

Unit 9 Q3.5

A. Lebedev

2022-10-27

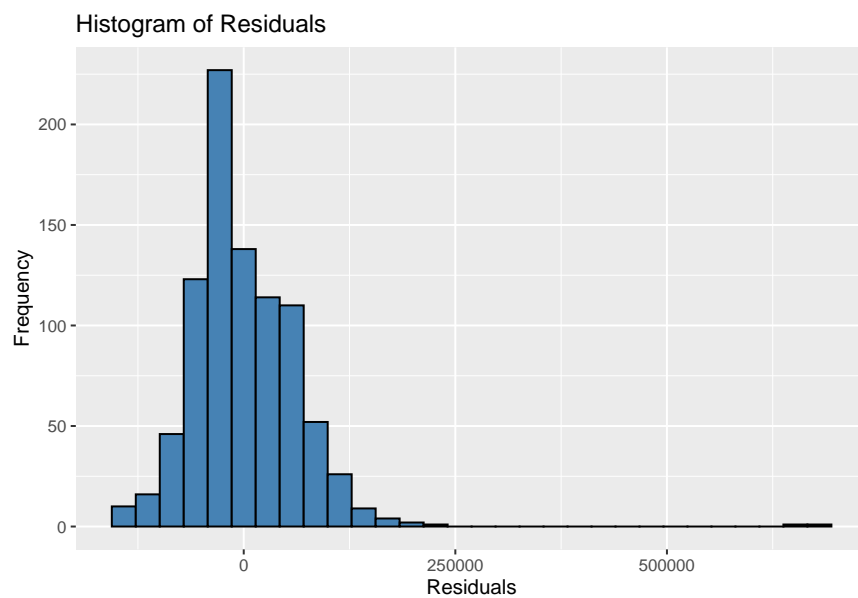
1. Load data

2. Filter data to remove NAs and generate the model

```
joint_df <- filter(joint_df, general_votes >=0)
my_model <- lm(formula = general_votes ~ ttl_disb + candidate_party, data = joint_df)
```

3. Evaluate residuals

```
ggplot(data = joint_df) +
  aes(x = my_model$residuals) +
  geom_histogram(fill = 'steelblue', color = 'black') +
  labs(title = 'Histogram of Residuals', x = 'Residuals', y = 'Frequency')
```



Two main assumptions for the large sample OLS model is i.i.d. sampling and the number of observations sufficient for the statistics to exhibit asymptotic behavior due to Central Limit Theorem. I.I.D. assumption is likely to be satisfied at least loosely, because the amount of money raised and the number of votes of one candidate. The residuals of the model plotted above are distributed almost normally, with a few rare exceptions. Additionally, there are almost a thousand observations. That gives us hope that large number of observations is sufficient for our statistics to be distributed almost normally.