PS3_CP

CP

2024-11-11

Question One : We are interested in knowing how the difference in campaign spending between incumbent and challenger affects the incumbent's vote share.

Sub-question 1 : Run a regression where the outcome variable is voteshare and the explanatory variable is difflog.

To do so, I start by downloading and loading my data.

```
url <- "https://raw.githubusercontent.com/ASDS-TCD/StatsI_Fall2024/main/datasets/incumbents_subset.csv"
data <- read.csv(url)
```

I can now run the linear regression.

```
model <- lm(voteshare ~ difflog, data = data)
summary(model)
```

```
##
## Call:
## lm(formula = voteshare ~ difflog, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.26832 -0.05345 -0.00377  0.04780  0.32749
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.579031   0.002251  257.19   <2e-16 ***
## difflog     0.041666   0.000968   43.04   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.07867 on 3191 degrees of freedom
## Multiple R-squared:  0.3673, Adjusted R-squared:  0.3671
## F-statistic:  1853 on 1 and 3191 DF,  p-value: < 2.2e-16
```

Sub-question 2 : Make a scatterplot of the two variables and add the regression line.
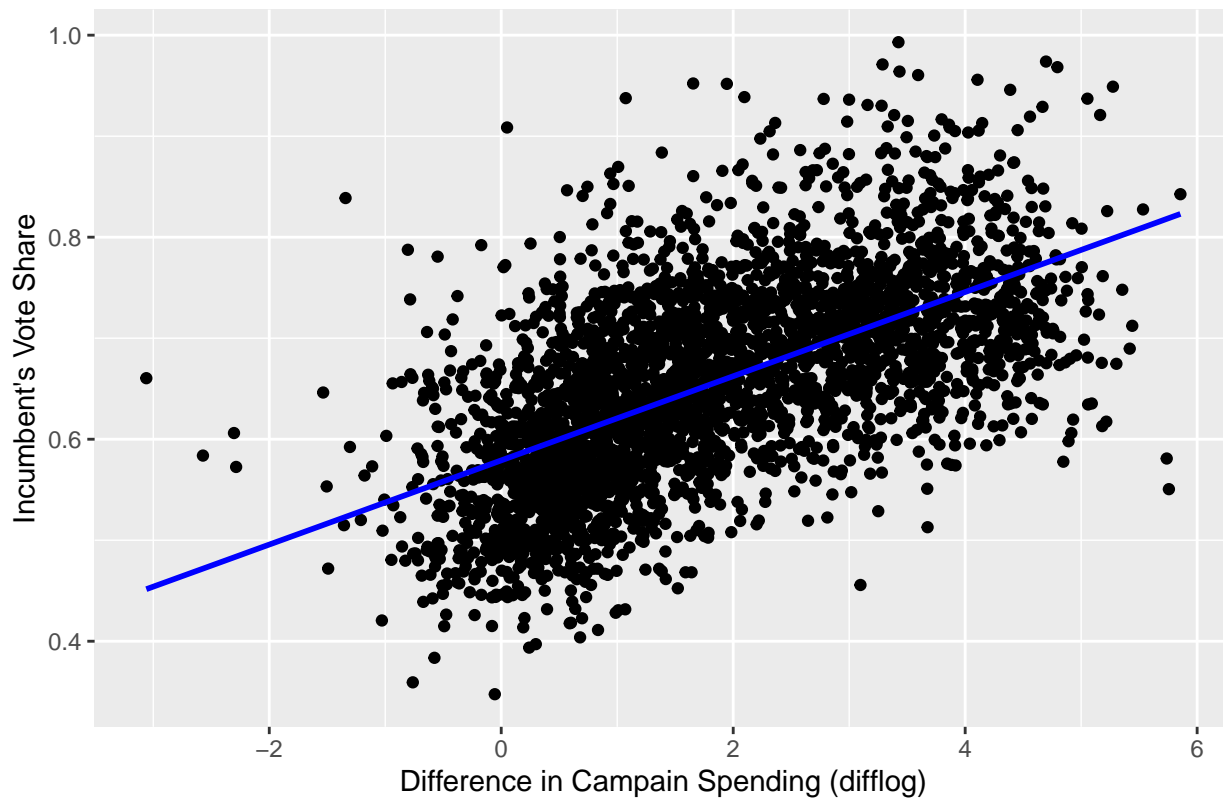
T do so, I create a scatterplot (including the regression line).

(I load and use ggplot).

Loading required package: ggplot2

`geom_smooth()` using formula = 'y ~ x'

Effect of Difference in Campaign Spending on Incumbent's Vote Share

Sub-question 3 : Save the Residuals in a Separate Object.

```
residuals_model <- resid(model)
```

Sub-question 4 : Write the prediction equation.

(To do so, I start by extracting coefficients before dysplaying the perdiction equation).

```
intercept <- coef(model)[1]
slope <- coef(model)[2]
cat("Prediction Equation: voteshare =", round(intercept, 3), "+", round(slope, 3), "* difflog\n")
```

Prediction Equation: voteshare = 0.579 + 0.042 * difflog

Question Two : We are interested in knowing how the difference between incumbent and challenger's spending and the vote share of the presidential candidate of the incumbent's party are related.

Part 1: Relationship Spending Difference/Presidential Candidate Vote Share.

Sub-question 1 : Run a regression where the outcome variable is presvote and the explanatory variable is difflog.

```r
url <- "https://raw.githubusercontent.com/ASDS-TCD/StatsI_Fall2024/main/datasets/incumbents_subset.csv"
data <- read.csv(url)
model1 <- lm(presvote ~ difflog, data = data)
summary(model1)
```

```
##
## Call:
## lm(formula = presvote ~ difflog, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.32196 -0.07407 -0.00102  0.07151  0.42743
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.507583   0.003161  160.60   <2e-16 ***
## difflog     0.023837   0.001359   17.54   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1104 on 3191 degrees of freedom
## Multiple R-squared:  0.08795,    Adjusted R-squared:  0.08767
## F-statistic: 307.7 on 1 and 3191 DF,  p-value: < 2.2e-16
```

Sub-question 2 :Make a scatterplot of the two variables and add the regression line.

(Once again, I use ggplot).

```r
library(ggplot2)
ggplot(data, aes(x = difflog, y = presvote)) +
  geom_point() +
  geom_smooth(method = "lm", color = "red", se = FALSE) +
  labs(title = "Effect of Spending Difference in Campain Spending on Presidential Vote Share",
       x = "Difference in Campain Spending (difflog)",
       y = "Presidential Candidate's Vote Share (presvote)")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

Effect of Spending Difference in Campain Spending on Presidential Vote Sh
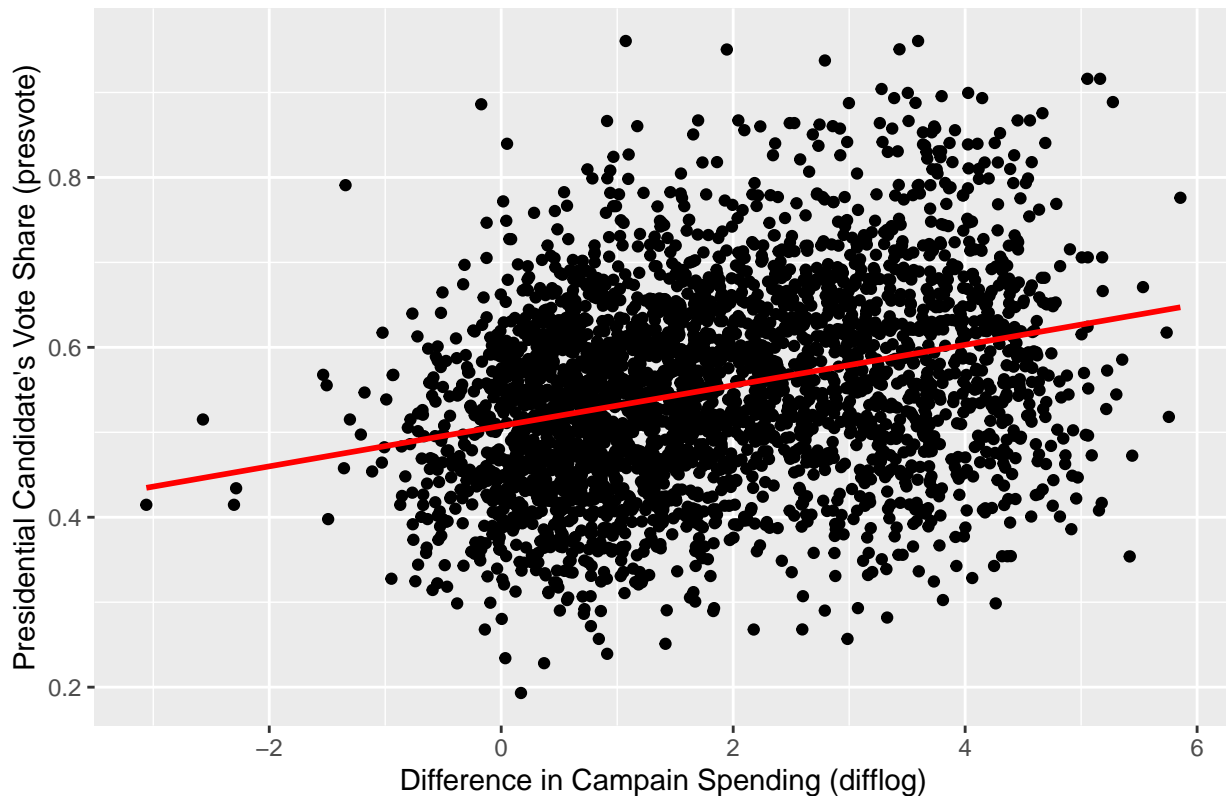
Sub-question 3 : Save the residuals of the model in a separate object.

```
residuals_model1 <- resid(model1)
```

Sub-question 4 : Save the residuals of the model in a separate object.

```
intercept1 <- coef(model1)[1]
slope1 <- coef(model1)[2]
cat("Prediction Equation for presvote:", round(intercept1, 3), "+", round(slope1, 3), "* difflog\n")
```

```
## Prediction Equation for presvote: 0.508 + 0.024 * difflog
```

Part 2 : Relationship Presidential Candidate Vote Share/Incumbent's Vote Share.

Sub-question 1 : Run a regression with voteshare as the outcome and presvote as the explanatory variable.

```
model2 <- lm(voteshare ~ presvote, data = data)
summary(model2)
```

```
##
## Call:
## lm(formula = voteshare ~ presvote, data = data)
##
```

```
## Residuals:
##      Min      1Q   Median      3Q      Max
## -0.27330 -0.05888  0.00394  0.06148  0.41365
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.441330    0.007599   58.08   <2e-16 ***
## presvote    0.388018    0.013493   28.76   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.08815 on 3191 degrees of freedom
## Multiple R-squared:  0.2058, Adjusted R-squared:  0.2056
## F-statistic:    827 on 1 and 3191 DF,  p-value: < 2.2e-16
```

Sub-question 2 : Create a Scatterplot with the Regression Line.

```
ggplot(data, aes(x = presvote, y = voteshare)) +
  geom_point() +
  geom_smooth(method = "lm", color = "yellow", se = FALSE) +
  labs(title = "Effect of Presidential Candidate Vote Share on Incumbent's Vote Share",
       x = "Presidential Candidate Vote Share (presvote)",
       y = "Incumbent's Vote Share (voteshare)")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```


Effect of Presidential Candidate Vote Share on Incumbent's Vote Share
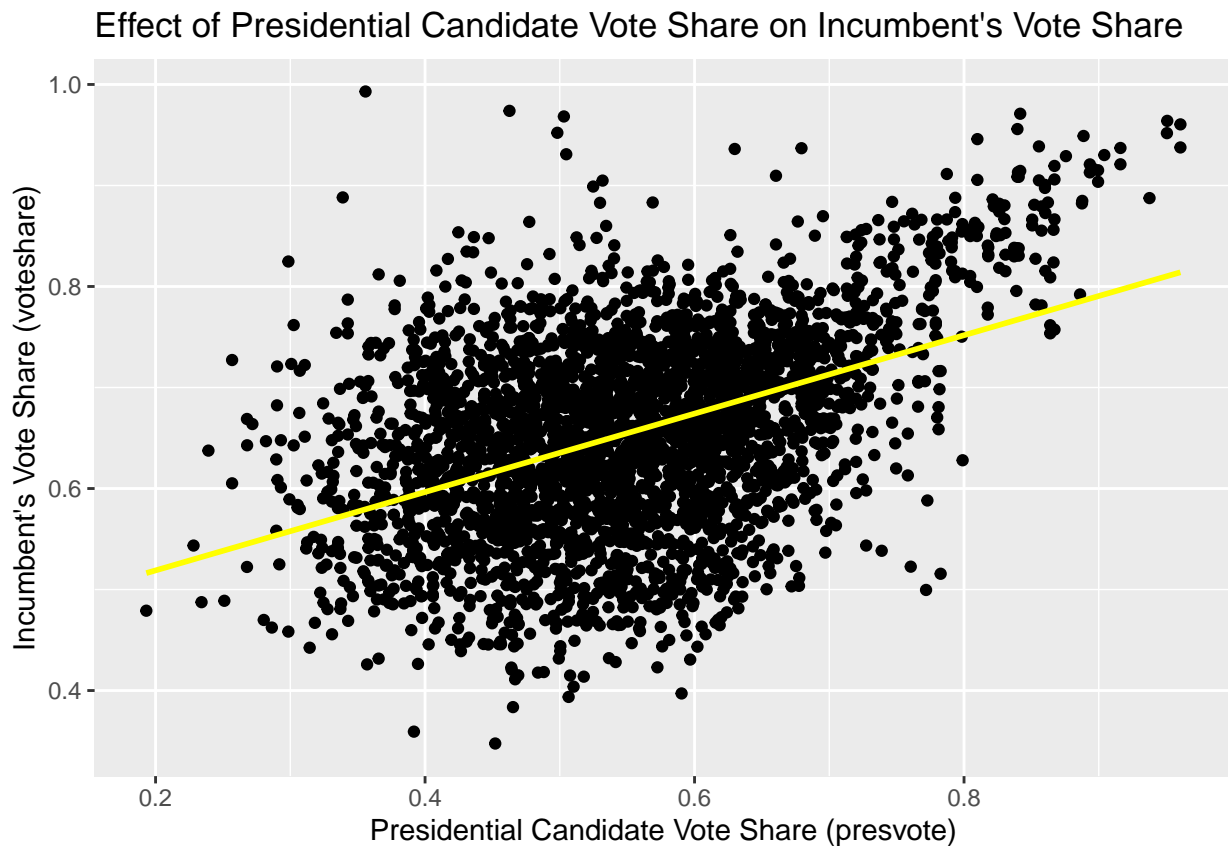
Sub-question 3 : Write the Prediction Equation.

```
intercept2 <- coef(model2)[1]
slope2 <- coef(model2)[2]
cat("Prediction Equation for voteshare:", round(intercept2, 3), "+", round(slope2, 3), "* presvote\n")

## Prediction Equation for voteshare: 0.441 + 0.388 * presvote
```

Question Three : We are interested in knowing how the vote share of the presidential candidate of the incum- bent's party is associated with the incumbent's electoral success.

I use the residual found and saved from question One and question Two.

```
# From Question 1 I have residuals of voteshare ~ difflog
residuals_voteshare_difflog <- resid(model)

# From Question 2 I have residuals of presvote ~ difflog
residuals_presvote_difflog <- resid(model1)
```

Sub-question 1 : Run a regression where the outcome variable is voteshare and the explanatory variable is presvote.

```
model_residuals <- lm(residuals_voteshare_difflog ~ residuals_presvote_difflog)
```

Sub-question 2 : Create a Scatterplot with Regression Line.

```
ggplot(data = NULL, aes(x = residuals_presvote_difflog, y = residuals_voteshare_difflog)) +
  geom_point() +
  geom_smooth(method = "lm", color = "pink", se = FALSE) +
  labs(title = "Relationship between Residuals of incumbent voteshare and residential condidate presvote
       x = "Residuals of residential condidate presvote ~ difflog",
       y = "Residuals of incumbent voteshare ~ difflog")

## `geom_smooth()` using formula = 'y ~ x'
```

Relationship between Residuals of incumbent voteshare and residential co
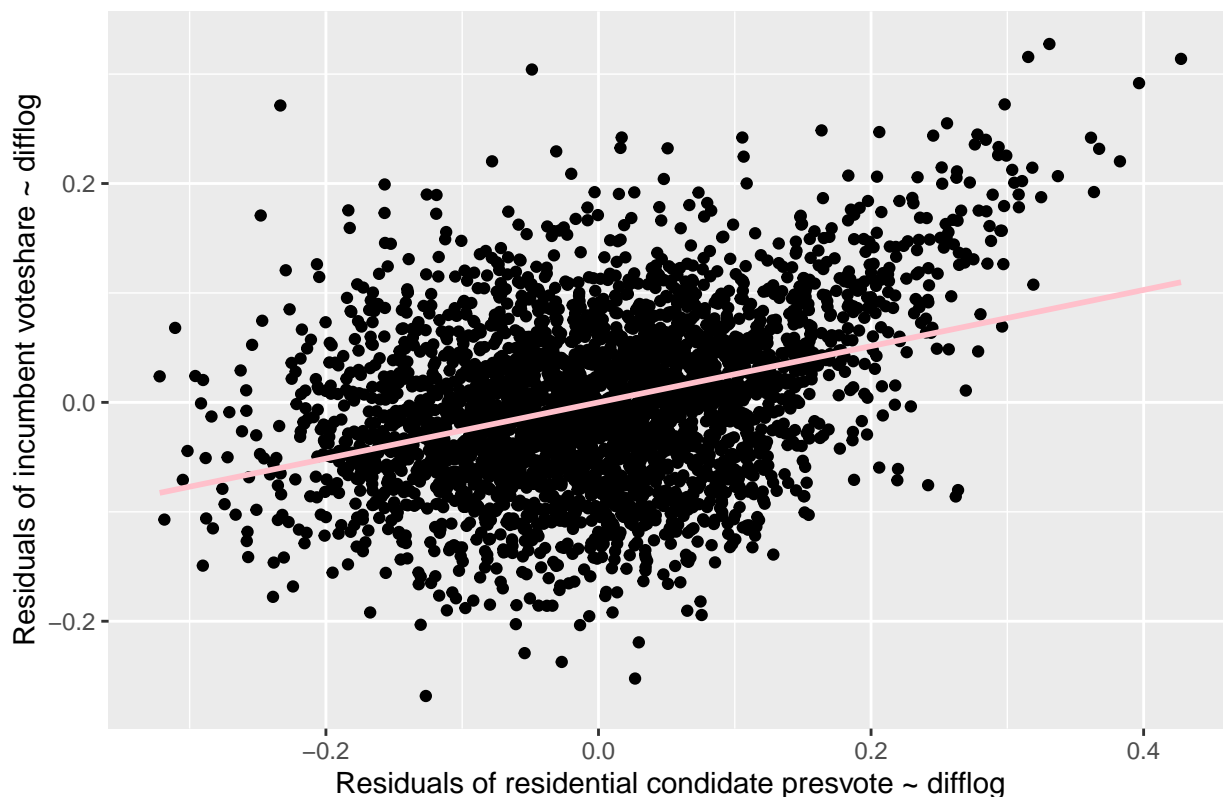
Sub-question 3 : Write the Prediction Equation.

```
intercept_residuals <- coef(model_residuals)[1]
slope_residuals <- coef(model_residuals)[2]
cat("Prediction Equation for residuals of voteshare:", round(intercept_residuals, 3), "+", round(slope_
```

```
## Prediction Equation for residuals of voteshare: 0 + 0.257 * residuals of presvote
```

Question Four : The residuals from part (a) tell us how much of the variation in voteshare is not explained by the difference in spending between incumbent and challenger. The residuals in part (b) tell us how much of the variation in presvote is not explained by the difference in spending between incumbent and challenger in the district.

Sub-question 1 : Run a regression where the outcome variable is the residuals from Question 1 and the explanatory variable is the residuals from Question 2.

```
model_combined <- lm(voteshare ~ difflog + presvote, data = data)
```

Sub-question 2 :Write the prediction equation.

```
intercept_combined <- coef(model_combined)[1]
slope_difflog <- coef(model_combined)[2]
slope_presvote <- coef(model_combined)[3]
cat("Prediction Equation for voteshare:", round(intercept_combined, 3), "+", round(slope_difflog, 3), ":
```

```
## Prediction Equation for voteshare: 0.449 + 0.036 * difflog + 0.257 * presvote
```

Sub-question 3 :Identify Similarity to Question 4 and Explain Why

```
summary(model_combined)
```

```
##
## Call:
## lm(formula = voteshare ~ difflog + presvote, data = data)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.25928  -0.04737  -0.00121   0.04618   0.33126
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.4486442  0.0063297   70.88   <2e-16 ***
## difflog     0.0355431  0.0009455   37.59   <2e-16 ***
## presvote    0.2568770  0.0117637   21.84   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.07339 on 3190 degrees of freedom
## Multiple R-squared:  0.4496, Adjusted R-squared:  0.4493
## F-statistic:  1303 on 2 and 3190 DF,  p-value: < 2.2e-16
```