

# Midterm Coursework

## Introduction to Quantitative Research Methods (PUBL0055)

### Instructions

- The coursework will be posted on Moodle on **Friday 6th November 2020 at 2pm**, and is due on **Wednesday 11th November 2020 at 2pm**. Please follow all designated SPP submission guidelines for online submission as detailed on the PUBL0055 Moodle page. Standard late submission penalties apply.
- This is an assessed piece of coursework (worth 25% of your final module mark) for the PUBL0055 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 tutors or teaching fellows questions about the coursework.
- Along with the coursework itself, the datasets for the coursework can be found in the PUBL0055 page on Moodle.
- Coursework should be submitted via the ‘Turnitin Assessment - PUBL0055 - Midterm Assessment’ 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 coursework.
- The coursework consists of five questions; you must complete each part of each question to achieve full marks. Each question is worth 20 marks in total.
- Unless otherwise stated, answers should be written in complete sentences. Be sure to answer all parts of the questions posed and interpret the results.
- The word count for this assessment is 1000 words. This does not include the appendix.
- Please submit your type-written (numbered) answers in a single document. Create an appendix section at the end which contains all the R code needed to reproduce your results (you do not need to include the code that failed to run, but just the cleaned-up version. Your code has to work when we run it).
- 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 explain how they apply to answering the question.
- Round all numbers to two digits after the decimal point.
- Do not copy and paste *any* R output (e.g. the output from running `lm(y ~ x)`) into your answers. Create a formatted table that is easy to read.
- All variable names in the coursework are written in **this\_font**.

# Female Leadership and Public Health Outcomes During COVID-19

Some countries have been more successful than others in terms of public health outcomes during the COVID-19 pandemic. In particular, deaths associated with the virus have been unevenly distributed cross-nationally. Of the many factors that might be responsible for these differences, a great deal of media attention has focused on the idea that countries led by female politicians may have been more successful at dealing with COVID-19 than countries led by male politicians. For instance, a recent article in the Guardian newspaper asked “Are female leaders more successful at managing the coronavirus crisis?” The New York Times went further, asking “Why are Women-Led Nations Doing Better With COVID-19?” Reporting of this sort relates to ongoing debates in the academic literature on differences in the efficacy of male and female political leaders.

In this section, you will investigate these ideas empirically by examining the relationship between female political leadership and COVID-19 death-rates in a sample of countries. This exercise is loosely based on papers by [Purkayastha et al](#) and [Coscieme et al](#), both of which use data on COVID-19 fatalities to estimate the effects of female leadership on public health outcomes.

The data file you will use, which can be downloaded on the PUBL0055 Moodle page, is titled `covid_country_data.csv` and contains information from 180 countries. The data includes the following variables:

Variable name	Description
<code>countryiso3</code>	Unique country identifier
<code>country</code>	Country name
<code>deaths_per_100k</code>	Total number of deaths recorded from COVID-19 per 100,000 people in the population
<code>female_leader</code>	TRUE if the country has a female leader and FALSE otherwise
<code>ghs_index</code>	The Global Health Security index score, a 2019 measure that aimed to predict how prepared each country was for handling epidemics and pandemics
<code>gdp_percap</code>	GDP per capita in current US dollars, measured in 2016
<code>pct_urban</code>	The percentage of the population living in urban agglomerations of more than 1 million, measured in 2016
<code>health_gdp_pct</code>	Current health expenditure as a percentage of GDP, measured in 2016
<code>democracy</code>	TRUE if the country is a democracy and FALSE otherwise (Polity IV, dichotomised at 6)
<code>continent</code>	The continent in which the country is located

You can load the data by using the following command:

```
covid <- read.csv("data/covid_country_data.csv")
```

## Question 1 (20 marks)

- Begin your analysis by providing appropriate descriptive statistics on the two main variables of interest for this analysis, `female_leader` and `deaths_per_100k`. Present summaries of both variables, either in graphical or tabular form. Interpret your results.
- Calculate the difference in mean deaths for countries with and without female leaders. Interpret this difference in means in substantive terms. Is this the causal effect of female leadership on public health outcomes? Why or why not?

## Question 2 (20 marks)

- Estimate two multiple regression models with `deaths_per_100k` as the dependent variable. For the first model, include `female_leader` as the only explanatory variable. For the second model, include `female_leader` and three other variables of your choice. Do not include `ghs_index`. If you decide to include

`gdp_percap`, use a log transformation by including `log(gdp_percap)` in the model formula. Interpret your results, making sure to compare and contrast the coefficient on female leadership in the two models.

**b.** How does controlling for these other variables affect your answer on causality from question 1.b? Describe two additional variables that are not included here that you might also want to control for to strengthen the evidence for a causal effect.

### Question 3 (20 marks)

In 2019, before the pandemic, researchers constructed an index which was designed to measure how prepared different countries are for global health emergencies. The scores for each country are stored in the variable `ghs_index`.

**a.** Present descriptive statistics relating to the `ghs_index` variable and produce a plot to show the relationship between that variable and `deaths_per_100k`. Interpret your plot.

**b.** Use multiple regression to assess whether/how `ghs_index` is predictive of COVID-19 deaths. Include the same variables that you included in Question 2 plus `ghs_index`. Interpret your results.

# Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

## Children's Television and Educational Performance



Figure 1: Sesame Street

Can educational television programmes improve children's learning outcomes? Sesame Street is a long-running American television programme aimed at young children. The creators of Sesame Street decided from the very beginning of the show's production that a central goal would be to educate as well as entertain its audience. In addition to building the show around a carefully constructed educational curriculum, the producers also worked closely with educational researchers to determine whether the show's content was effectively improving its young viewers' numeracy and literacy skills.

The dataset contained in `sesame_experiment.csv` includes information on 240 children who were randomly assigned to two groups. The treatment of interest here is not *watching* Sesame Street, because it was not possible to force children to watch or to refrain from watching a freely available TV show. Instead, researchers randomized whether children were *encouraged* to watch the show. The parents of children in the treatment group were encouraged to show Sesame Street to their children on a regular basis, while parents of the children in the control group were given no such encouragement.

In this section, you will analyse data from this experiment. The data file you will use, which can be downloaded on the PUBL0055 Moodle page, is titled `sesame_experiment.csv` and contains information from 240 children who participated in the experiment. The data includes the following variables:

Variable name	Description
<code>encouraged</code>	TRUE if the child was encouraged to watch Sesame Street, FALSE otherwise
<code>watched</code>	TRUE if the child watched Sesame Street, FALSE otherwise
<code>letters</code>	The score of the child on a literacy test (from 0 to 100)
<code>age</code>	Age of the child (in months)
<code>female</code>	TRUE if the child is female, FALSE otherwise

You can load the data by using the following command:

```
sesame <- read.csv("data/sesame_experiment.csv")
```

### Question 4 (20 marks)

a. Use the data from the experiment to calculate the following quantities:

1. The proportion of children who were encouraged to watch Sesame Street.
2. The proportion of children who watched Sesame Street.

3. The proportion of children who watched Sesame Street among those who were encouraged to watch.
4. The proportion of children who watched Sesame Street among those who were not encouraged to watch.

What do these figures tell you about the effectiveness of the encouragement?

- b.** Calculate the difference in mean literacy scores between children who were encouraged to watch Sesame Street and those who were not. Interpret your results. Does the difference in means estimate the causal effect of encouraging parents to have their children watch Sesame Street? Why, or why not?
- c.** Estimate two regression models, both of which should have **letters** as the outcome variable. In the first model, include **encouraged** as the only explanatory variable. In the second model, include **encouraged**, **female**, and **age**. Interpret the results, making sure that you compare and contrast the results of your models. Explain any differences or similarities that you observe.

### Question 5 (20 marks)

- a.** Adapt the second model from question 4b to estimate a regression model that allows you to determine whether the effect of the encouragement depends on the gender of the child. Interpret your results.
- b.** Discuss the strengths and weaknesses of this experiment for answering the research question posed at the beginning of this section (“Can educational television programmes improve children’s learning outcomes?”). What alternative research designs might be used to improve our understanding of the effects of educational television on child literacy?

# Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder