Economics Tripos Part 1 Paper 3
Quantitative Methods in Economics
Supervisor: Selma Telalagic (st390@cam.ac.uk)

Supervision 6 (statistics): Linear regression - theory
and practise

Questions

1. Consider a set of observations $\{y_1, ..., y_N\}$ that can be described
by the following equation:

$$y_i = \alpha + \varepsilon_i, \qquad i = 1, ..., N.$$

Suppose the value of α is unknown and needs to be estimated.
The criterion used for estimation is Ordinary Least Squares. In
other words, $\hat{\alpha}$ is chosen to minimise the sum of squared errors:

$$\min_{\alpha} \sum_{i=1}^{N} \varepsilon_i^2 \qquad \text{where } \varepsilon_i = y_i - \alpha.$$

(a) Using calculus, solve this minimisation problem and obtain
an expression for $\hat{\alpha}$. Explain what this expression represents.

(b) If the error term ε_i satisfies all the conditions usually required
for a linear regression model, what desirable properties does the
estimator $\hat{\alpha}$ have? What does this say about the desirability of the
sample mean as an estimator?

(c) Now consider the following equation:

$$y_i = \alpha + \beta x_i + \varepsilon_i, \qquad i = 1, ..., N.$$

Using the least squares criterion, derive both $\hat{\alpha}$ and $\hat{\beta}$. What do
these expressions represent?

2. (a) Describe the method of Ordinary Least Squares, making
clear the conditions under which it can be used.

(b) If the error term ε satisfies all the conditions usually required
for a linear regression, then

$$E(\varepsilon_i) = 0 \qquad var(\varepsilon_i) = \sigma^2 \qquad cov(\varepsilon_i, \varepsilon_j) = 0 \qquad \text{for all } i \neq j$$

Show that the least squares estimator of the slope coefficient $\hat{\beta}$ has the following properties:

    i. $E(\hat{\beta}) = \beta$

    ii. $var(\hat{\beta}) = \frac{\sigma^2}{\sum_{i=0}^{n}(x_i - \bar{x})^2}$.

(c) An economist interested in the relationship between price of overnight accommodation and proximity of that accommodation to the sea collects the following data for 12 hotels:

Hotel	Distance to the nearest beach (km)	Price of standard room (£)
1	0.3	99
2	0.9	95
3	1.3	83
4	2.2	65
5	2.4	70
6	3.3	60
7	3.9	65
8	5.0	59
9	5.2	52
10	6.2	48
11	6.8	61
12	8.0	65

Using this data, determine the parameters of the following model:

$$y_i = \alpha + \beta x_i + \varepsilon_i, \qquad i = 1, ..., 12$$

(y_i is the price of the hotel and x_i is the distance to the beach for hotel i)

(d) Using your results assess the claim that the proximity of a hotel to a beach is a significant factor in the price charged by hotels.

3. The correlation coefficient, ρ_{xy}, is a measure of linear association between variables x and y.

(a) Describe the following relationships and draw a diagram illustrating each situation:

    i. $\rho_{xy} = 1$

    ii. $\rho_{xy} = -1$

iii. $\rho_{xy} = 0$

(b) What is the difference between correlation and causation? Does either imply the other?

(c) Give an expression for the sample correlation coefficient in terms of variances and covariances.

(d) Show that $\beta = r\frac{S_y}{S_x}$, where r is the sample correlation coefficient and S_y and S_x are the sample standard errors of x and y.

(e) Use the expression in the previous answer to find the sample correlation coefficient for the following set of data:

x	y
10	3
14	0
15	8
18	0
0	3
14	0
7	3
14	9
13	7
8	11
9	10
19	4
2	2
0	2
4	10

(f) Produce a scatter plot of the data.

(g) What kind of relationship would you expect between x and y?

(h) Use normal equations to obtain the regression equation of y on x.

(i) Test the hypothesis that the slope coefficient is significantly different from zero at the 5% level of significance.

(j) Test the hypothesis that the slope coefficient is significantly different from one at the 5% level of significance.

(k) Calculate the value of R^2 and comment on the relationship between x and y.

4. Find data on the unemployment rate and inflation rate for a country of your choice. You will need at least 20 observations. When submitting your solutions please provide a printout of your dataset. If you do not know how to use the Excel functions needed for this question, they can be found in Tools –> Data Analysis. If it is not installed in your Excel, go to Tools –> Add-ins –> Analysis Toolpak.

(a) Present a table with the mean, variance, maximum and minimum for each series.

(b) Provide a scatter plot of the data.

(c) Calculate the sample correlation coefficient.

(d) Using this information, what kind of relationship would you expect between the two variables?

(e) Report the coefficient estimates for the following regression equation:

$$Inflation_t = \alpha + \beta * unemployment_t + \varepsilon_t$$

(f) Test whether the coefficients are statistically different from zero.

(g) What do the results suggest about the Phillips curve in this country?

5. Tripos 2009 D1

6. Tripos 2007 D2

7. Tripos 2006 D2