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	20	
Problem 2	20	
Total:	40	

1. The dataset `trips` contains data on randomly sampled ($n = 1991$) taxi trips that took place during 2015 in New York, USA. For each trip the data contains:

   tip_amount (numeric): The tip (in USD) given to the driver.

   trip_distance (numeric): Trip distance in miles.

   total_amount (numeric): Total amount (in USD) charged to the customer for the trip (includes the initial fee, fee for distance travelled, tolls, waiting/standing time, etc.).

   (a) (2 points) Create a new variable in the dataset `trips` called `trip_distance_km` that gives the trip distance in kilometers. Hint: one mile corresponds to about 1.6 km.

   (b) (2 points) Fit the linear regression model :

   $$\text{tip\_amount}_i = \beta_0 + \beta_1 \text{trip\_distance\_km}_i + u_i \quad (1)$$

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

   (c) (3 points) Take a look at the estimated regression coefficients. How would you explain the positive association between the amount of the tip and the distance travelled (short answer)?

(d) (5 points) Estimate the expected tip for a 10 km ride and give an approximate 95% confidence interval. Explain the meaning of the confidence interval.

(e) (2 points) What is the meaning of the intercept β_0 in this model (in the context of *this* data)?

(f) (2 points) Estimate the following linear regression model:

$$\text{tip\_amount}_i = \beta_0 + \beta_1 \text{trip\_distance\_km}_i + \beta_2 \text{total\_amount}_i + u_i \tag{2}$$

with the same assumptions about u_i as in model (1).

(g) (2 points) Interpret the intercept in model (2). Is this a meaningful quantity?

(h) (2 points) Estimate the expected tip for a 10 km ride that cost a total of 20 USD.

2. The dataset `homeCredit` is a random sample of $n = 2000$ loan records from Home Credit, a consumer finance provider that lends to people with little or no credit history. Assume that each row in the data corresponds to a single person who has received a loan from Home Credit. The dataset contains the following columns:

   credit (numeric): Amount of credit received (in USD).

   ownsRealty (character): Equals `"Y"` if the person owns real estate property and `"N"` otherwise.

   (a) (2 points) Create a new variable in the dataset `homeCredit` called `noProperty` that equals `TRUE` if the person owns real estate (i.e. `ownsPropery == "Y"`) and is `FALSE` otherwise.

   (b) (2 points) Fit the linear regression model:

   $$\text{credit}_i = \beta_0 + \beta_1 \text{noProperty}_i + u_i \tag{3}$$

   with $i = 1, \ldots, n$ and where u_i are independent random terms with zero mean and constant variance. Note that the `lm` function treats logical variables as 0/1 variables: `TRUE` corresponds to 1 and `FALSE` corresponds to 0.

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

   (d) (3 points) Let μ_{property} denote the expected amount of credit received by a person who owns real estate property. Estimate μ_{property} using the estimated regression coefficients.

   (e) (3 points) Let $\mu_{\text{no-property}}$ denote the expected amount of credit received by a person without real estate property. Estimate $\mu_{\text{no-property}}$ using the estimated regression coefficients.

   (f) (2 points) Give an approximate 95% confidence interval for the *difference* between the expected tip amount on card-paid trips μ_{property} and the expected tip amount on cash-paid trips $\mu_{\text{no-property}}$ using the estimated regression coefficients.

(g) (2 points) Formulate the hypothesis:

$$H_0 : \mu_{\text{property}} \geq \mu_{\text{no-property}}$$
$$H_1 : \mu_{\text{property}} < \mu_{\text{no-property}}$$

in terms of β_1 (note that the alternative is one-sided).

(h) (3 points) Write down the t-test statistic for the hypothesis from (g) using the regression output. Compute the p-value of the test in `R` and explain your decision to reject or not to reject the null hypothesis at a 90% significance level (10% error probability).

(i) (1 point) How would you explain the difference between the average amounts of credit received by property owners and non-owners (short answer).