

Project Part 5

Data Description

Research Question

Does the percent of a state population who lives in urban environments affect that state's election results?

General Context and Sources

The urbanization data set comes from Iowa State University's Iowa Community Indicators Program (1). These data are US Decennial Census data of the percent of the population living in urban areas during the years 1980-2010 in all 50 US states and Washington, DC. These data are appropriate for answering the research question because they are the defined predictor.

The Senate election results data set are a population that includes the US states and which comes from MIT's Election Data and Science Lab (2). The data describes outcomes of US Senate elections in all 50 US states, including, notably, year, candidate, party, number of votes received, and number of votes cast. These data are appropriate for answering the research question because they are the defined response.

I chose to use only data from 2010 because data from the same state in different years is generally dependent.

Spearman Correlation Test

Test Selection

I chose the Spearman correlation test because I want to test the correlation between two paired continuous variables – voting affiliation and urban percent of the population, paired

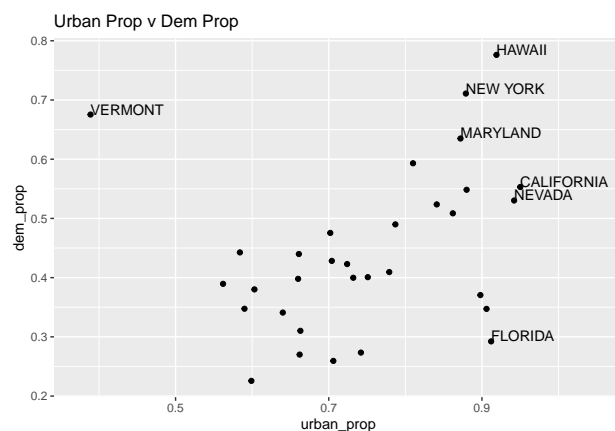
by state – and because my variables are not multivariate normally distributed. Thus, a non-parametric method (without a normality or similar distribution assumption) is the correct choice. (3)

Assumptions

The assumptions of the Spearman correlation test are that the data are 1) ordinal, interval, or ratio, 2) paired, and 3) monotonically related. (3)

- 1) The data are ratio.
- 2) The data are paired by state.
- 3)

```
data.2010 %>% ggplot(aes(x=urban_prop, y=dem_prop)) +
  geom_point() + labs(title="Urban Prop v Dem Prop") +
  geom_text(aes(label=ifelse(
    urban_prop > quantile(urban_prop, 0.9) | dem_prop > quantile(dem_prop, 0.9),
    as.character(state),
    ""
  )), hjust=0, vjust=0) + xlim(0.375, 1.05)
```



Vermont is a major outlier but I can leave it because Spearman's Correlation is not very sensitive to outliers.

The data appear to be approximately monotonically related.

Hypotheses

$H_0 : \rho = 0$ There is no monotonic correlation between the proportion of the population that lives in urban areas and the proportion of the population that vote Democrat. (3)

$H_A : \rho \neq 0$ There is a monotonic correlation between the proportion of the population that lives in urban areas and the proportion of the population that vote Democrat. (3)

Execution

```
result <- cor.test(data.2010$urban_prop, data.2010$dem_prop, method = "spearman") # (4)
result

##
## Spearman's rank correlation rho
##
## data: data.2010$urban_prop and data.2010$dem_prop
## S = 3388, p-value = 0.03316
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## 0.3790323
```

Result

Since the p-value of 0.033 is less than the alpha of 0.05, we reject the null hypothesis.

Conclusions

In answer to my research question, there does appear to be a monotonic relationship between the percent of the population that lives in urban areas and the percent of the population that votes Democrat.

Generalization

In future US Senate elections, the percent urban population could be a good metric to help predict the outcome of the election. It could be a useful talking point for election night news anchors.

References

1. Iowa State University Iowa Community Indicators Program <https://www.icip.iastate.edu/tables/population/urban-pct-states>
2. MIT Election Data and Science Lab <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/PEJ5QU>
3. Spearman's Rank-Order Correlation (cont. . .) <https://statistics.laerd.com/statistical-guides/spearmans-rank-order-correlation-statistical-guide-2.php>
4. Correlation Test Between Two Variables in R <http://www.sthda.com/english/wiki/correlation-test-between-two-variables-in-r#spearman-rank-correlation-coefficient>