

Problem Set 3

Rafaela Alves

November 11, 2024

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`.

```
1 model1 <- lm(voteshare ~ difflog, data=inc.sub)
2 summary(model1)
```

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 ***

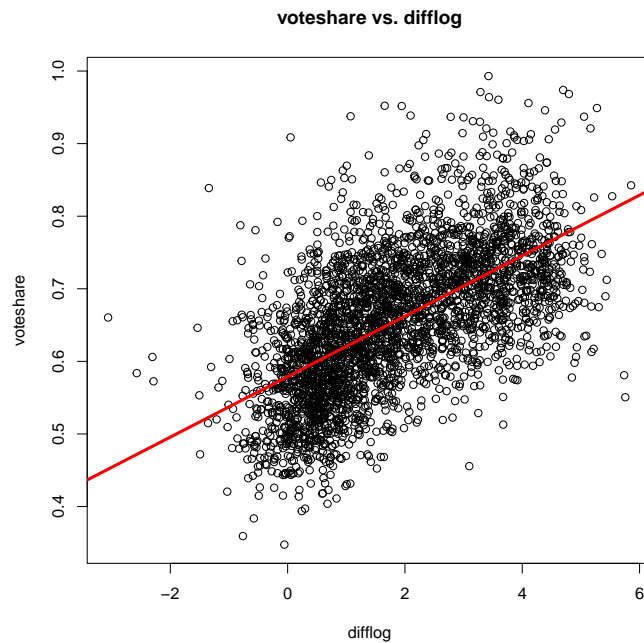
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

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

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



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

```
1 residuals_model1 <- residuals(model1)
2 head(residuals_model1)
```

1	2	3	4	5	6
-0.00042	-0.03168	-0.00455	0.03866	0.03552	0.03228

4. Write the prediction equation.

$$\text{voteshare} = 0.579 + 0.041 * \text{difflog}$$

For every one-unit increase in difflog, the predicted voteshare increases by approximately 0.041 units. The starting point for voteshare is 0.579 when difflog is zero.

```
1 coefficients(model1)
```

(Intercept)	difflog
0.57903071	0.04166632

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`.

```
1 model2 <- lm(presvote ~ difflog, data=inc.sub)
2 summary(model2)
```

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 ***

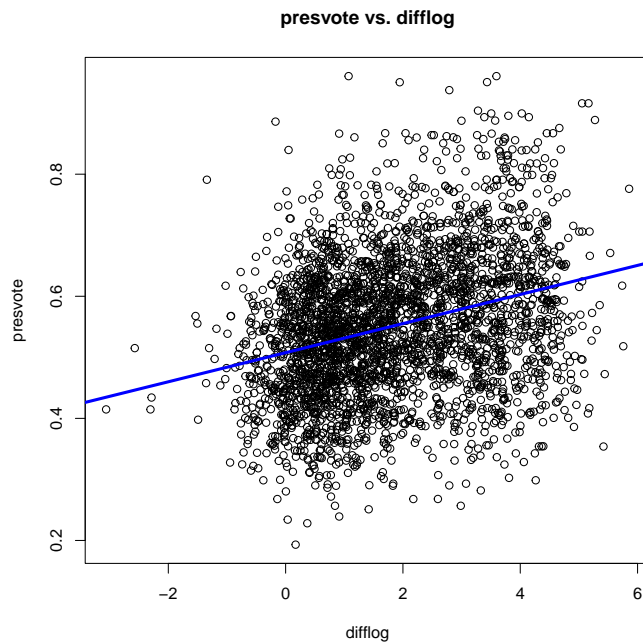
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

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

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



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

```
1 residuals_model2 <- residuals(model2)
2 head(residuals_model2)
```

1	2	3	4	5	6
0.00560	0.03757	-0.05313	-0.05299	-0.04584	0.07433

4. Write the prediction equation.

$$\text{presvote} = 0.507 + 0.023 * \text{difflog}$$

For every one-unit increase in difflog, the predicted presvote increases by approximately 0.023 units. The starting point for presvote when difflog is zero is around 0.507.

```
1 coefficients(model2)
```

(Intercept)	difflog
0.50758333	0.02383723

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**.

```
1 model3 <- lm(voteshare ~ presvote, data=inc.sub)
2 summary(model3)
```

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 ***

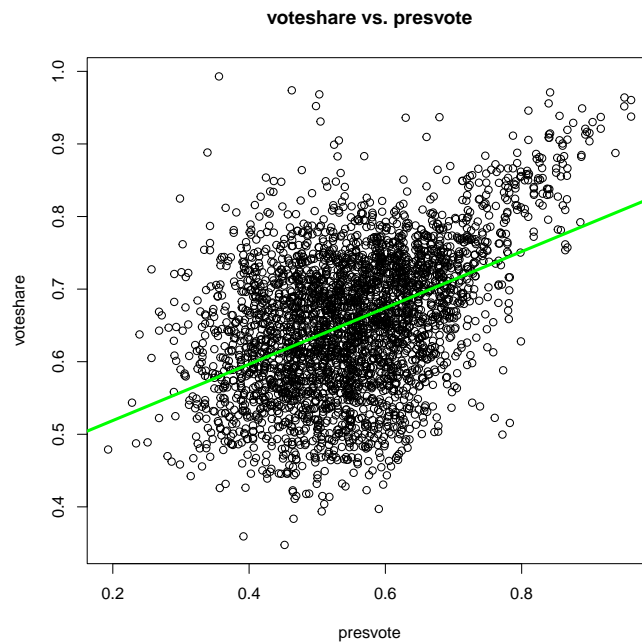
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

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

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



3. Write the prediction equation.

$$\text{voteshare} = 0.441 + 0.388 * \text{presvote}$$

For every one-unit increase in presvote, the predicted voteshare increases by approximately 0.388 units. The starting point for voteshare when presvote is zero is around 0.441.

```
1 coefficients(model3)
```

```
(Intercept)    presvote
0.4413299    0.3880184
```

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.

Y outcome = residuals_model1

X predictor = residuals_model2

```
1 reg_residuals <- lm(residuals_model1 ~ residuals_model2)
2 summary(reg_residuals)
```

Call:

```
lm(formula = residuals_model1 ~ residuals_model2)
```

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_model2	2.569e-01	1.176e-02	21.84	<2e-16 ***

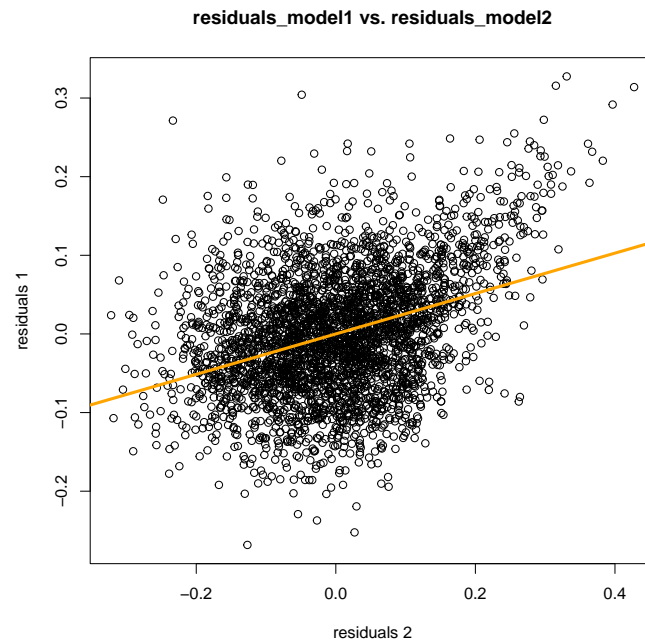
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

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

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



3. Write the prediction equation.

$$\text{residuals_model1} = -5.934 + 2.568 * \text{residuals_model2}$$

For each unit increase in the residuals_model2, the residuals_model1 increase by approximately 2.568 units. The starting point for residuals_model1 is negative, -5.934, when residuals_model2 is zero.

```
1 coefficients(reg_residuals)
```

(Intercept)	residuals_model2
-5.934078e-18	2.568770e-01

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`.

Y = `voteshare`

X1 = `difflog` (campaigning spending)

X2 = `presvote` (president's popularity)

```
1 model5 <- lm(voteshare ~ difflog + presvote, data=inc.sub)
2 summary(model5)
```

Call:

```
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

2. Write the prediction equation.

$$\text{voteshare} = 0.448 + 0.035 * \text{difflog} + 0.256 * \text{presvote}$$

For each unit increase in difflog, the voteshare increases by approximately 0.035 units, assuming presvote remains constant. For each unit increase in presvote, the voteshare increases by approximately 0.256 units, assuming difflog remains constant. The starting point for voteshare is 0.448 when both, difflog and presvote are zero.

```
1 coefficients(model5)
```

(Intercept)	difflog	presvote
0.44864422	0.03554309	0.25687701

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

The RESIDUALS in Question 4 are identical to the ones in Question5.

In the regression in Question 4, we're regressing the residuals from model1 (what remains in **voteshare** after accounting for **difflog**) and model2 (what remains in **presvote** after accounting for **difflog**). It checks if there is a relationship between these two residuals, which corresponds to the part of **voteshare** explained by **presvote** after controlling for **difflog**.

In Question 5 regression, we use **difflog** and **presvote** as predictors of **voteshare**. The residuals in this case are what remains in **voteshare** after accounting **difflog** and **presvote**.

Summarizing, both sets of residuals are showing the "unexplained" part of **voteshare** after accounting the effect of **difflog** and **presvote**.