Case Study 1: Your Informative Title Here

Miya Dang, Mia Tran

2025-03-05

Introduction

The U.S. presidential election in 2000 between George W. Bush and Al Gore was one of the closest in history, with its final outcome dependent on the results in the state of Florida. Ultimately, Bush secured victory by a margin of fewer than 400 votes. However, Democratic voters in Palm Beach County argued that a confusing "butterfly" ballot layout led them to mistakenly vote for Reform Party candidate Pat Buchanan instead of Gore. This claim was supported by evidence showing that Buchanan received an unusually high percentage of votes in the county, along with a significant number of discarded ballots where voters had marked two choices.

In this case study, we will model the relationship between Bush and Buchanan votes in 2000 across Florida (excluding Palm Beach County to assess its anomaly) to answer two questions:

- *Was the number of votes for Buchanan in Palm Beach County statistically unusual compared to other Florida counties?*
- *What is the estimated number of votes intended for Gore but cast for Buchanan in Palm Beach County?*

Data Description

The dataset contains the vote counts for Buchanan and Bush in 2000 across all 67 counties in Florida.

Summary Statistics:

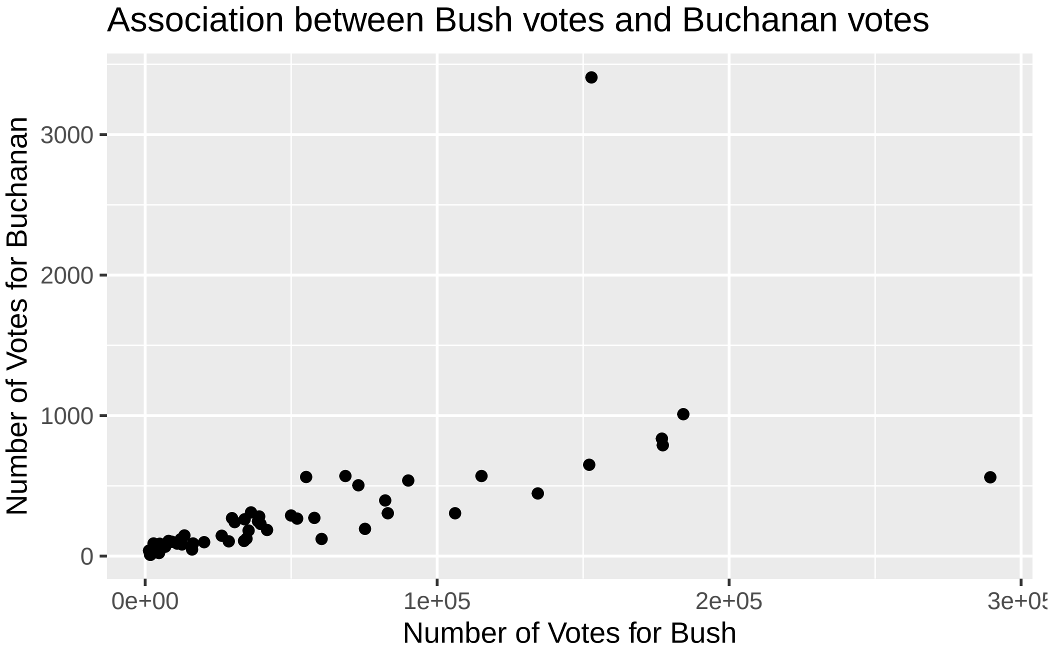
```
    Bush2000       Buchanan2000
 Min.   :  1316   Min.   :   9.0
 1st Qu.:  4746   1st Qu.:  46.5
 Median : 20196   Median : 114.0
 Mean   : 43356   Mean   : 258.5
 3rd Qu.: 56542   3rd Qu.: 285.5
 Max.   :289456   Max.   :3407.0
```

- Summary statistics indicate that, overall, Bush received significantly higher votes than Buchanan.

- Both variables have means much higher than their medians, indicating a strong right skew. This suggests that most counties had relatively low vote counts, while a small number had exceptionally high vote counts.
- The maximum number of Buchanan votes was 3,407, far exceeding the third quantile of 285.5. This vote count came from Palm Beach County, suggesting that it is an outlier.

Accompanying Graph:

The scatterplot below visualizes the relationship between Bush votes (x-axis) and Buchanan votes (y-axis) for all 67 counties:



There is a general positive trend, as counties with more Bush votes also tended to have more Buchanan votes. Palm Beach County (the mark on the top of the graph) is a notable outlier with unexpectedly high Buchanan votes (3,407).

Modeling Process and Results

Let *Buchanan* denote the votes for Buchanan in Palm Beach County and *Bush* denote the the votes for Bush in Palm Beach County. Our final linear regression model is:

$$E[log(Buchanan)|log(Bush)] = \beta_0 + \beta_1 (Bush) .$$

The prediction interval is $[365, 831]$. We are 95% confident that the votes for Buchanan when there are 152846 votes for Bush will fall between 365 votes and 831 votes.

Buchanan's real vote from Palm Beach County in 2000 election is 3407 votes. Compared to the prediction interval above, we can calculate that an estimate for the likely number of votes intended for Gore but cast for Buchanan in Palm Beach County falls between $[2576, 3042]$.

Summary and Conclusions

Key Findings

- The Buchanan vote count in Palm Beach County during the 2000 election was statistically unusual.
- An estimated 2,576 to 3,042 votes intended for Gore were likely miscast for Buchanan.
- These results suggest that the ballot design may have altered the election outcome. Given that Gore lost Florida by fewer than 400 votes, the potential miscast votes in Palm Beach County alone far exceed this margin.

Limitations

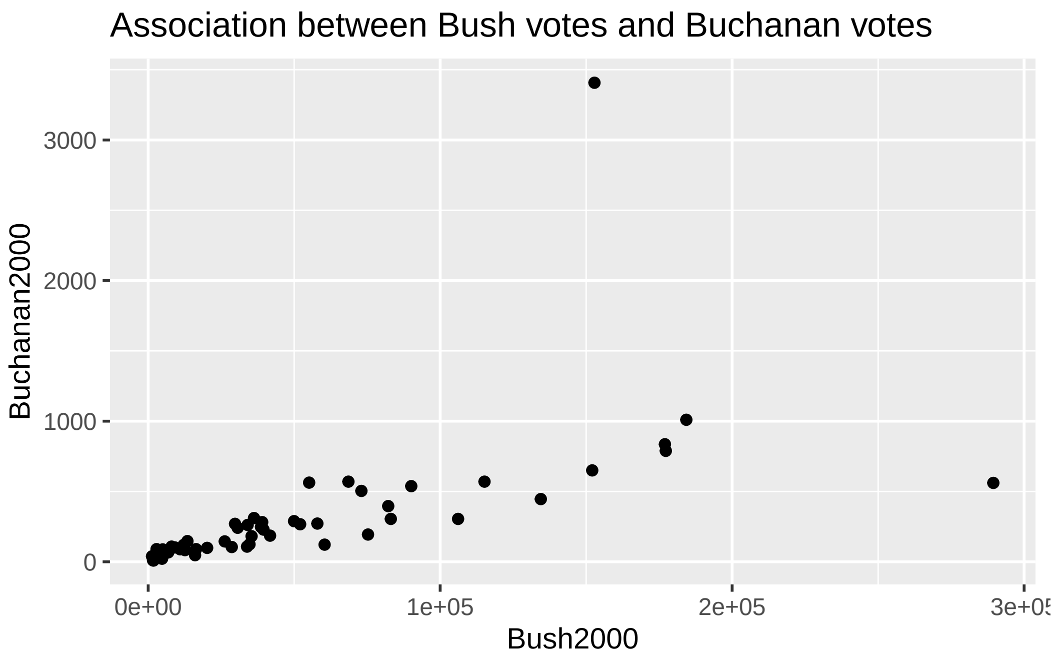
- The log-log transformation assumes a multiplicative relationship between Bush and Buchanan votes. While this improved linearity, residual diagnostics revealed mild heteroscedasticity and a slightly left-skewed residual distribution. These imply that the prediction interval may underestimate uncertainty, particularly for counties with high Bush vote count like Palm Beach.
- The model assumes that voter behavior is similar across all counties. However, many other factors could contribute to regional differences between voter behavior, such as candidate campaigns or other demographic factors.
- The analysis demonstrate association and not causation. Further investigation would be necessary to definitively prove that the butterfly ballot caused the excess Buchanan votes.

R Appendix

```r
# Loading necessary packages
library(tidyverse)
library(Sleuth2)
library(broom)
library(kableExtra)

# Loading the data for case study one
election <- Sleuth2::ex0825

# Summary statistics
summary(election %>% select(Bush2000, Buchanan2000))
```

```
    Bush2000       Buchanan2000
 Min.   :  1316   Min.   :   9.0
 1st Qu.:  4746   1st Qu.:  46.5
 Median : 20196   Median : 114.0
 Mean   : 43356   Mean   : 258.5
 3rd Qu.: 56542   3rd Qu.: 285.5
 Max.   :289456   Max.   :3407.0
```

```
# Creating a scatterplot for the relationship between Bush and Buchanan votes
election |>
  ggplot(aes(x = Bush2000,
             y = Buchanan2000)) +
  geom_point() +
  ggtitle("Association between Bush votes and Buchanan votes")
```



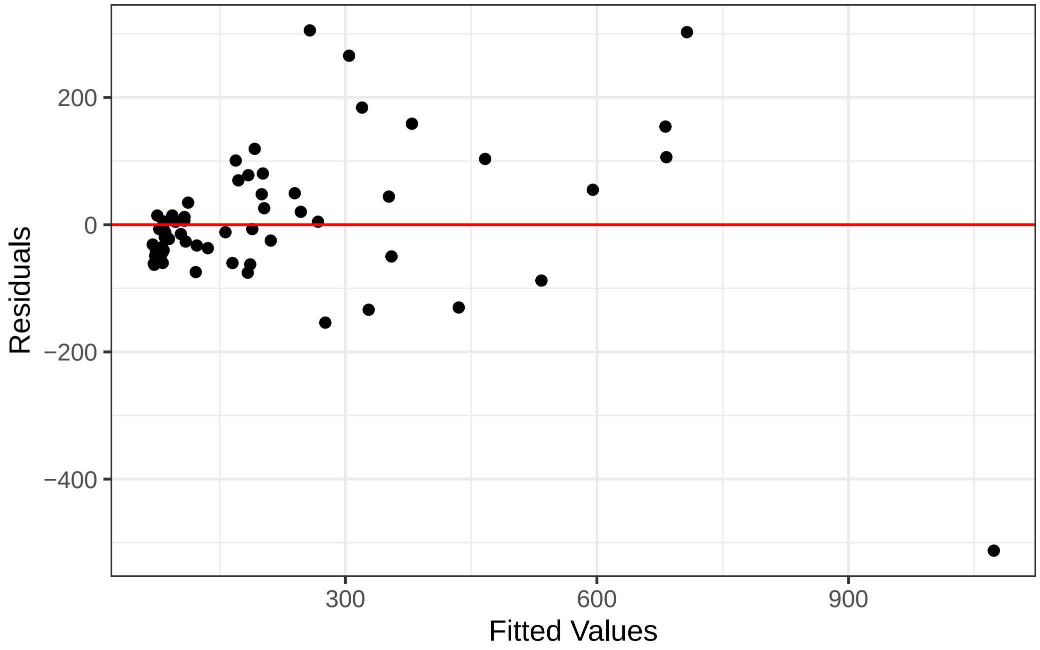
```
# Creating a second dataset with Palm Beach County excluded
election_wo_pb <- election |>
  filter(County != "Palm Beach")

# Fitting and summarizing the regression line for mean Buchanan votes
# as a function of Bush votes
election_lm <- lm(Buchanan2000 ~ Bush2000, data = election_wo_pb)
summary(election_lm)$coefficients
```

```
                Estimate   Std. Error   t value     Pr(>|t|)
(Intercept) 65.573496362 1.733043e+01  3.783721 3.427131e-04
Bush2000     0.003481898 2.500903e-04 13.922562 4.916245e-21
```

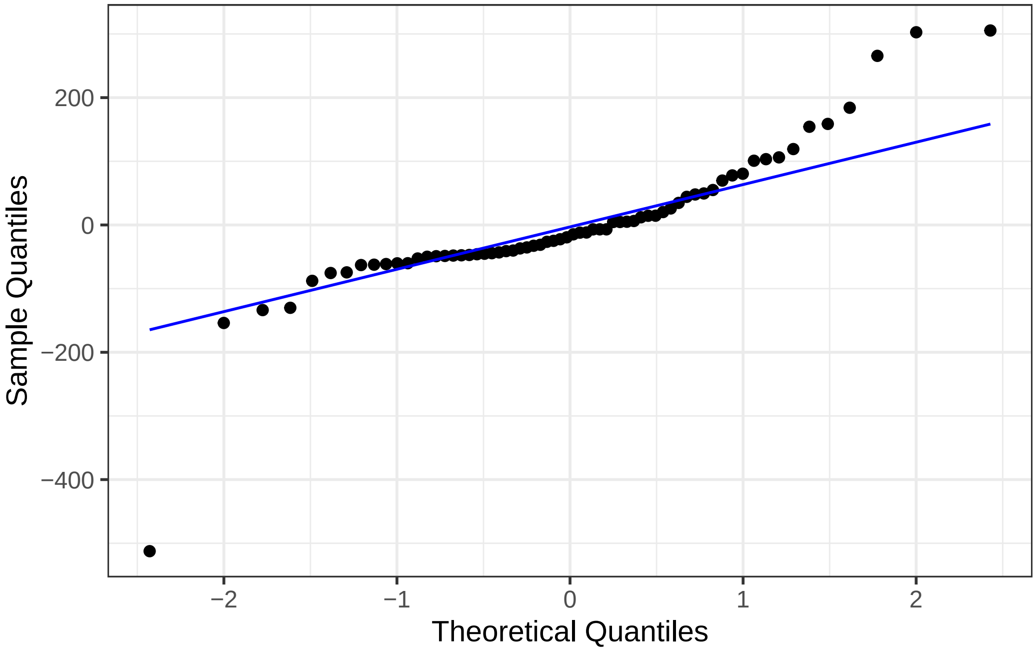
```
election_lm_table <- summary(election_lm)$coefficients

# Creating the table summarizing the estimated coefficients of the model
# and their corresponding standard errors
election_lm_table |> kbl(col.names = c("Estimate", "Std. Error",
                                       "t value", "Pr(>|t|)"),
                         align = "c",
                         booktabs = T,
                         linesep="",
                         digits = c(4, 4, 4, 4)) |>
  kable_classic(full_width = F, latex_options = c("HOLD_position"))
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	65.5735	17.3304	3.7837	3e-04
Bush2000	0.0035	0.0003	13.9226	0e+00

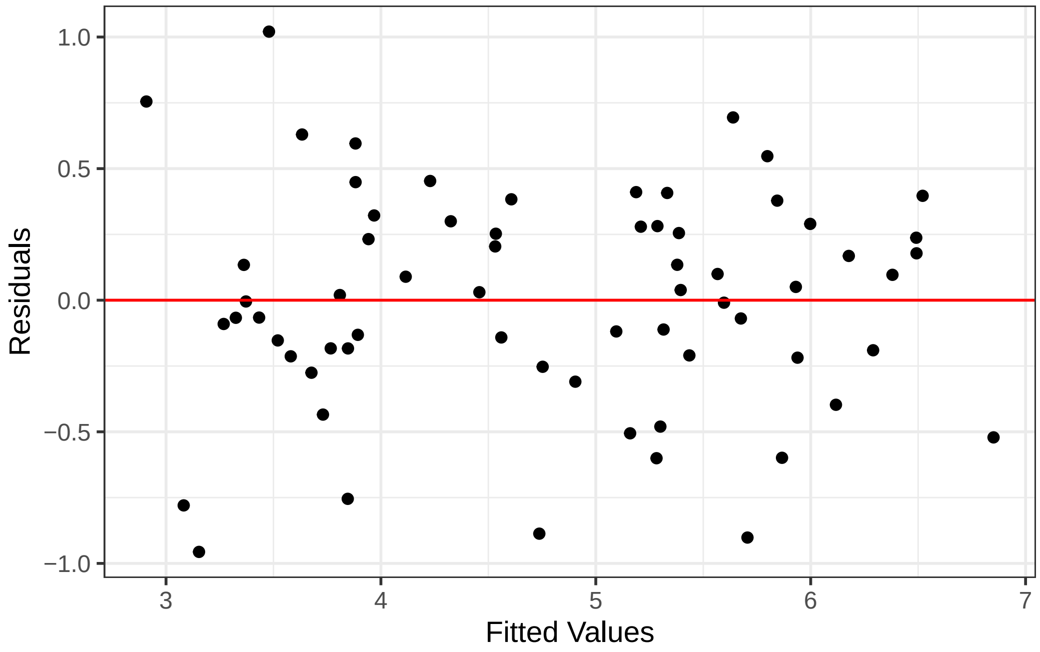
```
# Creating the residuals-fitted plot to check
# Linearity and Equal Variance for the election model
election_lm |>
  augment() |>
  ggplot(aes(x = .fitted, y = .resid)) +
  geom_point() +
  geom_hline(yintercept = 0, col = "red") +
  xlab("Fitted Values") +
  ylab("Residuals") +
  theme_bw()
```



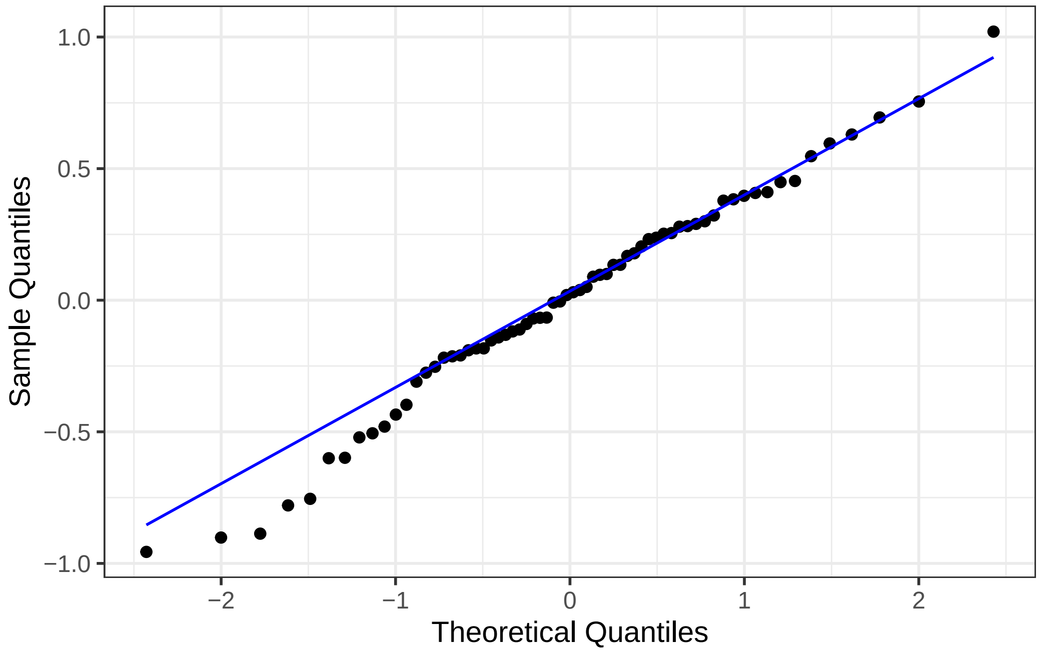
```
# Creating the Q-Q plot as check Normality for the election model
election_lm |>
  augment() |>
  ggplot(aes(sample = .resid)) +
  geom_qq() +
  geom_qq_line(col = "blue") +
  xlab("Theoretical Quantiles") +
  ylab("Sample Quantiles") +
  theme_bw()
```



```
#Transformation
transformed_election_lm <- lm(log(Buchanan2000) ~ log(Bush2000), election_wo_pb)

# Creating the residuals-fitted plot to check
# Linearity and Equal Variance for the transformed election model
transformed_election_lm |>
  augment() |>
  ggplot(aes(x = .fitted, y = .resid)) +
  geom_point() +
  geom_hline(yintercept = 0, col = "red") +
  xlab("Fitted Values") +
  ylab("Residuals") +
  theme_bw()
```



```
# Creating the Q-Q plot as check Normality for the transformed election model
transformed_election_lm |>
  augment() |>
  ggplot(aes(sample = .resid)) +
  geom_qq() +
  geom_qq_line(col = "blue") +
  xlab("Theoretical Quantiles") +
  ylab("Sample Quantiles") +
  theme_bw()
```



```
# Filter the original dataset to only Palm Beach Country
# to create prediction interval
election_pb <- election |> filter(County == "Palm Beach")

# Creating a 95% prediction interval for the number of
# predicted Buchanan votes in Palm Beach County
# after knowing the votes for Bush from Palm Beach County
new_election <- data.frame(Bush2000 = 152846)
election_lm |>
  augment(newdata = new_election,
          interval = "prediction",
          conf.level = 0.95)
```

```
# A tibble: 1 x 4
  Bush2000 .fitted .lower .upper
     <dbl>   <dbl>  <dbl>  <dbl>
1   152846    598.   365.   831.
```