

Econometrics 2018, Final Exam, Time to complete : 60 minutes

Minimum points required for a positive grade: 20

Name and student ID _____

Signature _____

This exam contains 4 pages (including this cover page) and 2 questions.
Total of points is 40.

Grade Table (for teacher use only)

Question	Points	Score
Problem 1	22	
Problem 2	18	
Total:	40	

1. The dataset `internet.csv` contains measurements on internet penetration and Facebook use for 25 randomly selected countries.

country (character): Country name

penetration : Internet penetration (share of the population with internet access).

facebook : Percent of the population using Facebook.

- (a) (1 point) Calculate the average share of Internet users in the sample of countries and write it down.

- (b) (1 point) Create a new variable in the data.frame `internet` called `facebookUseClass` that equals 'low' if the share of Facebook users for the country is lower than 32% and 'high' otherwise.
- (c) (1 point) Create a new variable called `lowFacebookUse` that is `TRUE` if `facebookUseClass` equals 'low' and `FALSE` otherwise.
- (d) (1 point) How many countries in the sample have a low (as defined above) share of Facebook users? Write down the answer.

- (e) (3 points) Estimate the linear regression model:

$$\text{penetration}_i = \beta_0 + \beta_1 \text{lowFacebookUse}_i + u_i$$

where u_i , $i = 1, \dots, n$ are independent random terms with zero mean and constant variance. Write down the estimated regression line.

- (f) (3 points) Interpret the estimated coefficients in the model.

- (g) (6 points) A friend of yours asserts that the average internet penetration is higher in countries with lower shares of Facebook users. Use the estimated model to test her hypothesis at a 95% significance level. Express her hypothesis in terms of the model, compute and write down the value of the test-statistic and the p-value of the test. Explain your decision to reject or to not reject the hypothesis. (Your calculations need to be reproducible!)

- (h) (6 points) The friend of yours is not sure about the theoretical foundations of your test and asks you to explain the assumptions behind the test. Due to the small sample size you know that you need to examine the normality of the data. Use a graphical check to assess the plausibility of the normality assumption. Write down

your conclusion.

2. The dataset `houses.csv` contains data on houses in the USA sold between 2006 and 2010.

SalePrice : Sales price in USD.

LotArea : Lot size in square feet.

GarageCars : Size of garage in car capacity.

- (a) (1 point) Create a new variable **Price** by dividing **SalePrice** by 1000.
(b) (2 points) Fit the linear regression model:

$$\text{Price}_i = \beta_0 + \beta_1 \text{LotArea}_i + \beta_2 \text{GarageCars}_i + u_i$$

with $i = 1, \dots, n$ and where u_i are independent random terms with zero mean and constant variance.

- (c) (2 points) Write down the estimated regression equation.

- (d) (5 points) Interpret the regression coefficients (pay attention to the scales of the variables). What is the meaning of the intercept in this model?

- (e) (2 points) Create a new variable `LotArea1` in the data.frame `houses` that equals `LotArea - 9500` and fit the model:

$$\text{Price}_i = \beta_0 + \beta_1 \text{LotArea1}_i + \beta_2 \text{GarageCars}_i + u_i$$

- (f) (6 points) A client of yours has a house for sale that has a lot size of 9500 square feet and has no garage. He would like to get an estimate of the expected sales price of her house. Give an estimate of the expected sales price together with an approximate 95% confidence interval for the expected price. Explain the meaning of the confidence interval.