

Problem Set 3

Ana Sofia Jesus 19327602

19/11/2023

Question 1

1)

After uploading the data in R, a regression was performed with the outcome variable being the incumbent's vote share and the explanatory variable being the difference in campaign spending between incumbent and challenger (difflog).

The hypotheses are as follows:

Null Hypothesis (H0): The coefficient for the predictor is equal to zero.

Alternative Hypothesis (Ha): The predictor's coefficient is not equal to zero.

The code used and the results obtained are as follows:

```
model_base <- lm(votesshare~difflog, data=inc.sub)
summary(model_base)
```

Call:

```
lm(formula = votesshare ~ 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

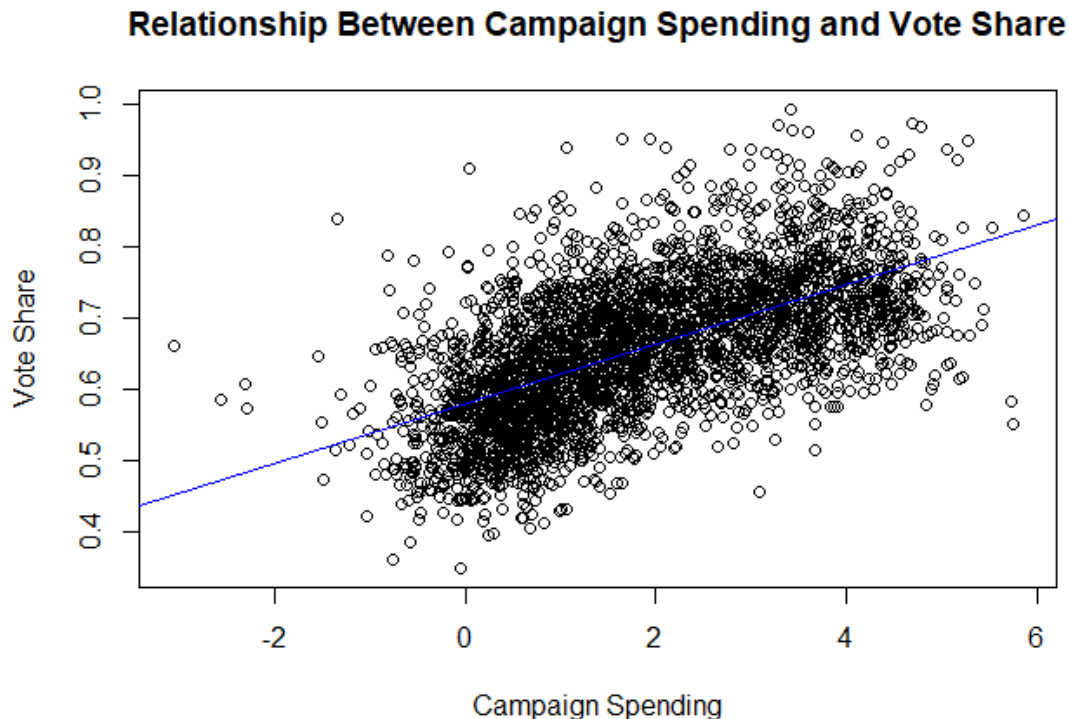
F-statistic: 1853 on 1 and 3191 DF, p-value: < 2.2e-16

According to the results of this model, for each unit increase in the difference in campaign spending between incumbent and challenger, an average positive increase of 0.041666 in the incumbent's vote share is predicted. The linear regression model is statistically significant, providing evidence against the null hypothesis and suggesting that campaign spending has a statistically significant impact on the incumbent's vote share. The model explains approximately 36.71 per cent of the variability in the dependent variable, and the large value of the F-statistic associated with a small p-value reinforces the overall significance of the model.

2)

In this section, a scatterplot of the two variables with a regression line is produced using the following code:

```
plot(inc.sub$difflog, inc.sub$voteshare, main="Relationship Between Campaign Spending and  
Vote Share",  
      xlab="Campaign Spending", ylab="Vote Share")  
abline(lm(voteshare ~ difflog, data=inc.sub), col="blue")
```



This graph reinforces the previous conclusions of a positive effect of campaign spending difference on the incumbent's vote share.

3)

As requested, residuals were saved in a separate object using the following code:

```
residuals <- resid(model_base)
```

4)

The prediction equation was inserted in R using the following code and yielding the following equation:

```
intercept <- coef(model_base)[1]  
slope <- coef(model_base)[2]  
prediction_equation_base <- paste("voteshare = ", round(intercept, 4), " + ", round(slope, 4), "  
* difflog")  
cat("Prediction Equation:\n", prediction_equation_base, "\n")
```

Prediction Equation: voteshare = 0.579 + 0.0417 * difflog

Question 2

1)

A simple regression was performed to assess the relationship between the outcome variable: vote share of the presidential candidate of the incumbent's party and the predictor: the difference between incumbent and challenger's campaign spending.

The hypotheses are as follows:

Null Hypothesis (H0): The coefficient for the predictor is equal to zero.

Alternative Hypothesis (Ha): The predictor's coefficient is not equal to zero.

The code used and the results obtained are as follows:

```
model_pres<-lm(presvote~difflog, data=inc.sub)
```

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

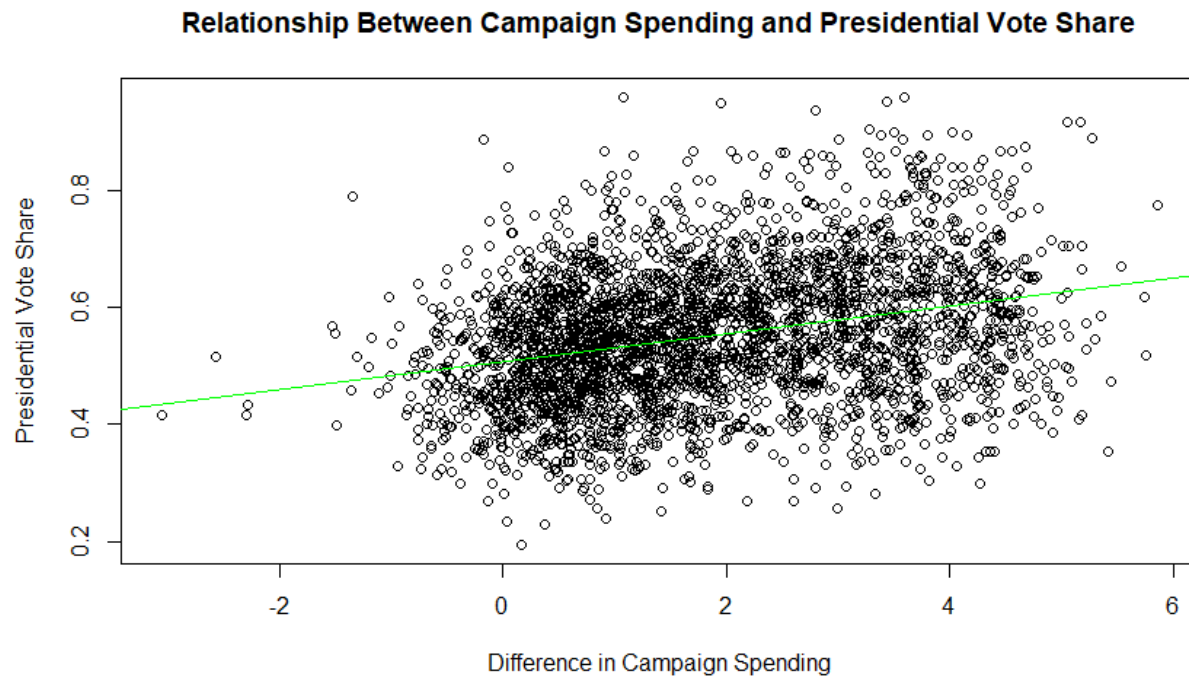
F-statistic: 307.7 on 1 and 3191 DF, p-value: < 2.2e-16

The estimated intercept 0.5076 represents the expected value of the outcome variable (presvote) when the independent variable (difflog) is zero. The estimated coefficient for difflog is 0.0238. This represents the estimated change in the outcome variable (presvote) for a one-unit increase in the predictor (difflog). In other words, on average, a one-unit increase in the difference in campaign spending is associated with a 0.0238 increase in presidential vote share. This regression model suggests that there is a statistically significant relationship between the difference in campaign spending (difflog) and presidential vote share (presvote), providing evidence against the null hypothesis. However, the R-squared value indicates that the model explains a small portion of the variability in the outcome variable (presidential vote share).

2)

In this section, a scatterplot of the two variables with a regression line is produced using the following code:

```
plot(inc.sub$difflog, inc.sub$presvote, main="Relationship Between Campaign Spending and  
Presidential Vote Share",  
      xlab="Difference in Campaign Spending", ylab="Presidential Vote Share")  
abline(lm(presvote ~ difflog, data=inc.sub), col="green")
```



This graph reinforces the previous conclusions of a positive effect of campaign spending difference on the vote share of the presidential candidate of the incumbent's party.

3)

As requested, residuals were saved in a separate object using the following code:

```
residuals_pres <- resid(model_pres)
```

4)

The prediction equation was inserted in R using the following code and yielding the following equation:

```
intercept_pres <- coef(model_pres)[1]
slope_pres <- coef(model_pres)[2]
prediction_equation_pres <- paste("presvote = ",
round(intercept_pres, 4), " + ", round(slope_pres, 4), " * difflog")
cat("Prediction Equation:\n", prediction_equation_pres, "\n")
```

Prediction Equation: $\text{presvote} = 0.5076 + 0.0238 * \text{difflog}$

Question 3

1)

A simple regression was performed to assess the relationship between the outcome variable: incumbent's electoral success (voteshare) and the predictor: vote share of the presidential candidate of the incumbent's party (presvote)

The hypothesis are as follows:

Null Hypothesis (H0): The coefficient for the predictor is equal to zero.

Alternative Hypothesis (Ha): The predictor's coefficient is not equal to zero.

The code used and the results obtained are as follows:

```
model_inc<-lm(votesshare~presvote, data=inc.sub)

Call:
lm(formula = votesshare ~ 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
```

These results suggest that, on average, for each one-unit increase in the vote share of the presidential candidate, the incumbent's vote share is expected to increase by approximately 0.388018, assuming a linear relationship between the two variables. The model is statistically significant, indicating that presidential vote is a meaningful predictor of the incumbent's vote share, providing evidence against the null hypothesis.

2)

In this section, a scatterplot of the two variables with a regression line is produced using the following code:

```
plot(inc.sub$presvote, inc.sub$votesshare, main="Relationship Between Presidential Vote Share
and incumbent's electoral success",
     xlab="Presidential vote share", ylab="Incumbent's success")
abline(lm(votesshare ~ presvote, data=inc.sub), col="Purple")
```

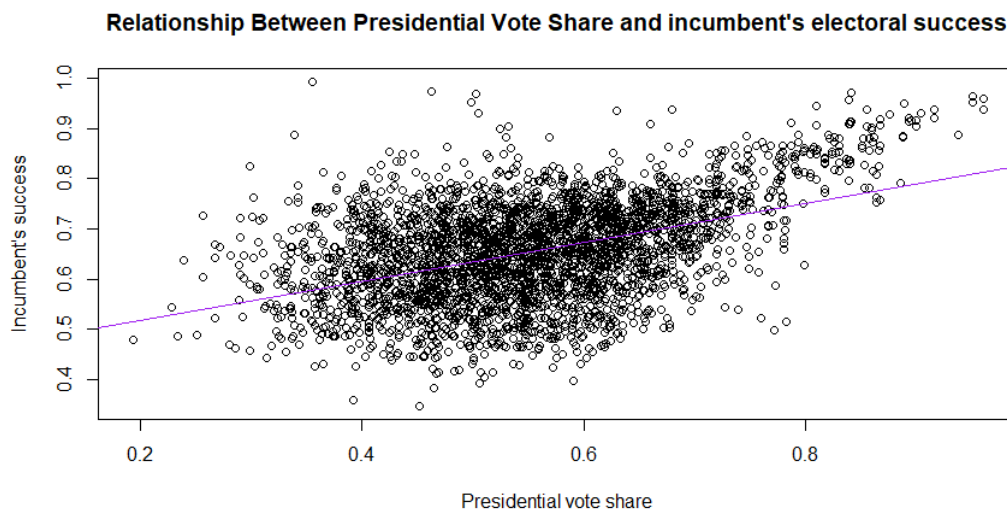
This graph reinforces the previous conclusions of a positive effect of the vote share of the presidential candidate on the electoral success of incumbent's party.

3)

The prediction equation was inserted in R using the following code and yielding the following equation:

```
intercept_inc <- coef(model_inc)[1]
slope_inc <- coef(model_inc)[2]
prediction_equation_inc <- paste("votesshare = ", intercept_inc, " + ", slope_inc, " * presvote")
cat("Prediction Equation:\n", prediction_equation_inc, "\n")
```

Prediction Equation: votesshare = 0.441329881204297 + 0.38801844338744 * presvote



Question 4

1)

A simple regression was performed to assess the relationship between the outcome variable:residuals 1 and the predictor: residuals 2.

The hypotheses are as follows:

Null Hypothesis (H0): The coefficient for the predictor is equal to zero.

Alternative Hypothesis (Ha): The predictor's coefficient is not equal to zero.

The code used and the results obtained are as follows:

```
model_residuals<-lm(residuals~residuals_pres, data=inc.sub)
```

Call:

```
lm(formula = residuals ~ residuals_pres, 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
residuals_pres	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

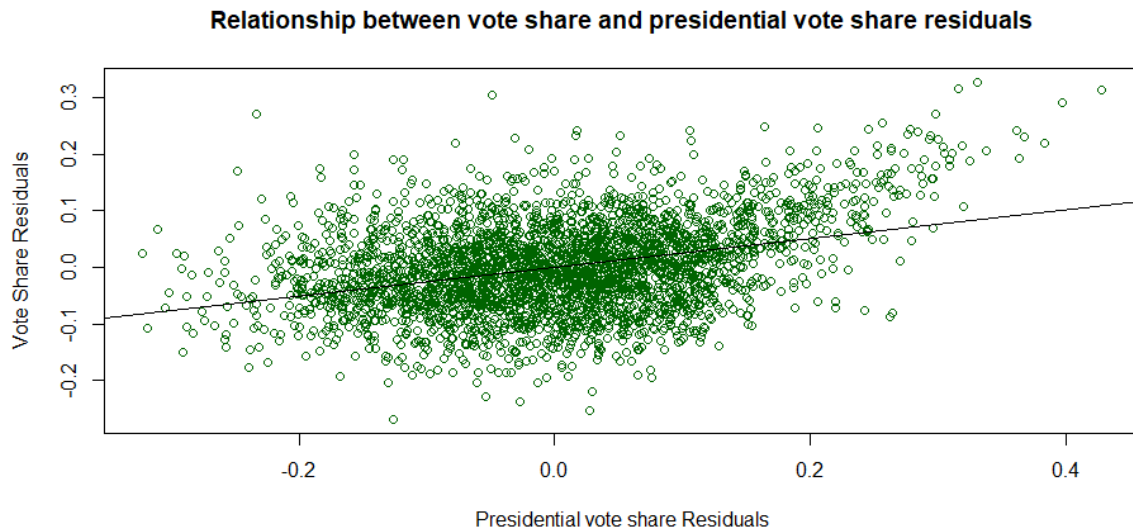
F-statistic: 477 on 1 and 3191 DF, p-value: < 2.2e-16

These results suggest that, on average, for each one-unit increase in residuals 2 (residualspres), residuals 1 (residuals) are expected to increase by approximately 0.2569, assuming a linear relationship between the two variables. The model is statistically significant (since p-value associated with the predictor's coefficient is extremely small) indicating that residuals resulting from the interaction between campaign spending and presidential vote share are a meaningful predictor of residuals resulted from the relationship between campaign spending and incumbent's vote share, providing evidence against the null hypothesis.

2)

In this section, a scatterplot of the two variables with a regression line is produced using the following code:

```
plot(residuals_pres, residuals, main="Relationship between vote share and presidential vote  
share residuals",  
     xlab="Presidential vote share Residuals", ylab="Vote Share Residuals",  
     col = "darkgreen")  
abline(lm(residuals ~ residuals_pres), col="black")
```



This graph reinforces the previous conclusions of a positive effect between the residuals of both variables.

3)

The prediction equation was inserted in R using the following code and yielding the following equation:

```
intercept_res <- coef(model_residuals)[1]  
slope_res <- coef(model_residuals)[2]  
intercept_formatted <- formatC(intercept_res, format = "e", digits = 10)  
slope_formatted <- formatC(slope_res, format = "f", digits = 10)  
prediction_equation_res <- paste("residuals = ", intercept_formatted, " + ", slope_formatted,  
  " * residuals_pres")  
cat("Prediction Equation:\n", prediction_equation_res, "\n")
```

Prediction Equation: residuals = -5.9340779240e-18 + 0.2568770127 * residuals_pres

Question 5

1)

A multiple linear regression was performed to assess the relationship between the outcome variable: incumbent's electoral success (voteshare) and two predictors: the difference in spending between incumbent and challenger (difflog) and the vote share of the presidential candidate of the incumbent's party (presvote).

The hypotheses were formulated as follows:

H0: All of the predictor's coefficients in the model are equal to zero.

Ha: At least one coefficient in the model is not equal to zero.

The code used and the results obtained are as follows:

```
model_total<-lm(votesshare ~ difflog + presvote, data = inc.sub)
```

Call:

```
lm(formula = votesshare ~ 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

The value found for the intercept (0.4486) corresponds to the estimated value of the outcome (voteshare) when both predictors (difflog and presvote) are zero. In this context, it represents the expected vote share when there is no difference in campaign spending and presidential vote share.

The coefficient for Difflog (0.03554) represents the average estimated change in vote share for a one-unit increase in difflog, holding presvote constant. This coefficient indicates the impact of the difference in campaign spending on voteshare, assuming the presidential vote share is constant.

The coefficient for Presvote (0.2569) represents the average estimated change in vote share for a one-unit increase in presvote, holding difflog constant. This coefficient represents the impact of the presidential vote share on the incumbent's voteshare, assuming the difference in campaign spending is constant.

From these results it is possible to verify that all coefficients are statistically significant since p-values associated with all coefficients are very close to zero, suggesting strong evidence against the null hypothesis that the true coefficients are zero. The F-statistic (1303) is quite high and is associated with a low p-value (< 2.2e-16) suggesting that the model as a whole is statistically significant, indicating that both predictors are meaningful in predicting our outcome variable voteshare.

2)

The prediction equation was inserted in R using the following code and yielding the following equation:

```
coefficients <- coef(model_total)
prediction_equation_total <- paste("votesshare = ", round(coefficients[1], 4),
                                   " + ", round(coefficients[2], 4), " * difflog",
                                   " + ", round(coefficients[3], 4), " * presvote")
cat("Prediction Equation:\n", prediction_equation_total, "\n")
```

Prediction Equation: voteshare = 0.4486 + 0.0355 * difflog + 0.2569 * presvote

3)

In comparing the regression outputs for Question 4 and Question 5, it is observed that the residuals in both models are identical. This can be attributed to the fact that both analyses utilize the exact same data and variables. In Question 4, we examined the residuals from the relationship between voteshare, campaign spending (difflog), and the existing presidential vote share ('presvote'). Question 5, on the other hand, includes these same variables as factors in the model. Residuals represent the amount of variability in our outcome variable that is not explained by our predictor(s). Hence, the residuals obtained in Question 4 represent the unexplained variation in voteshare after accounting for spending differences and the existing presidential vote share. When these same variables are included as predictors in Question 5, the model aims to explain essentially the same variation, resulting in identical residuals between the two analyses.