# ECOM20001: Econometrics 1

## Assignment 3

---

*Student Information*

**You must fill out this table and include it as a front cover page for your assignment. Only students whose name and student ID number are included on the cover page will receive marks. Groups of up to 3 students are allowed.**

| Name | Student ID Number |
|---|---|
| **Sally Probability** | 422552 |
| **Yin Statistics** | 653223 |
| **Ahmed Hypothesis** | 883190 |

*Due Date and Weight*

- **Submit via LMS by 8am on 11 May 2020**
- No late assignments will be accepted.
- This assignment is worth 10% of your final mark in ECOM20001.
- There are 50 marks in total.

*What You Must Submit via LMS*

- **Assignment answers**, no more than 9 A4 pages with 12 point font and 1 cm margins (max). 5/50 marks deducted if any of these restrictions are violated.
- The **R code** that generates your results. Specifically, copy-and-paste your R code in an Appendix at the end of your assignment document (e.g., in the .docx file) so that it can be viewed and tested by markers. The R code Appendix does not count toward your 10 page answer limit. You may alter and shrink the R code font to less than 12 point font so that it is easier to read. **2 marks will be deducted if you do not include your R code.**

*Additional Instructions*

- You may submit this assignment in groups of up to three students. Students in a group are allowed to be in different tutorials. You are also most welcome to submit your assignment not as part of a group.
- You must complete the assignment in no more than 9 A4 pages with 12 point Arial, Times New Roman, Helvetica, Cambria or Calibri font. The assignment cover page does not count as one of the 9 A4 pages, nor does the R code in the Appendix.
- To save time, you may cut and paste RStudio output directly into your answers in reporting empirical results. You are also free to create your own better-formatted tables based on your RStudio output, which is of course good practice in learning how to present empirical results.
- Figures may also may be copied and pasted directly into your assignment answers. They may be scaled down in size to meet the 8 page limit, but please ensure that 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.
- This R code in the Appendix at the end of your assignment (as discussed on the previous page) must be clearly commented and easy for the subject tutors to follow. If the code is not well commented and 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 transferred to the exam at the following link:
    - https://students.unimelb.edu.au/admin/special/

*Getting Started*

Please create an Assignment3 folder on your computer, and go to the Canvas site for ECOM 20001 and download the following data file into the Assignment3 folder:

- as3_crime.csv

It is the same dataset from Assignment 2, which recall has the following 9 variables:

- state: name of state in the United States
- year: year
- robbery_rate: number of robberies in the state per 100,000 people
- assault_rate: number of assaults in the state per 100,000 people
- burglary_rate: number of burglaries in the state per 100,000 people
- black: percentage of the population that is black/African American
- income: average household income
- age: average individual age
- female: percentage of the population that is female

In total, this dataset contains 11 years (2000-2010 inclusive) of crime and demographic data for each of the 50 American states, yielding 550 (state, year) observations in total.[1]

*About the Assignment*

Questions 1-9 (7 pages max) of the assignment continue to study what determines the level of crime in a given state, as measured in terms of the rates of robberies, assaults, and burglaries, in light of a substantial body of research in economics on large black/white disparities in the United States that identifies where they exist (e.g, crime, income, education), why they exist (e.g., discrimination, poverty traps), and what policy can do to mitigate them (affirmative action policies, anti-discrimination laws). In the assignment, we explore non-linearities in the relationship between crime and the share of the population in a state that is black, estimate elasticities, and see if the relationship evolves over time.

Question 10 (2 pages max) asks how you would go about evaluating the impact of relaxing social distancing rules on the spread of COVID19.

---

[1] The crime data comes from Doleac, Jennifer L. (2017): "The Effects of DNA Databases on Crime," *American Economic Journal: Applied Economics*, 9(1), 165-201.

*Questions*

1. **(2 marks)** Using the ggplot() command in R, produce two first scatter plots where robbery_rate is on the vertical axis, and black is on the horizontal axis. In the first scatter plot, present a quadratic regression line to visualise the relationship between the two variables. In the second scatter plot, present a cubic regression line to visualise the relationship. Does the relationship appear to be nonlinear and are there any differences, visually, of note between the quadratic and cubic regression lines?

2. **(3 marks)** Construct the following new variables in your dataset:

   - income_scale=income/10000
   - black_sq=black x black
   - black_cu=black x black x black
   - Dummy variables using the as.numeric() command in R for all years in the dataset d2000, d2001, d2002, d2003, d2004, d2005, d2006, d2007, d2008, d2009, d2010 (e.g., as you did in Assignment 2)

Using these regressors, run the following three regressions with robbery_rate being the dependent variable in each, and the independent variables being:

- Reg(1): black, black_sq, black_cu, income_scale, age, female, d2001, d2002, d2003, d2004, d2005, d2006, d2007, d2008, d2009, d2010
- Reg(2): black, black_sq, income_scale, age, female, d2001, d2002, d2003, d2004, d2005, d2006, d2007, d2008, d2009, d2010
- Reg(3): black, income_scale, age, female, d2001, d2002, d2003, d2004, d2005, d2006, d2007, d2008, d2009, d2010

Construct your table using the stargazer() command in R and for each regression report heteroskedasticity-robust standard errors[2] for each coefficient estimate, the adjusted R-squared for model fit, and the number of observations used in running the regression. Based on the regression results, report the results from an appropriate hypothesis test of your choice (using a 5% significance level) as to whether there exists a non-linear relationship between robbery_rate and black.

3. **(4 marks)** Report estimates and 95% confidence intervals of the following two partial effects from Reg (1) in question 2, holding all other regressors fixed:

   - Changing black from 0.05 to 0.10.

[2] Assume heteroskedastic errors throughout the entire assignment in conducting all hypothesis tests.

- Changing black from 0.10 to 0.15

Briefly explain why you obtain differences in these partial effects, despite the change in black being 0.05 in each case.

4. **(4 marks)** Create another collection of new variables in your dataset:
    - Logarithmic variables
        - log_robbery_rate = log(robbery_rate)
        - log_black = log(black)
    - Dummy variables for start (2000-2003), middle (2004-2007), end of sample period (2008-2010)
        - start = 1 if year<=2003 and 0 otherwise
        - middle = 1 if 2004 <= year <= 2007 and 0 otherwise
        - end = 1 if year>=2008 and 0 otherwise
    - Interactions variables
        - log_black_income = log_black x income_scale
        - log_black_start = log_black x start
        - log_black_middle = log_black x middle
        - log_black_end = log_black x end

Using these new variables and your existing ones, run the following set of regressions where the dependent variable is log_robbery_rate and the sets of independent variables are :

- Reg(1): black, income_scale, age, female, d2001, d2002, d2003, d2004, d2005, d2006, d2007, d2008, d2009, d2010
- Reg(2): log_black, income_scale, age, female, d2001, d2002, d2003, d2004, d2005, d2006, d2007, d2008, d2009, d2010
- Reg(3): log_black, log_black_middle, log_black_end, income_scale, age, female, d2001, d2002, d2003, d2004, d2005, d2006, d2007, d2008, d2009, d2010
- Reg(4): log_black, log_black_income, income_scale, age, female, d2001, d2002, d2003, d2004, d2005, d2006, d2007, d2008, d2009, d2010

Construct your table using the stargazer() command in R. For each regression report heteroskedasticity-robust standard errors, the adjusted R-squared for model fit, and

the number of observations used in running the regression. Questions 5-8 ask questions based on the regression results in this table.

5. **(2 marks)** Interpret the coefficients on black and log_black in Reg(1) and Reg(2), respectively, and comment on whether they are each individually statistically significant at the 5% level.

6. **(2 marks)** Interpret the individual coefficients on log_black_middle and log_black_end in Reg(3) and comment on whether they are each individually statistically significant at the 5% level. Offer a plain-language interpretation of what your findings imply in words.

7. **(3 marks)** Based on your estimates from Reg(3) conduct a joint-hypothesis test of the equality of the coefficients on log_black_middle and log_black_end. Report the F-statistic and p-value for the test, along with the degrees of freedom of the F-distribution for the F-statistic you report. Offer a plain-language interpretation of what your findings imply in words.

8. **(3 marks)** Based on your estimates from Reg(4) compute the elasticity of robbery_rate with respect to black for a state with income with income=$30,000 and income=$50,000. Also report the 95% CI for each elasticity. Offer a plain-language interpretation of what your findings imply in words.

9. **(2 marks)** R-code: we will review and mark your R code according to the following scheme:

    - 2/2 if R code is correct and organised and commented like the solution code for the assignment.
    - 1/2 if R code is correct, but hard to follow or not well commented.
    - 0/2 if R code is incorrect and/or a complete mess, or not submitted.

**10. (25 marks)** Evaluating the Impact of Relaxing Social Distancing Laws on the Spread of COVID19

**Background.** As of 11 April 2020, Australian states appear to be suppressing COVID19 through social distancing rules, which, among other things, include:

- Closing international and state borders
- Prohibiting social gatherings of more than 2 people if they are not family
- Closing schools, cafes, restaurants
- Restricting movements to the following activities: (1) shopping for food and essential supplies; (2) medical care; (3) exercise; (4) work and study only if you can't work or learn remotely

**Policy Question.** A pressing policy question is this: what do we do once COVID19 is suppressed? In particular to what extent are we able to relax particular aspects of social distancing rules without setting off another wave of COVID19 infections?

**What Would You Do as an Econometrician?** I want you to provide a **two-page write-up** of a recommendation to the Australian Government outlining:

- *Research design.* How you would go about relaxing a social distancing restriction across time and space in a way that would allow you to evaluate its impact on COVID19 spread after the restriction has been relaxed.
- *Data.* What data you would collect for your evaluation, being explicitly about the level of aggregation (e.g., an individual person, postcode, city/town, state?) and time frequency (e.g,. daily, weekly, monthly?) you would want to collect the data at. Also state over what sample period you would collect the data.
- *Econometric model.* The econometric model you would estimate to evaluating the impact of the spread of COVID19 after a restriction has been relaxed.

To simplify the analysis, focus on relaxing a single restriction and evaluating its impact, holding all other restrictions fixed.[3] You choose the restriction to relax.

**What We Will Do.** After receiving the classes' two-page write-ups, A/Prof Byrne and the tutors will consolidate them into a ECOM20001 report and submit it on behalf of the class to the Prime Minster and Treasurer of Australia, Hon Scott Morrison MP and Hon Josh Frydenberg MP.

[3] An example of a relaxation of a restriction might be, for example, re-opening cafes and restaurants and allowing groups of up to 10 to be together. Another example could be re-opening schools for all children.