

Categorical Data Applications

YOUR NAME

09 June, 2020

Exercises

Create an Rmd file for the work including headers, file creation data, and explanation of your work. Make sure your plots have a title and the axes are labeled.

1. Views on immigration

910 randomly sampled registered voters from Tampa, FL were asked if they thought workers who have illegally entered the US should be (i) allowed to keep their jobs and apply for US citizenship, (ii) allowed to keep their jobs as temporary guest workers but not allowed to apply for US citizenship, or (iii) lose their jobs and have to leave the country.

The data is in the `openintro` package in the `immigration` data object.

- a. How many levels of *political* are there?
- b. Create a table using `tally`.
- c. What percent of these Tampa, FL voters identify themselves as conservatives?
- d. What percent of these Tampa, FL voters are in favor of the citizenship option?
- e. What percent of these Tampa, FL voters identify themselves as conservatives and are in favor of the citizenship option?
- f. What percent of these Tampa, FL voters who identify themselves as conservatives are also in favor of the citizenship option? What percent of moderates and liberal share this view?
- g. Create a stacked bar chart.
- h. Using your plot, do political ideology and views on immigration appear to be independent? Explain your reasoning.

2. **Views on the DREAM Act** The same survey from Exercise 1 also asked respondents if they support the DREAM Act, a proposed law which would provide a path to citizenship for people brought illegally to the US as children.

The data is in the `openintro` package in the `dream` data object.

- a. Create a `mosiac` plot.
- b. Based on the mosaic plot, are views on the DREAM Act and political ideology independent?

3. Heart transplants

The Stanford University Heart Transplant Study was conducted to determine whether an experimental heart transplant program increased lifespan. Each patient entering the program was designated an official heart transplant candidate, meaning that he was gravely ill and would most likely benefit from a new heart. Some patients got a transplant and some did not. The variable *transplant* indicates which group the patients were in; patients in the treatment group got a transplant and those in the control group did not. Another variable called *survived* was used to indicate whether or not the patient was alive at the end of the study.

- a. Create a **mosiac** plot.
- b. Based on the mosaic plot, is survival independent of whether or not the patient got a transplant? Explain your reasoning.
- c. Using *survtime* create side-by-side boxplots.
- d. What do the box plots suggest about the efficacy (effectiveness) of transplants?