Problem Set 3

Applied Stats/Quant Methods 1

Due: November 11, 2024

Instructions

- Please show your work! You may lose points by simply writing in the answer. If the problem requires you to execute commands in `R`, please include the code you used to get your answers. Please also include the `.R` file that contains your code. If you are not sure if work needs to be shown for a particular problem, please ask.
- Your homework should be submitted electronically on GitHub.
- This problem set is due before 23:59 on Sunday November 11, 2024. No late assignments will be accepted.

In this problem set, you will run several regressions and create an add variable plot (see the lecture slides) in `R` using the `incumbents_subset.csv` dataset. Include all of your code.

Question 1

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

1. Run a regression where the outcome variable is `voteshare` and the explanatory variable is `difflog`.

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

```
Call:
lm(formula = voteshare ~ difflog, data = inc.sub)

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
```

overall the model explains the difference in campaign spend between incumbents and challengers have a statistically significant and positive effect on vote share, with 36.73 percent of the variance in vote share explained by campaign spending differences.

2. Make a scatterplot of the two variables and add the regression line.

```r
plot(x=inc.sub$difflog, y=inc.sub$voteshare) # Scatter plot # Either specify single value (v for vertical)
abline(0.579031,0.041666) # Or intercept and slope
abline(model) # Use intercept and slope in model object
abline(model, col="red")
```

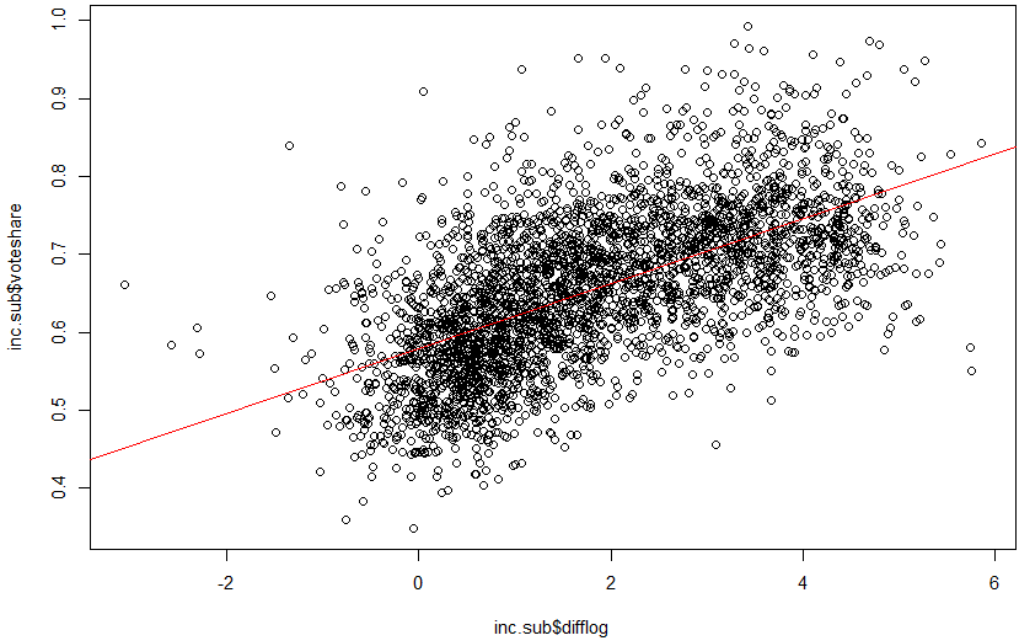

Figure 1: Scatterplot of difflog vs. voteshare with regression line

3. Save the residuals of the model in a separate object.

```
model_residuals<-model$residuals
model_residuals
```

4. Write the prediction equation

$$y = \beta_0 + \beta_1 x \tag{1}$$

- y: voteshare
- β_0: intercept
- β_1: slope
- x: difflog

$$voteshare = 0.579031 + 0.041666 * difflog \tag{2}$$

```
intercept<- model$coefficients[1]
intercept
coefficients_difflog<- model$coefficients[2]
coefficients_difflog
voteshare = intercept+coefficients_difflog
```

Question 2

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.

1. Run a regression where the outcome variable is 'presvote' and the explanatory variable is 'difflog'.

```
inc.sub <- read.csv("https://raw.githubusercontent.com/ASDS-TCD/StatsI_Fall2024/main/datasets/incumbents_subset.csv")
inc.sub
model_2<- lm(presvote~ difflog , data = inc.sub)
summary(model_2)
```

```
Call:
lm(formula = presvote ~ difflog, data = inc.sub)

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
```

Overall, the model shows that the difference in campaign spending between incumbents and challengers has a statistically significant and positive effect on the presidential candidate's vote share. However, only 8.7 percent of the variance in the presidential vote share is explained by campaign spending differences. Specifically, a one-unit increase in the difference in campaign spending between the incumbent and challenger leads to a 2.38 percentage point increase in the presidential vote share. The low R-squared value indicates that campaign spending is not a strong predictor of presidential vote share.

2. Make a scatterplot of the two variables and add the regression line.

```
plot(x = inc.sub$difflog, y = inc.sub$presvote)
abline(0.507583,0.023837)
abline(model_2,col='red')
```

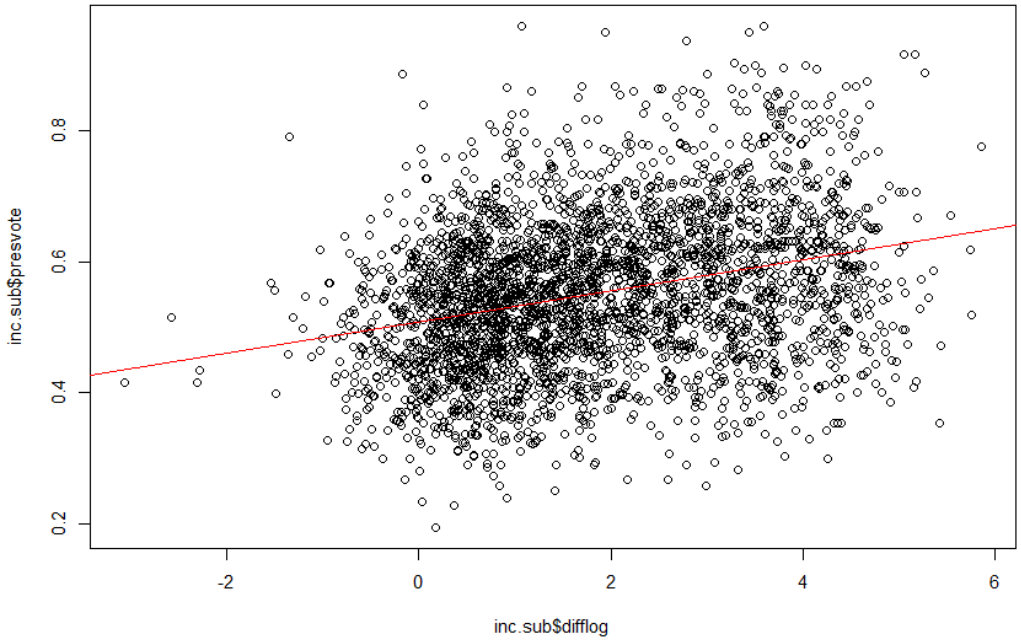

Figure 2: Scatterplot of presvote vs. difflog with regression line

3. Save the residuals of the model in a separate object.

```
model_residual_2 = model_2$residuals
model_residual_2
```

4. Write the prediction equation.

$$y = \beta_0 + \beta_1 x \tag{3}$$

- y: presvote
- β_0: intercept
- β_1: slope
- x: difflog

$$presvote = 0.507583 + 0.023837 * difflog \tag{4}$$

```
intercept<- model_2$coefficients[1]
intercept
coefficients_difflog<- model_2$coefficients[2]
coefficients_difflog
presvote = intercept+coefficients_difflog
```

Question 3

We are interested in knowing how the vote share of the presidential candidate of the incumbent's party is associated with the incumbent's electoral success.

1. Run a regression where the outcome variable is `voteshare` and the explanatory variable is `presvote`.

```
model_3<-lm(voteshare ~ presvote,data=inc.sub)
summary(model_3)
```

```
> summary(model_3)

Call:
lm(formula = voteshare ~ presvote, data = inc.sub)

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
```

Overall, the model shows that presvote (presidential vote share) has a statistically significant effect on voteshare (the vote share of the incumbent's party). 20.58 percent of the variance in the vote share is explained by presvote. Specifically, a one-unit increase in presvote leads to a 38.8 percentage point increase in the voteshare.

2. Make a scatterplot of the two variables and add the regression line.

```
plot(x= inc.sub$presvote,y=inc.sub$voteshare)
abline(0.441330,0.388018)
abline(model_3,col='red')
```

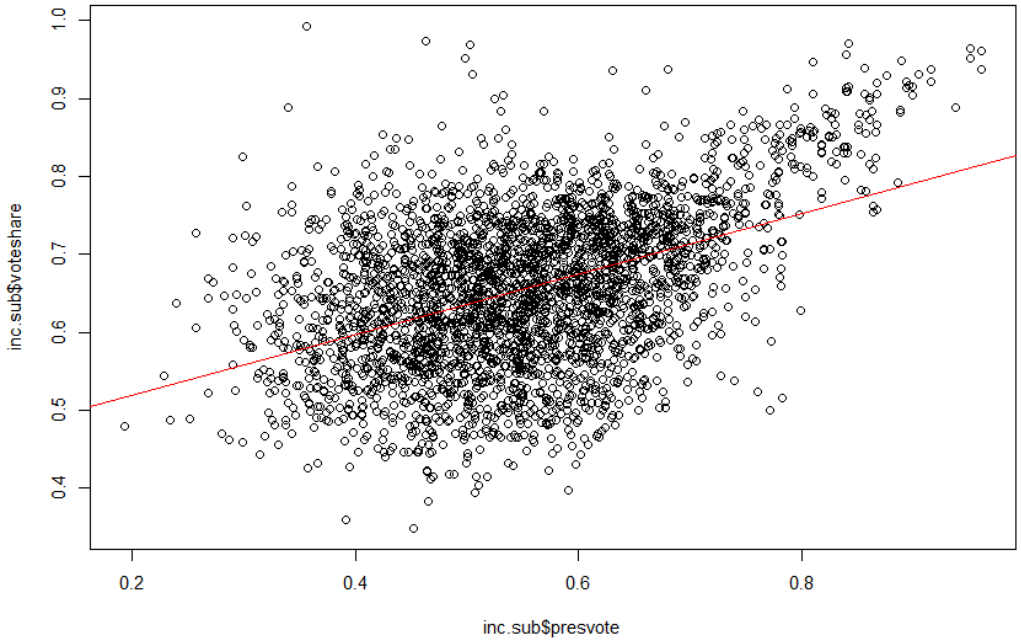

Figure 3: Scatterplot of voteshare vs. presvote with regression line

3. Write the prediction equation.

$$y = \beta_0 + \beta_1 x \tag{5}$$

- y: voteshare
- β_0: intercept
- β_1: slope
- x: presvote

$$voteshare = 0.441330 + 0.388018 * presvote \tag{6}$$

```
intercept<- model_3$coefficients[1]
intercept
coefficients_presvote<- model_3$coefficients[2]
coefficients_presvote
voteshare<-intercept+coefficients_presvote*presvote
```

Question 4

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.

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_4<- lm(model_residuals ~ model_residual_2,data=inc.sub)
summary(model_4)
```

```
Call:
lm(formula = model_residuals ~ model_residual_2, data = inc.sub)

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)      -5.934e-18  1.299e-03    0.00        1
model_residual_2  2.569e-01  1.176e-02   21.84   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.07338 on 3191 degrees of freedom
Multiple R-squared:   0.13,        Adjusted R-squared:  0.1298
F-statistic:   477 on 1 and 3191 DF,  p-value: < 2.2e-16
```

The model indicates a statistically significant effect of the residuals of presvote on the residuals of voteshare, with 13 percent of the variance in the voteshare residuals explained by the presvote residuals. The positive coefficient of 0.256 suggests that as the residuals of presvote increase, the residuals of voteshare also tend to increase.

2. Make a scatterplot of the two residuals and add the regression line.

```
plot(x = model_residual_2,y= model_residuals)
abline(model_4,col='red')
```

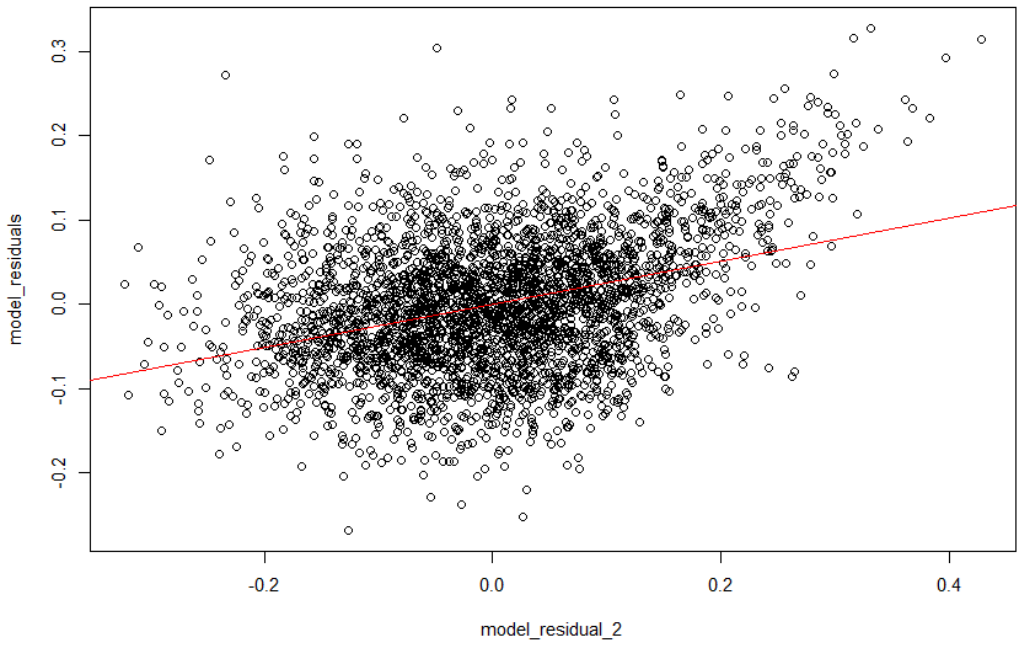

Figure 4: Scatterplot of model residual 2 vs. model residual with regression line

3. Write the prediction equation.

$$y = \beta_0 + \beta_1 x \tag{7}$$

- y: residuals of voteshare
- β_0: intercept
- β_1: slope
- x: residuals of presvote

$$\text{model\_residuals} = -5.934e-18 + 2.569e-01 * \text{model\_residual\_2} \tag{8}$$

```
intercept<- model_4$coefficients[1]
intercept
coefficients_residual_2<- model_4$coefficients[2]
coefficients_residual_2
model_residuals<- intercept+coefficients_residual_2*model_residual_2
```

Question 5

What if the incumbent's vote share is affected by both the president's popularity and the difference in spending between incumbent and challenger?

1. Run a regression where the outcome variable is the incumbent's `voteshare` and the explanatory variables are `difflog` and `presvote`.

```
model_5<- lm(voteshare ~ difflog+presvote,data=inc.sub)
summary(model_5)
```

```
lm(formula = voteshare ~ difflog + presvote, data = inc.sub)

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
```

here the multiple regression model suggests that both difflog (the difference in campaign spending) and presvote (the presidential vote share) are statistically significant predictors of incumbent's voteshare, where a one unit increase in difflog leads to 3.55 percent increase in voteshare, and a one unit increase in presvote leads to 25.68 percentage increase in voteshare. overall 44.93 percentage of variance in voteshare can be explained by both the predictors.

2. Write the prediction equation.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \tag{9}$$

- y: voteshare
- β_0: intercept
- β_1: coefficient. of difflog

- β_1: coefficient of presvote
- x_1: difflog
- x_2: presvote

$$voteshare = 0.4486442 + 0.0355431 * difflog + 0.2568770 * presvote \quad (10)$$

```
intercept<- model_5$coefficients[1]
intercept
coefficients_difflog<- model_5$coefficients[2]
coefficients_difflog
coefficients_presvote<-model_5$coefficients[3]
coefficients_presvote
voteshare = intercept+coefficients_difflog*difflog+coefficients_presvote*
    presvote
```

3. What is it in this output that is identical to the output in Question 4? Why do you think this is the case?

   - **Common Residuals**
     In Model 4, we used the residuals of presvote and voteshare to model their relationship. inorder to explain the variations in voteshare using the residuals of presvote. In Model 5, presvote is used as an independent variable in a multiple regression model with voteshare as the dependent variable. Because both models work with voteshare's unexplained variance, they end up with similar residuals. Therefore a common residuals in bot the models are observed,as both involve capturing variations in voteshare associated with presvote.

   - **Coeffiecnts of prevote**
     In Model 4, the residuals of presvote were used as the independent variable to explain the residuals of voteshare, while in Model 5, presvote itself is one of the predictors for voteshare. Since residuals of presvote in Model 4 help explain voteshare residuals, and presvote itself is a predictor in Model 5, thereby leading to a similar pattern in coefficients. there exists a strong effect of presvote on voteshare in both the models, with a significant p-value (¡ 2e-16).

   - **Standardsied residuals**
     Both models analyze the residuals of voteshare, which is why they show similar standard residual errors. In each model, the residuals represent the remaining variance in voteshare after accounting for the independent variables. Both models are designed to capture how presvote impacts voteshare: directly in Model 5 as an independent variable, and indirectly in Model 4 through the residuals of presvote. therefore similar standardized residual errors across the two models.

     In conclusion, presvote has a statistically significant positive effect on vote share.