

Assignment 1: Who Wants to be a Billionaire?

Student Information

To receive an assignment grade, you must fill out the information in this table and include this page as your assignment cover page.

Name	Student ID Number	Tutor	Tutorial Day & Time	Tutorial Location
Sally Probability	422552	Richard Hayes	Tue 10:15am	The Spot 4452
Markus Statistics	653223			

Due Date and Weight

- **Submit via the LMS by 3pm on Friday, 14 September.**
- No late assignments will be accepted.
- This assignment is worth 10% of your final mark in ECOM20001.
- There are 30 marks in total.

What You Must Submit via the LMS

- Assignment answers, no more than 10 pages
- The .R code that generates your assignment results

Additional Instructions

- You may submit this assignment on your own, or in groups of two, to which you have been assigned in your tutorials by your tutor. Groups of people attending different tutorials are not allowed.
- You must complete the assignment in no more than 10 A4 pages with 12-point Arial, Times New Roman, Helvetica, Cambria or Calibri font, at least 1.15 points interline space, and with at least 2.54 cm margins on all sides of the page. The assignment cover page does not count as one of the 10 A4 pages.
- To save time, you may cut and paste RStudio output directly into your answers in reporting empirical results.
 - Note: Question 6 requires you to construct tables of results; submitting RStudio output alone for these questions will result in deducted marks.
- Figures may also be copied and pasted directly into your assignment answers. They may be scaled down in size to meet the 10-page limit, but please ensure your figures are readable. If they are not, marks will be deducted.
- Marks will be deducted if interpretations of results are incorrect, imprecise, unclear, or not well-scaled. Similarly, marks will be deducted if figures or tables are incorrect, unclear, not properly labeled, not well-scaled, or missing legends. Remember to always clearly label the x-axis and the y-axis of your figures, and to add legends if there is more than one line in the figure.
- In addition to your assignment answers, you must submit the .R code that generates your assignment results. This must be clearly commented and easy for the subject tutors to follow. If the code does not run through, is not well commented, or is not easy to follow, marks will be deducted.
- Students with a genuine reason for not being able to submit the assignment on time can apply for special consideration to have the assignment mark transfer to the exam at the following link:
 - <https://students.unimelb.edu.au/admin/special/>

Who Wants to be a Billionaire?

In this assignment, we explore which countries have more billionaires, and in particular whether countries open to trade have more billionaires.

Getting Started

Please create an Assignment1 folder on your computer, and then go to the LMS site for ECOM 20001 and download the following files into the Assignment1 folder:

- [Billionaires_clean.csv](#)

This dataset contains the following 8 variables:¹

- **country** --- Name of the country
- **numbil0** --- Number of billionaires in country, 0 if none in Forbes list
- **gattwto** --- Number of years that the country has been a member of the GATT or the WTO
- **open** --- Whether the country has been a member of the General Agreement of Tariffs and Trade (GATT) or the World Trade Organization (WTO) for over 30 years (dummy variable, =1 if 'Yes', 0 if 'No')
- **roflaw** --- Rule of law index
- **topint** --- Top marginal income tax rate in the country
- **gdppc** --- GDP per capita of the country, in thousands of current US Dollars
- **pop** --- Population of the country, in millions of people

In total, there is complete data for 119 countries.

¹ The reference for the dataset is Treisman, D. (2016). Russia's Billionaires. *American Economic Review*, 106(5), 236-41. The data is from selected countries in 2013.

Assignment Questions

1. Report summary statistics for the entire dataset, and separately for countries classified as closed and open to trade (as defined by the variable **open**). What does a typical country look like? How do countries closed to trade compare to countries open to trade? Report and discuss summary stats only for **numbil0**, **open**, **gattwto**, **gdppc**, **pop**. Your discussion should be no more than **four sentences** long.

(3 marks)

2. Plot and describe the densities of **numbil0** and **gattwto**. Briefly discuss your findings in no more than **two sentences**.

(3 marks)

3. Using conditional densities, explore the relationship between **numbil0** and **open** (i.e., provide densities of **numbil0** for open and closed countries in one graph). Briefly discuss your findings in no more than **two sentences**.

(2 marks)

4. Using scatter plots, provide preliminary evidence on the relationship between **numbil0** and **gattwto**. Scatter plots must include a line of best fit from a single linear regression of **numbil0** on **gattwto**. Also, provide similar graphs for open and closed countries separately (so, in total, 3 graphs). Briefly discuss your findings in no more than **four sentences**.

(4 marks)

5. Conduct two-sample t-tests for the null that the mean of **numbil0** for countries open to trade is equal to the mean of **numbil0** for closed countries. Interpret the results, and discuss the 95% confidence intervals for the difference in means in no more than **three sentences**.

(3 marks)

6. Estimate the following five multiple linear regression models where **numbil0** is the dependent variable in each regression. The list of independent variables in each respective regression is:

- **gattwto**
- **gattwto roflaw**
- **gattwto roflaw topint**
- **gattwto roflaw topint pop**
- **gattwto roflaw topint pop gdppc**

Present your results in a table that has 6 columns.

- The first column contains the independent variable names
- The second through sixth columns contain the regression results for each of the five multiple linear regression models listed above. These columns should be numbered (1) through (5) for ease of reference.
- Each regression coefficient estimate should be reported with its homoscedastic standard error below it in parentheses ().
- Put ** and * markers on the coefficient estimates to indicate statistical significance of a two-sided test of the null that the coefficient equals 0 at the 1% and 5% levels, respectively.
- At the bottom of each column the number of observations, the adjusted R-Squared, and the regression F-statistic should be reported.
- The last page of this document provides an example of the table structure you must follow. Be sure to include a table title and footnote.

Your tasks are to:

- a. In no more than **six sentences**, discuss the main results from the table, focusing on the coefficient estimate on **gattwto** and its statistical significance and magnitude. As part of your discussion, provide an example where omitted variable bias in the regression impacted the coefficient estimate on **gattwto** and the intuition behind the change in the coefficient once the extra variable was controlled for in the regression.
- b. In no more than **two sentences**, provide a clear interpretation of the coefficients of **pop**, and **gdppc** in the final column of your table.
- c. In no more than **two sentences**, interpret the adjusted R-squared from the final column in your table and the changes in adjusted R-squares from the different models in your table. Which variable seems to explain the largest share of the variation in **numbil0**?

7. Your R Code will be graded as part of the assignment.

- 5/5 if the code is as clear as the code from the tutorials
- 3/5 if anything in the code is unclear or uncommented
- 0/5 if the code is an incomprehensible mess
- If the code does not run and/or the results of the assignment cannot be replicated by your tutor, you could fail the assignment!

(5 marks)

Table Structure for Answering Question 6

FAKE DATA EXAMPLE; FOR ILLUSTRATIONS PURPOSES ONLY

Billionaires and country trade openness

	(1)	(2)	(3)	(4)	(5)
gattwto	0.159** (0.043)	0.120** (0.043)	0.115* (0.046)	0.044 (0.038)	0.030 (0.035)
roflaw		2.907** (0.933)	2.762** (1.021)	4.097** (0.836)	0.497 (1.223)
topint			0.031 (0.085)	-0.012 (0.068)	-0.073 (0.065)
pop				0.119** (0.017)	0.119** (0.015)
gdppc					0.301** (0.079)
_cons	-0.380 (1.545)	1.111 (1.552)	0.207 (2.958)	0.319 (2.363)	-1.268 (2.238)
N	93	93	93	93	93
adj. R ²	0.119	0.196	0.188	0.482	0.551

Dependent variable is numbil0, the number of billionaires in the country.

Homoscedastic standard errors in parentheses.

Statistical significance from two-sided tests of the null of no effect marked as * for 5% and ** for 1%