

3.6 Evaluate large sample assumptions

Unit 9 Coding Homework - W203 Section 8

Chase Madson

Contents

```
##
## Call:
## lm(formula = general_votes ~ ttl_disb + candidate_party, data = .)
##
## Coefficients:
##              (Intercept)              ttl_disb
##              2.698e+04              1.326e-02
## candidate_partyDemocrat candidate_partyRepublican
##              1.131e+05              1.196e+05
```

For OLS in a large sample ($n > 100$), we need to verify a few assumptions.

1. The data must be IID

- Each observation is drawn from political candidates running in a 2016 election. We may see violations of IID in this circumstance given that there is strong regional clustering of a candidate's success based on their party and geographic location. If there are two candidates from the same state, there is likely some covariance between their political parties. Therefore we would consider this assumption likely violated.

2. A unique BLP exists

- The existence of a BLP may be in question since the predictor `ttl_disb`, which we saw before in **Q3.2** was severely right-skewed. This causes the covariance between this predictor and the other variables to be infinite. Therefore we would consider this assumption likely violated, and log-transformation may be an appropriate remedy. However, if there is a lot of skew in the data, then the n must be large enough to overcome it. Here we have $n > 1000$ in addition to a strong skew, which may be enough.
- The uniqueness of the BLP appears to be valid - it does not appear reasonable to write one predictor as a linear combination of another. The budget of a campaign may have some relationship to the party (especially for third-party candidates), but this is not extreme enough to invite perfect collinearity.