 390 \quad \sum_{i=1}^n Y_i = 334.4 \quad \sum_{i=1}^n X_i Y_i = 13760.5$$

$$\sum_{i=1}^n X_i^2 = 16250 \quad \sum_{i=1}^n Y_i^2 = 11550.7 \quad \bar{X} = 32.5 \quad \bar{Y} = 27.866$$

- $\sum u_i^2 = \sum (Y_i - \beta_1 X_i)^2 = S$
 $\frac{\partial S}{\partial \beta_1} = -2 \sum (Y_i - \beta_1 X_i) X_i = 0 \rightarrow \sum (Y_i X_i - \beta_1 X_i^2) = 0$
 $\rightarrow \sum Y_i X_i - \beta_1 \sum X_i^2 = 0 \rightarrow \text{normal equation } \hat{\beta}_1 = \frac{\sum X_i Y_i}{\sum X_i^2}$
- $\hat{\beta}_1 = \frac{\sum X_i Y_i}{\sum X_i^2} = \frac{13760.5}{16250} = 0.84126$
- $\hat{\sigma}^2 = \frac{\sum u_i^2}{n-1} = \frac{\sum (Y_i - \hat{Y}_i)^2}{n-1}$ and we have $\hat{Y}_i = \hat{\beta}_1 X_i$ from here; $\frac{\sum (Y_i - \hat{\beta}_1 X_i)^2}{n-1} = \hat{\sigma}^2$
 $\hat{\sigma}^2 = \frac{\sum Y_i^2 - 2\hat{\beta}_1 \sum Y_i X_i + \hat{\beta}_1^2 \sum X_i^2}{n-1} \rightarrow \hat{\beta}_1 = \frac{\sum X_i Y_i}{\sum X_i^2}$ then,
 $\hat{\sigma}^2 = \frac{\sum Y_i^2 - \hat{\beta}_1 \sum Y_i X_i}{n-1} = \frac{11550.7 - (0.84126)13670.5}{12-1} = 4.56865$

Question 5

In a study of credit unions, the following regression was fitted to eighty-seven quarterly observations:

$$\log y = 0.47 + 0.31 \log x \quad \bar{R}^2 = 0.86 \quad (0.13)$$

where y = output of loans, x = expenditure on capital equipment

The figure in the parenthesis under the least square parameter estimate is the estimated standard error.

- Interpret the estimated regression coefficients.
- Construct and interpret a 95% confidence interval for the true coefficient on $\log x$.
- Interpret the coefficient of determination.

Solution:

- A 1% increase in capital equipment expenditure leads to an expected 0.31% increase in loan output.

- Let β_1 be the coefficient of $\log x$. We will construct a 95% confidence interval for β_1 .

$$\hat{\beta} \pm t_{n-k, 0.025} \hat{\sigma}_{\hat{\beta}_1}$$

$$n = 87 \quad k = 2$$

$$\hat{\beta}_1 = 0.31 \quad \hat{\sigma}_{\hat{\beta}_1} = 0.13 \quad t_{n-2, 0.025} \approx 1.99$$

$$0.05 < \beta_1 < 0.57$$

- Variations in expenditure on capital equipment explain the 86% of the variations in output of loans. A good fit.

Question 6

Assume the population regression function is as follows:

$$Y_i = \beta \ln(X_t) + u_t$$

$$E(u_t) = 0 \quad \text{where } t = 1, \dots, T$$

$$E(u_t^2) = \sigma^2 \quad \text{where } t = 1, \dots, T$$

$$E(u_t u_s) = 0 \quad \text{for all } t \neq s$$

- Derive the OLS estimator of β .
- Derive the variance of β .

Solution:

- By $\frac{\partial SSR}{\partial \beta} = -2 \sum (Y_t - \hat{\beta} \ln(X_t)) \ln(X_t) = 0$

Then we will have $\sum y_t \ln(X_t) - \beta \sum (\ln(X_t))^2 = 0$ which implies

$$\sum y_t \ln(X_t) = \hat{\beta} \sum (\ln(X_t))^2 \text{ then we can find the OLS estimator as } \hat{\beta} = \frac{\sum Y_t \ln(X_t)}{\sum (\ln(X_t))^2}$$

b. First notice that $Y_t = \hat{\beta}X_t + u_t$ then substitute this into our estimator to get

$$\hat{\beta} = \frac{\beta \sum (\ln(X_t))^2 + \sum \ln(X_t) u_t}{\sum (\ln(X_t))^2}$$

then simplify to get

$$\hat{\beta} = \beta + \frac{\sum \ln(X_t) u_t}{\sum (\ln(X_t))^2}$$

Now we are ready to compute $E(\hat{\beta})$ then notice that $\frac{1}{\sum (\ln(X_t))^2}$ is a constant and $E(\sum \ln(X_t) u_t) = 0$ so we are left with $E(\hat{\beta}) = \beta$. $Var(\hat{\beta}) = E(\hat{\beta} - \beta)^2$

To compute this first define $a_t = \frac{\ln(X_t)}{\sum (\ln(X_t))^2}$ then we get the following

$$\hat{\beta} - \beta = \sum a_t u_t \text{ then } Var(\hat{\beta}) = E(\sum a_t u_t)^2$$

Then to compute this expression notice that

$$E(a_1 u_1 + \dots + a_t u_t)^2 = E(a_1^2 u_1^2 + \dots + a_t^2 u_t^2) + \text{cross products}$$

Then note that the cross products will be 0 since $E(u_i u_j) = 0$ when $i \neq j$.

$$\text{So } E(a_1 u_1 + \dots + a_t u_t)^2 = E(a_1^2 u_1^2 + \dots + a_t^2 u_t^2) = E(\sum a_t^2 u_t^2) = \sum a_t^2 E(u_t^2) = (\sum a_t^2) \sigma^2$$

$$\text{Finally notice } Var(\hat{\beta}) = \frac{\sigma^2}{\sum (\ln(X_t))^2}$$