```
---
title: "Problem Set 2"
name: "Clare Porter"
subtitle: "Due date: 23 September"
format:
  html:
    self-contained: true
toc: true
editor: visual
execute:
  echo: true
  warning: false
---
```

Please upload your completed assignment to the ELMs course site (under the assignments menu) and to your class Github repository. You need to upload both your Quarto document (with R code, as needed) and its rendered output (either as a PDF or HTML file).

Total points: 25

Question 1

Points: 10

Suppose I want to examine the hypothesis that consuming political news coverage from cable network channels (e.g., CNN, MSNBC, Fox News, etc.) causes citizens to have less trust in government than if they were to consume it from mainstream print media (e.g., New York Times, Wall Street Journal, etc.). What is more, suppose that I believe that this effect is greatest among "young" citizens. Design the contours of an experiment to test this proposition (and be sure to sketch the key elements of the experimental design). Second, briefly describe any significant shortcomings that the experimental design might have.

In order to understand how political news coverage (independent variable) and age (independent/interaction variable) impacts trust in government (dependent variable), we would randomly sample 5,000 citizens across the United States with varying demographics to measure trust before and after watching or reading certain media.

The sample should include citizens across a range of political beliefs, states, and demographics, including age. Such metadata will be collected for each participant. At the beginning of the study, participants will also be asked to rate their trust in government on a scale of 1 to 5 in order to establish a baseline. Participants will then be randomly assigned a news outlet to watch or read for 30 minutes each day: MSNBC, Fox News, New York Times, Wall Street Journal, or no news. Participants which were assigned a media outlet to consume are the treatment group, and participants who were not assigned a media subscription are the control group. Participants will be surveyed for their trust in the government on a scale of 1 to 5 each month for the next year.

This experiment will elicit a data frame with 12 columns ranking trust in government at the beginning of the study and end of each month (continuous dependent variables), plus age (categorical independent variable: 18–29, 30–44, 45–59, 60 years or older), political affiliation (categorical independent variable: democrat, republican, independent, other), and participant (FEMA) regions of residence (categorical independent variable: 1–10); and 5,000 rows representing the associated data for each individual in the study.

Following the data collection, we can compare the level of trust in governments that participants held after consuming political news coverage from cable network channels with those assigned to consume political news from mainstream print media and those who did not consume any political news. We can find on average the trust in government to have increased or decreased with time depending on media type (or no media) and how consistently trust levels were changed across groups. By comparing the mean, median, and variation of trust levels across media types, we can accept or reject the alternative hypothesis that cable network channels more negatively impact trust in government than mainstream print media.

In order to understand how age impacts trust in governments, we can next compare difference in trust levels across age groups. By comparing the mean, median, and variation of age groups, we can accept or reject the alternative hypothesis that cable network channels more negatively impact trust in government amongst younger people (defined in this study as ages 18–29) compared to more seniored participants.

While this study may enable us to draw some general conclusions about the impact of media on trust in government, it is not without shortcomings. Some shortcomings are due to factors beyond our control. For example, ideally in a parallel universe, participants would first consume one media type, and then go back in time to consume another, thereby creating a perfect comparison. Also, in an ideal study, political beliefs, political changes happening in real life, and other external forces would not hold sway to a participant's trust in government over the course of the study. Of course, there are also shortcomings in this experiment which could be overcome by a different study design. For example, participants who share political beliefs with the party in power at the time of the study may see smaller variability in their trust in government than those who do not share the same political beliefs with the party in power. A different study may consider measuring trust in government for the sample across multiple administrations in order to test the impact of the political party in power. Similarly, a different study may consider different time steps to measure trust (e.g., on a daily or annual basis), in order to understand volatility of trust with time. This new study could even begin by measuring trust in government for some amount of time before administering the treatment (assignment of media type to consume) in order to better understand how participant's typical engagement with media and stability of their trust differs from the level and stability in their trust following the administration of the treatment.

Question 2

Points: 10

Below are four hypothetical research designs to determine the effects of democratic governance on the funding of educational institutions. Briefly evaluate each sampling design in terms of its: (1) sampling variability; and, (2) sources of potential bias.

A. sample of educational funding levels for every country in Europe for each of the last 100 years.

–   This study will elicit **low to moderate sampling variability** because of the large sample size (of every country in Europe for the last 100 years). The variability may be moderate however given the range in data over a 100-year time horizon.

–   **Assuming the population for the study was Europe,** the study would have **low bias** in which the sample well resembled the population.But **if the population for the study was the World,** this study would have **high bias** which over-represented Europe.

B. sample of educational funding levels in five randomly picked countries drawn from the global population of countries for each of the last 10 years.

–   This study will produce **high sampling variability** because repeated sampling of 5 countries across the entire world could produce a large range of results.

–   Five countries is also insufficient to generalize across the world, so this study would be **moderate to highly biased.**

C, survey of educational administrators that yields responses that allow for a comparison of educational funding levels in Russia and the United States for the six-year period from 2010 to 2015.

–   Repeated surveying across all or a large number of administrators (assuming sample size is representative) in Russia and the US from 2010 to 2015 would produce **low variability**.

–   This study could be **highly biased** towards educational administrators, over–

representing one perspective of many on educational funding levels.

D. sample of 75 randomly selected countries' educational funding levels for each of the past 25 years.

-    Given the large number of countries selected in each sample, this study would produce **low sampling variability.**

-    A range of countries should be well-represented in this sample, resulting in **low bias.**

Question 3

Points: 5

The COVID-19 pandemic wreaked havoc on the global economy. People became sick or died, supply chains snapped, and we all sheltered behind strict national borders or in our homes. We can see the effect of these changes when we look at how the growth of countries' Gross Domestic Product (GDP) changed between 2019 and the global pandemic's start in 2020.

::: callout-note
GDP is a popular measure of a country's annual economic output. Its annual growth tends to be approximately Normally distributed.
:::

```{r}
library(tidyverse)
library(wbstats)

gdp_growth_df <- wb_data("NY.GDP.MKTP.KD.ZG",
                         start_date = 2019,
                         end_date = 2020,
                         return_wide = F)

ggplot(gdp_growth_df, aes(x = value, fill = factor(date))) +
  geom_density(alpha = 0.75) +
  theme_minimal() +
  theme(plot.title = element_text(face = "bold"),
        plot.title.position = "plot") +
  labs(title = "Countries' annual percentage growth rate of GDP in 2019 and 2020",
       x = "Annual growth in GDP (%)",
       y = "Density",
       fill = "Year",
       caption = "Source: World Bank")
```

Part A

Points: 2

Describe the distribution of all countries' annual GDP growth rates in 2019 (found in the `gdp_growth_df` dataframe). Do the same for 2020. Include in those descriptions their five-number summaries.

Using the code below, I find the following five-number summary of 2019 growth rates:

-    Min: -11.44

-    P25: 1.17

-    Median: 2.82

-    P75: 4.92

    –   Max: 23.41

The distribution of all countries' annual GDP growth rates in 2019 is unimodal and very slightly right skewed (Mean \> Median), and clustered around a tall peak.

Using the code below, I find the following five-number summary of 2020 growth rates:

–   Min: -54.34

–   P25: -8.2

–   Median: -3.81

–   P75: -1.17

–   Max: 43.48

The distribution of all countries' annual GDP growth rates in 2020 is unimodal but very slightly left skewed (Mean \< Median), with a larger spread and shorter peak.

```{r}

gdp_growth_df %>%
  group_by(date) %>%
  skim()
```

Part B

Points: 2

Calculate the mean and standard deviation of all countries' annual GDP growth rates in 2019. Do the same for 2020.

Using the code in Part A, I find the following statistics of 2019 growth rates:

–   Mean: 2.9

–   SD: 3.54

Using the code in Part A, I find the following statistics of 2020 growth rates:

–   Mean: -5.11

–   SD: 8.69

Part C

Points: 1

Compare the mean and standard deviation of GDP growth rates in 2019 with those in 2020. What, generally, happened to each country's economy? Do not use the words or phrases "mean" or "standard deviation" in your answer.

In 2019, most counties experienced some positive economic growth. Many countries' economic growth weas clustered around moderate increases, and the differences between country growth levels was relatively moderate. In 2020, however, economic growth in most countries suffered downturns. Variation and spread in economic outcomes was much larger across the world, with some countries still seeing growth while others' economies shrank dramatically.