

# Final Coursework

## Quantitative Data Analysis (POLS0083)

### Instructions

- The final assessment will be posted on Moodle on **17th December 2021 at 6pm**, and is due on **10th January 2022 at 2pm**. Please follow all designated SPP submission guidelines for online submission as detailed on the POLS0083 Moodle page. Standard late submission penalties apply.
- This is an assessed piece of coursework (worth 70% of your final module mark) for the POLS0083 module; **collaboration and/or discussion of the coursework with anyone is strictly prohibited**. The rules for plagiarism apply and any cases of suspected plagiarism of published work or the work of classmates will be taken seriously.
- As this is an assessed piece of work, you may not email/ask the course teaching team questions about the coursework.
- Along with the coursework questions, the necessary data sets for the coursework can be found on the POLS0083 page on Moodle.
- The coursework should be submitted via the ‘POLS0083 - 2,000 word Final (70%)’ link on the course Moodle page. You will need to click the ‘Submit Paper’ link at the bottom of the page. When presented with the ‘Submit Paper’ box, **the ‘Submission Title’ should be your candidate number**, and you should upload your document into the box provided. Please remember to state **ONLY** your candidate number on your coursework (your candidate number is made up of four letters and one number e.g. ABCD5). Your name and/or student number **MUST NOT** appear on your submission.
- The coursework consists of three separate sections, each with several questions. The marks allocated for each section are indicated in the text. You must complete each question to achieve full marks.
- Unless otherwise stated, answers should be written in complete sentences. Be sure to answer all parts of the questions posed and provide a substantive interpretation of the results.
- The word count for this assessment is 2,000 words. This does not include the code, your output, or any words (or numbers) contained within tables or figures.
- Please submit your type-written (numbered) answers in a single document (a word file or a pdf file). You can, but do not have to, use RMarkdown to produce such a file. If you use RMarkdown, you can integrate the code with the answers (make sure that it is completely visible). If you do not use RMarkdown, you can create an appendix section at the end which contains all the R code needed to reproduce your results. Your code has to work when we run it. You do not need to include the code that failed to run, but just the well-annotated, cleaned-up version. If you do not provide the code to a question, any written answer to that question will be disregarded.
- You may assume the methods you have used (e.g. difference in means, linear regression, etc) are understood by the reader and do not need definitions, but you do need to be able to explain what they do and how they apply to answering the question.
- Round all numbers to two digits after the decimal point.
- Do not screenshot or copy and paste *any* brute R output (e.g. `lm(y ~ x)`) into your answers. Create a formatted table that is easy to read.
- Assign every table and figure a title and a number and refer to the number in the text when discussing a specific figure or table.

# Section 1

## Public perceptions of benefit fraud

Trust in social welfare institutions relies on how the public perceives of the deservingness of recipients of such benefits. In particular, notions of so-called ‘benefit cheats’ erodes public confidence in social welfare.

Past surveys have shown that the public tend to greatly over-estimate the financial scale of benefit fraud. For example, a 2013 [Ipsos MORI study](#) found that on average, respondents guessed that £24 out of every £100 in benefit claims is done fraudulently, whereas the official estimate is around £0.70 of every £100.

In this section, we explore public perceptions about [fraudulently claiming benefits](#), and in particular, about how widespread false benefits claims are. More specifically, we will examine whether certain socio-demographic characteristics are associated with public perceptions about how widely false benefit claims are made in the UK.

We will use part of the British Social Attitudes (BSA) data set on poverty and wealth, which you can download as `bsa-poverty.csv` from the POLS0083 Moodle page. The data set contains the following variables:

Variable name	Description
<code>NatFrEst</code>	Answer to the question “Out of every 100 people receiving benefits in Britain, how many have broken the law by giving false information to support their claim?”
<code>leftrigh</code>	Five-point left-right ideological scale, with 0 to the left and 4 to the right
<code>RSex</code>	Sex of respondent, 1 if male or 0 if female
<code>HEdQual3</code>	Completed university degree, 1 if the respondent completed degree or 0 if not

You can load the data set by using the following command:

```
bsa <- read.csv("data/bsa-poverty.csv")
```

### Questions (42 Marks)

1. How many individuals are included in the data set?
2. Is the left-right ideology similar on average for those who did and not answer the item `NatFrEst`?
3. Calculate the median of the variable `NatFrEst`. What does this tell us about the distribution of perceptions about fraudulent benefits claims?
4. Create a histogram for `NatFrEst` and interpret it. What does this tell us about public perceptions about fraudulent benefits claims?
5. We are interested in seeing whether there is a relationship between a person’s left-right orientation and how widespread they think fraudulent benefit claims are. Fit the relevant simple linear regression model and interpret the substantive significance of the estimated slope coefficient. You do not need to discuss statistical significance.
6. State a null and an alternative hypothesis for the estimated slope coefficient, decide whether to reject the null hypothesis, and provide a conclusion.
7. How is your conclusion in Question 6 related to Type I and Type II error?
8. We now add whether the respondent completed a university degree and respondent sex to our analysis. Interpret the estimated coefficient for the left-right orientation. Does your answer to the question about the relationship between a person’s left-right orientation and `NatFrEst` change? If so, how?
9. Calculate and interpret the 99% confidence interval for the estimated coefficient for completing a university degree from the model you fitted in Question 8. What does the estimated standard error tell

us?

10. Finally, we are interested in whether the relationship between left-right orientation and perceptions of how widespread benefit fraud is also depends on an individual's education level.
  - a. Add the relevant interaction term to your regression model and display your results.
  - b. Assess the model fit for this regression model.
  - c. Interpret the estimated interaction term and discuss its substantive and statistical significance.
  - d. Interpret the intercept and its statistical significance. Is the intercept meaningful in this regression model?
  - e. Visually represent the results from the regression model using four lines (one each for male respondents with a university degree, female respondents with a university degree, male respondents without a university degree and female respondents without a university degree) and describe what your graph shows.

## Section 2

### Dystopian fiction and willingness to justify radical political action

We now look at part of the [recent study by Jones and Paris \(2018\)](#), who conducted a series of survey experiments to examine whether exposure to dystopian fiction affects an individual's political beliefs. They define dystopian fiction as “portray[ing] a dark and disturbing world dominated by an overwhelmingly powerful... controlling entity that acts to undermine core values”. In particular, the authors hypothesised that exposure to dystopian fiction would lead to higher support for more radical political action, especially for more violent action.

In the first study, the authors randomly assigned the US-based respondents into two groups. Respondents assigned to the treatment group first read an excerpt from the first book in *Hunger Games*, and then watched a 17-minute video with various violent scenes from the *Hunger Games* films. Respondents assigned to the control group were not exposed to any media.

The premise of *Hunger Games* is that an overwhelmingly powerful government forces individuals to take part in a contest where the participants are forced to kill each other until only one survivor remains.

The outcome variables are a series of attitudinal items about willingness to justify different disruptive activities: civil disobedience, damaging government property, cyberattacks on government websites, armed rebellion, and violent protest.

The data file you will use, which can be downloaded on the POLS0083 Moodle page, is titled `dystopia.csv`. The data includes the following variables:

Variable name	Description
hgindic	Treatment group, with 1 for those exposed to <i>Hunger Games</i> and 0 for the control group
j_disobed	How much civil disobedience can be justified, on a 0-1 scale
j_damage	How much damaging government property can be justified, on a 0-1 scale
j_cyber	How much cyberattacks on government websites can be justified, on a 0-1 scale
j_rebel	How much armed rebellion can be justified, on a 0-1 scale
j_violent	How much violent protest can be justified, on a 0-1 scale
female	Respondent sex, with 1 as female and 0 as male
ideo	Respondent left-right ideological orientation, with higher scores indicating the right, on a 1-6 scale

You can load the data set by using the following command:

```
dystopia <- read.csv("data/dystopia.csv")
```

### Questions (33 Marks)

1. What are the proportions of female respondents within the treatment and control groups?
2. We are interested in the outcome variables for civil disobedience and armed rebellion. For each of these outcome variables:
  - a. Create a boxplot for each treatment group and interpret the results.
  - b. Calculate difference in means by treatment group and interpret the results. You do not need to discuss statistical significance.
  - c. Conduct the appropriate hypothesis test and interpret the results for the difference in means.

3. In the current study, in which two ways can we understand the population from which the sample is drawn?
4. With what assumptions can we interpret the results from Question 2 as causal?
5. Fit a regression model for each outcome variable with the treatment group, sex, and ideology. Does the estimated difference in means change? What does this tell us about the randomisation of treatment in this study?
6. Do your results support the authors' hypothesis about the link between dystopian fiction and willingness to support radical political action? State your conclusions in terms of substantial and statistical significance.

## Section 3

### Direct democracy and citizens' local support

Does direct democracy (such as local referendums) increase satisfaction with local policies and governmental institutions? Marien and Kern (2018) explore this question in their article [“The Winner Takes It All: Revisiting the Effect of Direct Democracy on Citizens' Political Support”](#).

Drawing on the existing literature, the authors argue that direct democratic instruments have short-term positive effects on satisfaction with local policies and institutions, since they increase the perceived fairness of decision-making and ability to influence political outcomes among the electorate.

To test their hypothesis, the authors conducted multiple surveys during the spring of 2015, when the Belgian city Mechelen conducted a referendum on traffic circulation, after citizens had objected to the local government's initial plans. The researchers collected survey data before and after the referendum took place both from a sample of local residents in Mechelen who lived in neighbourhoods affected by the traffic diversion plans (treatment group) and from a sample of residents of a similar neighbourhood in Mechelen not affected by the traffic circulation plans (control group).

The data file you will use, which can be downloaded on the POLS0083 Moodle page, is titled `referendum.csv`. The data includes the following variables:

Variable name	Description
<code>idresp</code>	Respondent ID
<code>neighb</code>	Neighbourhood, 1 as the treatment group and 0 as the control
<code>w1_byear</code>	Respondent birth year
<code>primary</code>	Highest level of education completed - primary - with 1 Yes and 0 No
<code>secondary</code>	Highest level of education completed - secondary - with 1 Yes and 0 No
<code>tertiary</code>	Highest level of education completed - tertiary - with 1 Yes and 0 No
<code>w1_trust_general</code>	General level of trust on a 0-10 scale, with higher scores meaning more trust, before the referendum
<code>w1_pol_interest</code>	Political interest on a 0-10 scale, with higher scores meaning more interest, before the referendum
<code>w1_democracy_satis</code>	Democratic satisfaction on a 0-10 scale, before the referendum, with higher scores meaning higher satisfaction
<code>w2_democracy_satis</code>	Democratic satisfaction on a 0-10 scale, after the referendum, with higher scores meaning higher satisfaction

You can load the data set by using the following command:

```
referendum <- read.csv("data/referendum.csv")
```

### Questions (25 Marks)

1. Is the average level of democratic satisfaction before the referendum significantly different (statistically and substantively) between respondents in the treatment and control groups?
2. Provide some evidence that the treatment and control neighbourhoods are similar in terms of respondent characteristics before the referendum took place.
3. Focusing on the respondents living in a neighbourhood affected by the government's plan on traffic circulation, use a before-after design to estimate the average treatment effect. Is this difference significant both statistically and substantively?
4. With what assumptions can we consider the results of Question 3 as causal?

5. Calculate and interpret the difference-in-differences for democratic satisfaction for the two neighbourhoods and interpret your results.
6. Do the results in Question 5 support the authors' hypothesis about the link between direct democracy and citizens' perceptions about democracy? Why or why not?
7. With what assumptions can we consider the results of Question 6 as causal?