STAT 2300 - Homework 1

Data Exercises

1. `starwars`

These questions use the data set `starwars` from the `dplyr` library. It gives biographical information about the fictional characters in the Star Wars movies.

a. How many variables are in the `starwars` data set?
b. How many observations are in the `starwars` data set?
c. How tall is Chewbacca? Are there any characters taller than Chewbacca?
d. How many Human characters are there? Of these, how many are male, and how many are female?
e. What are the two most common `homeworld`s? How many characters came from each?
f. Body mass index (BMI) is mass divided by the square of height, and is a measure of obesity. The highest BMI in Star Wars is Jabba the Hutt (by a lot). Which *Human* character in Star Wars has the highest BMI? What is their BMI?

2. Lasers and airplanes

These questions use data on lasers pointed at airplanes in 2021, as reported by the FAA. This data is available on our course web page at https://turtlegraphics.org/stat2300/data/lasers21.csv

a. How many laser incidents were reported in 2021?
b. Find the top five states with the most laser incidents.
c. Make a histogram of the altitudes (in feet) of laser incidents.
d. What was the highest altitude reported incident? What do you think about that?
e. Convert the `Incident.Date` to an actual date using the `dmy` function from the `lubridate` package (which you may need to install). Count the number of incidents per date, then make a plot showing number of incidents by date.

Textbook Exercises - Chapter 1

Read Chapter 1.

Conceptual exercises

These have answers at the end of the chapter, so don't write up solutions.

Exercises # 1, 2, 3, 4, 10.

Computational exercise # 17:

Seven students volunteered for a comparison of study guides for an advanced course in math-ematics. They were randomly assigned, four to study guide A and three to study guide B. All were instructed to study independently. Following a two-day study period, all students were given an examination about the material covered by the guides, with the following results:

Study Guide A scores: 68, 77, 82, 85

Study Guide B scores: 53, 64, 71

Perform a randomization test by listing all possible ways that these students could have been randomized to two groups. There are 35 ways. For each outcome, calculate the difference between sample averages. Finally, calculate the two-sided p-value for the observed outcome.

Rather than list all 35 ways, use resampling to make 10000 random reassignments to the groups

Data problem # 26:

Environmental Voting of Democrats and Republicans in the U.S. House of Representatives. Each year, the League of Conservation Voters (LCV) identifies legislative votes taken in each house of the U.S. Congress-votes that are highly influential in establishing policy and action on environmental problems. The LCV then publishes whether each member of Congress cast a pro-environment or an anti-environment vote. Display 1.15 shows these votes during the years 2005, 2006, and 2007 for members of the House of Representatives. Evaluate the evidence supporting party differences in the percentage of pro-environment votes. Write a brief report of your conclusion, including a graphical display and summary statistics.