PS531 Pre-Analysis Plan Code Appendix

Zach Salazar

Loading in the LEAD dataset and selecting needed variables, followed by subsetting data to only include US Presidents.

```r
# Set working directory
setwd("/Users/zach/Documents/GitHub/531-pre-analysis-plan")

# Load packages
library(tidyverse)

# Load in LEAD dataset
lead <- load("leaders_datapaper_replication_final_9_10_15.RData")

# Rename dataset
lead <- x

# Choose relevant variables
my_vars1 <- c("idacr", "leadername",
              "startdate", "enddate",
              "milservice", "combat",
              "polity")

# Select only relevant variables
lead <- lead %>%
  dplyr::select(all_of(c(my_vars1)))

# Filter leaders to only US Presidents
lead_USA <- lead %>%
  filter(idacr == "USA")
```

Loading in the MIDs dataset and selecting relevant variables, followed by subsetting for MIDs only involving the US.

```r
# Load in MIDs dataset
mids <- read.csv("dyadic_mid_4.02.csv")

# Choose relevant variables
my_vars2 <- c("namea", "strtmnth", "strtday",
              "strtyr", "endmnth", "endday",
              "endyear", "rolea", "outcome",
              "cumdurat", "fatlev","fatleva",
              "revstata","hihost", "highmcaa",
              "war", "severity")

# Select only relevant variables
mids <- mids %>%
  dplyr::select(all_of(c(my_vars2)))

# Filter MIDs to only those involving US
mids_USA <- mids %>%
  filter(namea == "USA")
```

Loading in the National Material Capabilities dataset and filtering for US CINC scores.

```r
# Load in NMC dataset
cinc <- read.csv("NMC-60-abridged.csv")

# Choose relevant variables
my_vars3 <- c("stateabb", "year", "cinc")

# Select only relevant variables
cinc <- cinc %>%
  dplyr::select(all_of(c(my_vars3)))

# Filter for US CINC scores
cinc_US <- cinc %>%
  filter(stateabb == "USA")
```

Stratification

```r
# Load packages
library(dplyr)

# Generate fake data
set.seed(123)
n <- 100

# Stratum variable
stratum <- sample(c("pre-1898", "1898-1945", "post-WWII"), n, replace = TRUE)
# Outcome variable
outcome <- rnorm(n, mean = 10, sd = 2)

# Create a data frame
data <- data.frame(stratum, outcome)

# Stratify the data
stratified_data <- data %>%
  group_by(stratum)

# Perform analysis within each stratum
stratified_data <- stratified_data %>%
  summarize(mean_outcome = mean(outcome))

# Results
print(stratified_data)
```
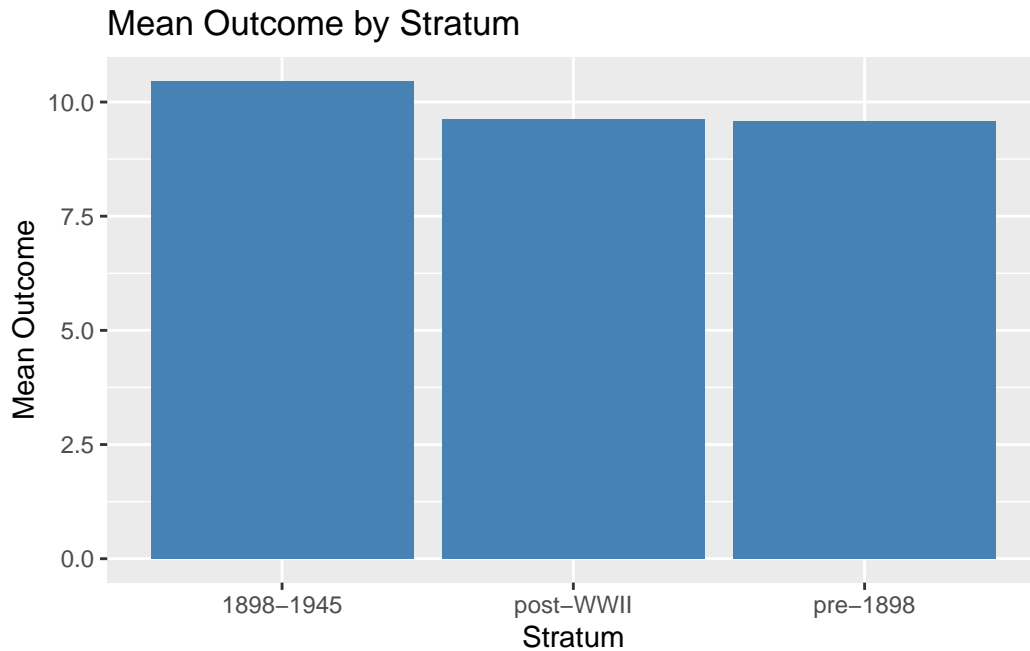
```
# A tibble: 3 x 2
  stratum    mean_outcome
  <chr>             <dbl>
1 1898-1945          10.5
2 post-WWII          9.62
3 pre-1898           9.59
```

```r
# Plot of results
ggplot(stratified_data, aes(x = stratum, y = mean_outcome)) +
  geom_bar(stat = "identity", fill = "steelblue") +
  labs(x = "Stratum", y = "Mean Outcome") +
  ggtitle("Mean Outcome by Stratum")
```

Mean Outcome by Stratum

Linear regression plots for MID initiations and outcomes

```r
# Load packages
library(ggplot2)

# Set working directory
setwd("~/Documents/GitHub/531-pre-analysis-plan")

# Load in US presidents dataset
pres <- read.csv("us_presidents.csv")

# Linear regression plot for total MID initiations
ggplot(pres, aes(x = rank, y = initiated)) +
      labs(title = "Total Initiations",
           x = "Rank", y = "Initiations") +
      geom_point() + stat_smooth(method = "lm",
                                 formula = y ~ x,
                                 geom = "smooth")
```

Total Initiations
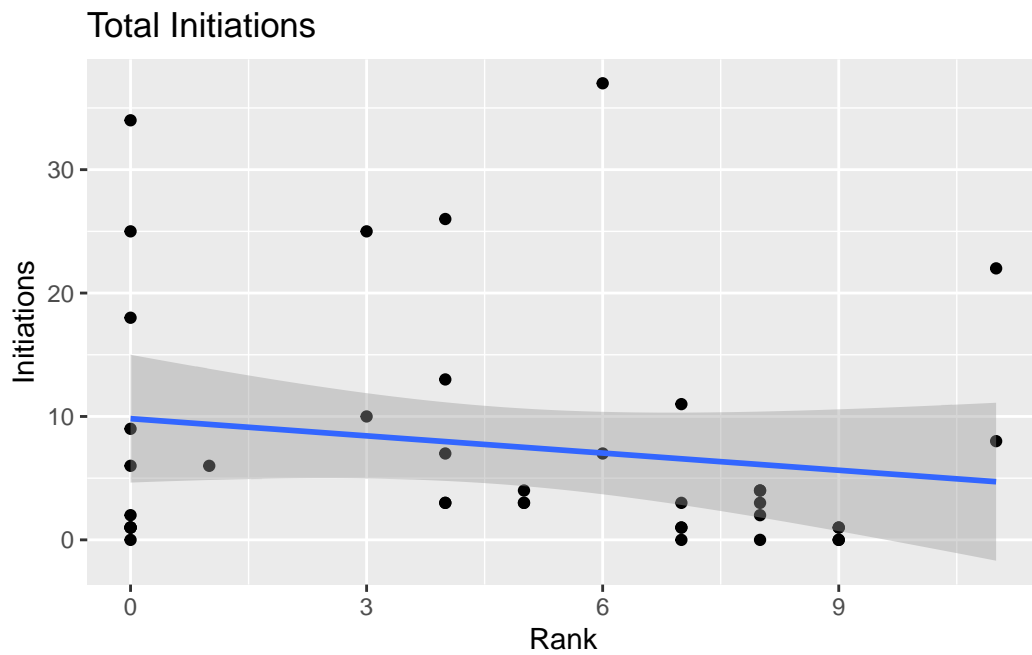

```r
# Linear regression plot for MID outcomes
ggplot(pres, aes(x = rank, y = pctwon)) +
      labs(title = "MID Wins",
           x = "Rank", y = "Wins") +
      geom_point() + stat_smooth(method = "lm",
                                 formula = y ~ x,
                                 geom = "smooth")
```

MID Wins
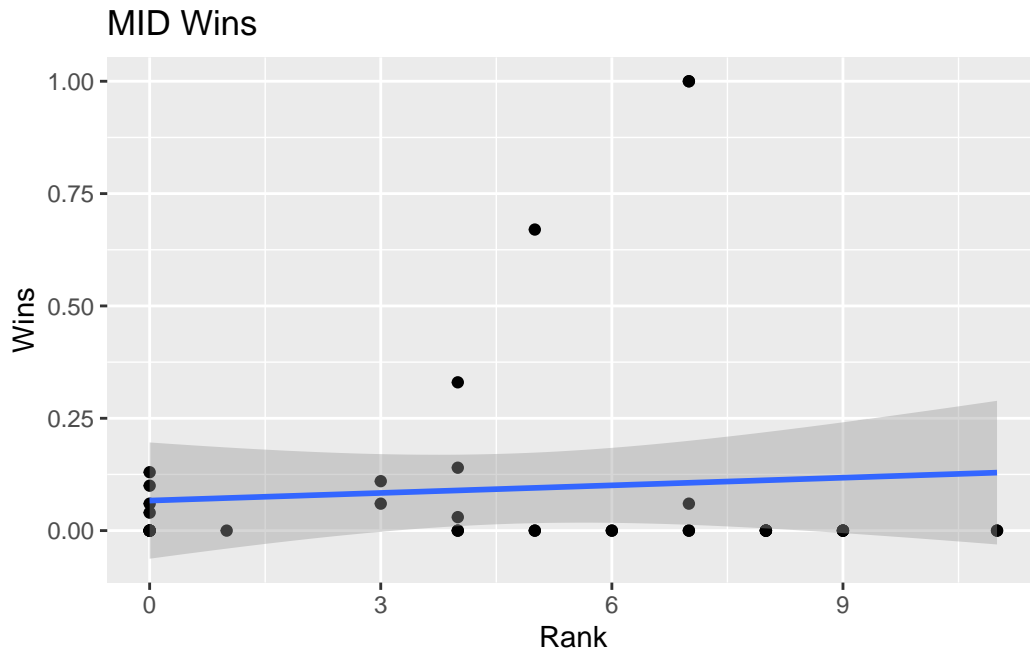
Difference-in-Differences, bias, and MSE

```r
# Load packages
library(dplyr)
library(ggplot2)

# Set seed
set.seed(123)

# Create a dataset with pre and post-treatment periods
pre_period <- data.frame(
  president = rep(c("Military", "Non-Military"), each = 5),
  time = rep("Pre", 10),
  foreign_policy = rnorm(10, mean = 5, sd = 1))

post_period <- data.frame(
  president = rep(c("Military", "Non-Military"), each = 5),
  time = rep("Post", 10),
  foreign_policy = rnorm(10, mean = 8, sd = 1))

# Combine pre and post periods
data <- rbind(pre_period, post_period)
```
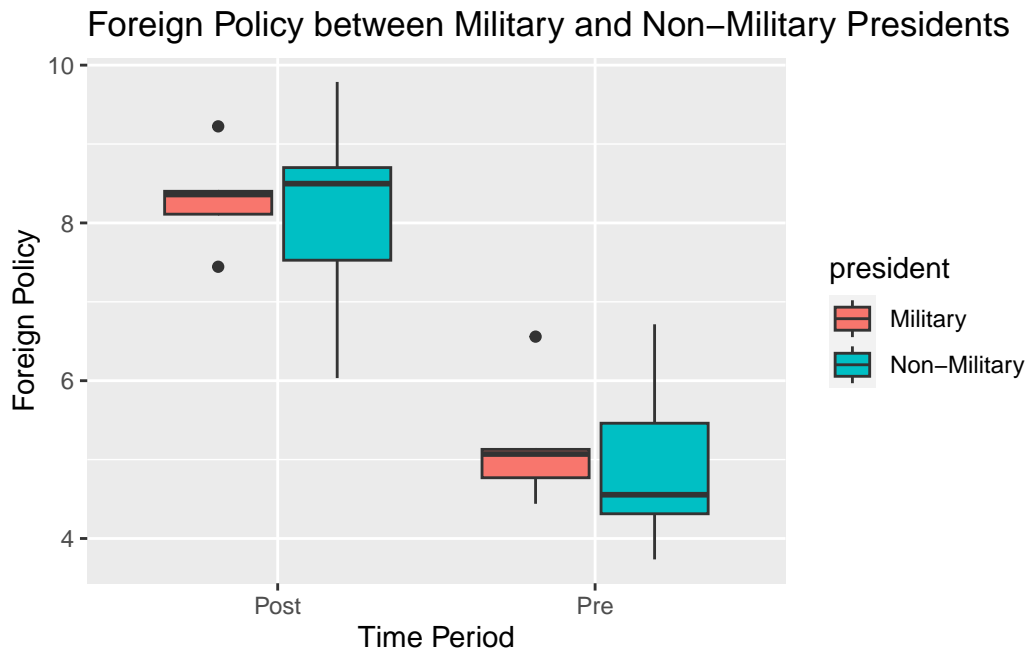
```
# Visualize the data
ggplot(data, aes(x = time, y = foreign_policy, fill = president)) +
  geom_boxplot() +
  labs(x = "Time Period", y = "Foreign Policy") +
  ggtitle("Foreign Policy between Military and Non-Military Presidents")
```

Foreign Policy between Military and Non−Military Presidents


```
# Difference-in-Differences analysis
model <- lm(foreign_policy ~ time * president, data = data)
summary(model)
```

```
Call:
lm(formula = foreign_policy ~ time * president, data = data)

Residuals:
    Min      1Q  Median      3Q     Max
-2.07596 -0.59723 -0.09367  0.52693  1.75938

Coefficients:
                            Estimate Std. Error t value Pr(>|t|)
(Intercept)                  8.30790    0.46965  17.690 6.29e-12 ***
```

```
timePre                         -3.11433    0.66418  -4.689 0.000246 ***
presidentNon-Military           -0.19856    0.66418  -0.299 0.768825
timePre:presidentNon-Military   -0.03933    0.93930  -0.042 0.967119
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.05 on 16 degrees of freedom
Multiple R-squared:  0.7366,    Adjusted R-squared:  0.6872
F-statistic: 14.92 on 3 and 16 DF,  p-value: 6.766e-05
```

```r
# Bias
bias <- coef(model)["timePost:presidentNon-Military"]
print(paste("Bias:", bias))
```

```
[1] "Bias: NA"
```

```r
# MSE
predicted_values <- predict(model)
mse <- mean((predicted_values - data$foreign_policy)^2)
print(paste("MSE:", mse))
```

```
[1] "MSE: 0.882283025085199"
```