Alignment Part 2

Peter Jones

2024-06-11

The Original Model

As indicated in the previous packet, the regression below (requested in point one of the recent memo) indicates a statistically significant relationship between our variables of interest. Specifically, the slope indicates that for each one percentage point increase in the electoral votes, the percentage of the popular vote for the winning candidate decreases by approximately 0.5603 percentage points. This negative relationship suggests that as the percentage of electoral votes increases, the percentage of the total vote won by the winning candidate tends to slightly decrease. This may indicate that in states with higher electoral vote percentages, the popular vote might be more competitive or closer.

```
# Fit a linear regression model
regression_model <- lm(Winner_Vote_Percentage ~ EC_percentage, data = elections)

# Summarize the regression results
summary(regression_model)
```

```
##
## Call:
## lm(formula = Winner_Vote_Percentage ~ EC_percentage, data = elections)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -21.428  -5.869  -1.396   3.990  40.814
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    58.6047     0.3971 147.597  < 2e-16 ***
## EC_percentage  -0.5603     0.1512  -3.707 0.000219 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.213 on 1293 degrees of freedom
## Multiple R-squared:  0.01051,    Adjusted R-squared:  0.009748
## F-statistic: 13.74 on 1 and 1293 DF,  p-value: 0.000219
```

Model 2

Point two of the memo asked for "Your prospect of voting for the winning candidate as a function of your fraction of the electoral college votes of your state, plus your gap (absolute value of the percent difference between the 1st and 2nd national candidates)"

Computing the "gap"

The below code creates a variable for the second-place candidate vote count and uses it to calculate the gap, as described above.

```
elections <- elections |>
  rowwise() |>
  mutate(Second_Vote_Count = sort(c_across(all_of(candidate_columns)), decreasing = TRUE)[2],
         Gap = abs(Winner_Vote_Count - Second_Vote_Count) / Total.Vote * 100)
```

Regression for Model 2

The below code runs a regression of Winner_Vote_Percentage on EC_percentage plus the Gap.

```
regression_model_2 <- lm(Winner_Vote_Percentage ~ EC_percentage + Gap, data = elections)
summary(regression_model_2)
```

```
##
## Call:
## lm(formula = Winner_Vote_Percentage ~ EC_percentage + Gap, data = elections)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -17.2805  -0.4719   1.1970   1.7853   2.4466
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   47.515625   0.165100 287.798   <2e-16 ***
## EC_percentage  0.065955   0.048959   1.347    0.178
## Gap            0.514832   0.004861 105.912   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.962 on 1292 degrees of freedom
## Multiple R-squared:  0.8978, Adjusted R-squared:  0.8976
## F-statistic:  5675 on 2 and 1292 DF,  p-value: < 2.2e-16
```

In the regression model above, the coefficient on EC_percentage indicates that for each additional percentage point of electoral college votes, the Winner_Vote_Percentage increases by approximately 0.066 percentage points. However, this effect is not statistically significant (p-value = 0.178).

The coefficient on the Gap indicates that for each additional percentage point in the gap between the first and second national candidates, the Winner_Vote_Percentage increases by approximately 0.515 percentage points. This effect is highly statistically significant (p-value < 2e-16).

Regression for Model 3

Model 3 is quite similar to Model 2, but it includes an interaction term for EC percentage and the gap. It is computed below.

```r
regression_model_3 <- lm(Winner_Vote_Percentage ~ EC_percentage + Gap
                          + EC_percentage * Gap, data = elections)
summary(regression_model_3)
```

```
##
## Call:
## lm(formula = Winner_Vote_Percentage ~ EC_percentage + Gap + EC_percentage *
##     Gap, data = elections)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -17.2498  -0.5897   1.2025   1.7785   2.6123
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)       47.283945   0.202374 233.646   <2e-16 ***
## EC_percentage      0.183134   0.076884   2.382   0.0174 *
## Gap                0.528000   0.008248  64.019   <2e-16 ***
## EC_percentage:Gap -0.007318   0.003705  -1.975   0.0485 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.959 on 1291 degrees of freedom
## Multiple R-squared:  0.8981, Adjusted R-squared:  0.8979
## F-statistic:  3793 on 3 and 1291 DF,  p-value: < 2.2e-16
```

As seen above, the coefficient on EC_percentage of 0.183133, statistically significant with p-value = 0.0174, indicates that for each additional percentage point of electoral college votes, the Winner_Vote_Percentage increases by approximately 0.183 percentage points, holding Gap constant.

Meanwhile, on Gap, the coefficient indicates that for each additional percentage point in the gap between the first and second national candidates, the Winner_Vote_Percentage increases by approximately 0.528 percentage points, holding EC_percentage constant. This effect is highly statistically significant (p-value < 2e-16)

For the interaction term, finally, the result here indicates that the effect of EC_percentage on Winner_Vote_Percentage changes by -0.0073 percentage points for each additional percentage point in the gap between the first and second national candidates. This effect is statistically significant (p-value = 0.0485). In other words, the influence of the electoral college percentage is moderated by the vote gap.

These results seem to follow from what one might intuitively assume — where the gap is bigger, the percentage of the votes that go to the winning candidate will be bigger (seeing as there can only be 100 percentage points and most statewide races are at least slightly competitive between two candidates). The effect here is bigger than that of the electoral college vote percentage, of course, although we still saw some modest statistically significant impact.

I hope this is useful! Let me know about next steps. I also was doing these computations somewhat blindly, so I'd love to know more about the thought process behind including this Gap variable. If there is nothing else more pressing, I will work now on breaking these computations down by time period.

- Peter