Applied Statistical Analysis - ProblemSet 3

Eoghan O'Sullivan

2024-11-12

1 Question 1

For this question, I formulated a research question: "do campaign spending differences affect voteshare?". At the same time, I used R scripting as the statistical method of analysis. The variables were provided in the brief. These were difflog and voteshare, from provided data. The explanatory variable, difflog, is used in the regression analysis as is the outcome variable, voteshare. I produced a regression analysis after determining that the dependent variable follows a normal distribution by showing a bell-shaped curve in a histogram, which I plotted and reviewed. Given it was bell-shaped, I proceeded with a linear regression and below are the results of the analysis.

1.1 Part 1

Run a regression where the outcome variable is voteshare and the explanatory variable is difflog. For this, I used the lm function in R and created a summary from that object. The two variables were used and the data incumbents_subset.csv which was read-in and made an object. The results of this regression are in figure 1.

```
Call:
lm(formula = incumbts$voteshare ~ incumbts$difflog, data = incumbts)

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 ***
incumbts$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
```

Figure 1: Regression statistics.

1.2 Part 2

Make a scatterplot of the two variables and add the regression line. This task was approached by me using a similar procedure as the last task. I plotted the scatter plot firstly to see if the distribution of data points could be described with a straight line. When I was confident it was, I fitted a linear model and proceeded to draw the regression line. Below in figure 2 is the resulting graph. To fit the model, I again used the lm function. I created an object called lm_model. To draw the line I used this object with the abline function for drawing the line and coloured it in red.

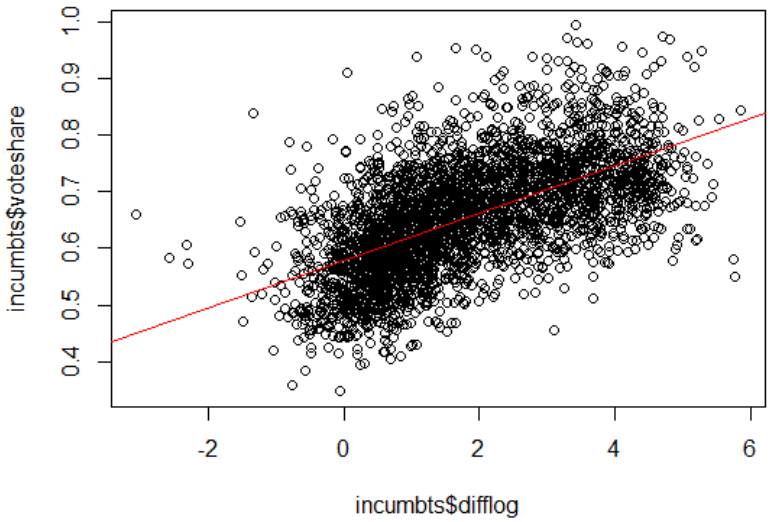

Figure 2: Scatter plot with regression line.

1.3 Part 3

Save the residuals of the model in a separate object. To save the calculated residuals, I made an object called question1_residuals to store the data for 3193 observations across 20 variables; again with the function lm.

1.4 Part 4

Write the prediction equation. I wrote the prediction equation as follows, having calculated the intercept as

$$\beta o \tag{1}$$

and the slope, as

$$\beta x \tag{2}$$

The equation is:

$$\hat{y} = 0.579 + 0.042 * difflog \tag{3}$$

2 Question 2

For this question, I formulated a research question: "do campaign spending differences affect presvote?". At the same time, I used R scripting as the statistical method of analysis. The variables were provided in the brief. These were difflog and presvote, from provided data. The explanatory variable, difflog, is used in the regression analysis as is the outcome variable, presvote. I produced a regression analysis and below are the results of the analysis.

2.1 Part 1

Run a regression where the outcome variable is presvote and the explanatory variable is difflog. The results of this regression are in figure 3.

```
Call:
lm(formula = incumbts$presvote ~ incumbts$difflog, data = incumbts)

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 ***
incumbts$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
```

Figure 3: Regression statistics.

2.2 Part 2

Make a scatterplot of the two variables and add the regression line. I was confident the scatter plot showed a a straight line and I then fitted a linear model and proceeded to draw the regression line. Below in figure 4 is the resulting graph. To fit the model, I again used the lm function. This time, I again created an object called lm_model. To draw the line I used this object with the abline function for drawing the line and coloured it in red.

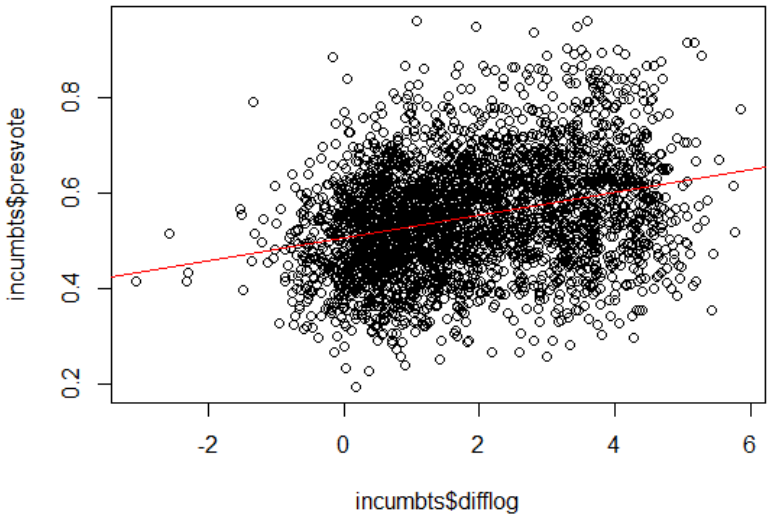

Figure 4: Scatter plot with regression line.

2.3 Part 3

Save the residuals of the model in a separate object. To save the calculated residuals, I made an object called question2_residuals to store the data for 3193 observations across 20 variables; again with the function lm.

2.4 Part 4

Write the prediction equation. I wrote the prediction equation as follows, having calculated the intercept as

$$\beta o \tag{4}$$

and the slope, as

$$\beta x \tag{5}$$

The equation is:

$$\hat{y} = 0.508 + 0.024 * diff log \tag{6}$$

3 Question 3

For this question, I formulated a research question: "how does the incumbent President's party vote share affect his/her electoral success?". The variables

were provided in the brief. These were presvote and voteshare, from provided data. The explanatory variable, presvote, is used in the regression analysis as is the outcome variable, voteshare. I produced a regression analysis after determining that the dependent variable follows a normal distribution by showing a bell-shaped curve in a histogram, which I plotted and reviewed. Given it was bell-shaped, I proceeded with a linear regression and below are the results of the analysis.

3.1 Part 1

Run a regression where the outcome variable is voteshare and the explanatory variable is presvote. The results of this regression are in figure 5.

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

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 ***
incumbts$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
```

Figure 5: Regression statistics.

3.2 Part 2

Make a scatterplot of the two variables and add the regression line. I was confident the scatter plot showed a straight line and I then fitted a linear model and proceeded to draw the regression line. Overleaf in figure 6 is the resulting graph. To fit the model, I again used the lm function. I created an object called lm_model. To draw the line I used this object with the abline function for drawing the line and coloured it in red.

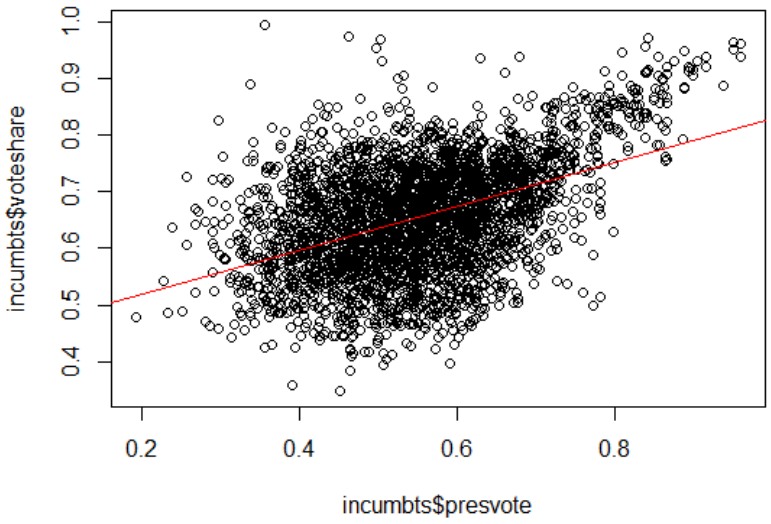

Figure 6: Scatter plot with regression line.

3.3 Part 3

Write the prediction equation. I wrote the prediction equation as follows, having calculated the intercept as

$$\beta o \tag{7}$$

and the slope, as

$$\beta x \tag{8}$$

The equation is:

$$\hat{y} = 0.441 + 0.388 * presvote \tag{9}$$

4 Question 4

This question I approached by using previously calculated residuals.

4.1 Part 1

Run a regression where the outcome variable is the residuals from Question 1 and the explanatory variable is the residuals from Question 2. The results of this regression analysis are below in figure 7.

```
Call:
lm(formula = question1_residuals ~ question2_residuals, data = incumbts)

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
question2_residuals  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
```

Figure 7: Regression statistics.

4.2 Part 2

Make a scatterplot of the two residuals and add the regression line. I approached this by using the lm function in R and wrote as arguments the residuals from questions 1 and 2 and the data itself. The graph in figure 8 refers.

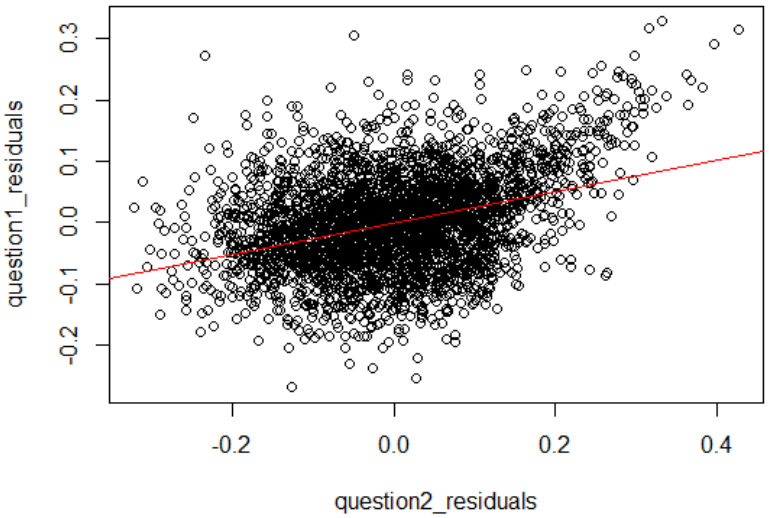

Figure 8: Scatter plot with regression line.

5 Question 5

For this question, I formulated a research question: "Is the incumbent's vote-share affected by both the president's popularity and spending differences?". At the same time, I used R scripting as the statistical method of analysis. The variables were provided in the brief. These were difflog, voteshare and presvote from provided data. The explanatory variables, presvote and difflog, are used in the regression analysis as is the outcome variable, voteshare. I reviewed assumptions and I proceeded with a multivariate regression and below are the results of the analysis.

5.1 Part 1

Run a regression where the outcome variable is the incumbent's vote-share and the explanatory variables are difflog and presvote. For this part, I used the lm function in R and created a summary from that object. The results of this regression are in figure 9.

```
Call:
lm(formula = Voteshare ~ Difference + Presvote)

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 ***
Difference  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
```

Figure 9: Regression statistics.

5.2 Part 2

Write the prediction equation. I wrote the prediction equation as 0.449 + 0.257 * difflog + 0.036 * presvote