



Assignment 4

(Release Date: 20/09/2017)

Answer the following questions and hand in your answers by **Sunday, September 24, 2017**. These should either be typed and sent electronically or handwritten, scanned and sent electronically (via E-mail at pkonstantinou@aueb.gr). Please make sure that your files do not exceed 5MB in size, otherwise the server at AUEB might block them!

Simple and Multiple Regression

1. Suppose that the following data are available:

x_i	3	12	6	20	14
y_i	55	40	55	10	15

- (a) Fit a straight line to the five data points in the table. Give the estimates of β_0 and β_1 . Plot the points and sketch the fitted line as a check on the calculations.
 - (b) Compute an estimate or the standard deviation of \hat{y}_0 when $x_0 = 8$ ($E(Y|X = 8^*)$).
 - (c) Find a 95% confidence interval for $E(Y|X = x^*)$ when $x^* = 8$.
 - (d) Estimate the standard deviation of an individual value of y when $x^* = 8$.
 - (e) Develop a 95% prediction interval for y when $x^* = 8$.
2. Fire damage in Greece amounts to millions of euros, much of it insured. The time taken to arrive at the fire is critical. This raises the question, Should insurance companies lower premiums if the home to be insured is close to a fire station? To help make a decision, a study was undertaken wherein a number of fires were investigated. The distance to the nearest fire station (in kilometers) and the percentage of fire damage were recorded (available in Assignemnt4Data.xlsx).
- (a) Determine the least squares line and interpret the coefficients.
 - (b) Predict with 95% confidence the percentage loss due to fire for a house that is 8 kilometers away from the nearest fire station.
 - (c) Estimate with 95% confidence the average percentage loss due to fire for houses that are 5 kilometers away from the nearest fire station.

3. To assess how the number of housing starts is affected by mortgage rates a colleague has collected data on the average mortgage rate and the number of housing starts in Greece for the past 10 years. These data are listed here tab Housing of Assignemnt4Data.xlsx.

- (a) Determine the least squares (regression) line.
- (b) What do the coefficients of the regression line tell you about the relationship between mortgage rates and housing starts?
- (c) Estimate with 90% confidence the mean monthly number of housing starts when the mortgage interest rate is 7%.

4. Using data on $n = 341$ families, I have estimated the following multiple regression where the depend variable is (annual) family income and the explanatory variables are AGE (age in years) of the main earner, EDUC (education in years) of the main earner, EARNRS (Number of family members earning money) and CHILDS (Number of children):

$$\widehat{\text{income}} = -78,202 + 426.6\text{AGE} + 5297\text{EDUC} + 3097\text{EARNRS} - 1605\text{CHILDS} \quad (1)$$

- (a) Interpret the estimated coefficients. Does the constant term make sense?
- (b) Now suppose that, I have re-estimated the above regression on a different sample using data on $n = 524$ families and found that

$$\widehat{\text{income}} = 62,202 + 412.3\text{AGE} + 6422\text{EDUC} + 2979\text{EARNRS} - 1504\text{CHILDS}. \quad (2)$$

Interpret the estimated coefficients. Does the constant term make sense?

- (c) Suppose that AGE =34 years, EDUC = 16 years, EARNRS = 1 and CHILDS=2. Using the results in (2), what is the expected income for this family?

5. Using data on 30 individuals we have calculated the correlation between their monthly income and their (self reported) level of happiness (ranging from 0 to 10), and found that $r = 0.91$. Is this correlation statistically significant in the population? Perform the test at the 5% significance level.