Math 58 / 58B - Introduction to (Bio)Statistics

your name here

due Jan 28, 2020

Homework 1

Important Note:

You should work to turn in assignments that are clear, communicative, and concise. Do **not** print pages and pages of output (your HW score will be marked down!). Additionally, you should remove these exact sentences and the information about HW scoring below.

Click on the *Knit* icon at the top of R Studio to run the R code and create a pdf document simultaneously.

Assignment Summary (Goals)

- Practice using R.
- R to work through the hypothesis testing ideas we've used in class.
- Do *not* expect the assignment to be obvious given what we've covered in class. (Most of the R code is done for you, try your best to understand what is going on.)
- use the `infer` syntax. See help sheets here: https://infer-dev.netlify.com/index.html

1. (Investigation 1.5: Kissing the Right Way, ISCAM)

   Most people are right-handed and even the right eye is dominant for most people. Researchers have long believed that late-stage human embryos tend to turn their heads to the right. German biopsychologist Onur Gunturkun (Nature 2003) conjectured that this tendency to turn to the right manifests itself in other ways as well, so he studied kissing couples to see if both people tended to lean to their right more often than to their left (and if so, how strong the tendency is). He and his researchers observed couples from age 13 to 70 in public places such as airports, train stations, beaches, and parks in the United States, Germany, and Turkey. They were careful not to include couples who were holding objects such as luggage that might have affected which direction they turned. We will assume these couples are representative of the overall decision making process when kissing.

   In total, 124 kissing pairs were observed with 80 couples leaning right.

(a) What are the

- observational units?
- the variable? is it categorical or quantitative?
- the statistic?
- the parameter?

(b) Do the data from the kissing study provide convincing evidence that the probability of leaning right is **smaller** than 0.8?

- State the hypotheses
- report the p-value [describing how you determined it]
- clarifying what it is in words / terms of the probability of ...

- interpret the strength of evidence against the null hypothesis).

To find the p-value, use the `infer` syntax. I've written much of the `infer` code below. You need to do two thing to the code: (1) toggle to `eval=TRUE` and (2) fill in all the blank spaces below

```r
# you may need to install infer
install.packages("infer") # run the intall only once in the console, then delete only this line
library(infer)

# to control the randomness
set.seed(47)

# first create a data frame with the kissing data
kissing <- data.frame(side = c(rep("right", ___ ), rep("left", ___ )))

# then find the proportion who kiss right
(p_obs_right <- kissing %>%
    specify(response = ___ , ___ = "right") %>%
    calculate(stat = "prop") )

# now apply the infer framework to get the p-value
null_kiss <- ___  %>%
  specify(response = ___ , success = "___ ") %>%
  hypothesize(null = "point", p = .8) %>%
  generate(reps = 1000, type = "simulate") %>%
  calculate(stat = "prop")

# visualize the null sampling distribution
visualize(___ ) +
  shade_p_value(obs_stat = ___ , direction = "___ ")

# calculate the actual p-value
null_kiss %>%
  get_p_value(obs_stat = ___ , direction = "___ ")
```

(c) Do the data from the kissing study provide convincing evidence that the probability of leaning right **differs** from 0.8? [See page 75-76 for a discussion on two-sided tests. Also, never be afraid to use the help files! Try: `?shade_p_value` and look at the options for direction.]

(Use the code and instructions from part (a).)

(d) Reconsider part (b): how would your results have differed if you had chosen "kiss left" instead of "kiss right"? Explain in words.

(e) Reconsider part (c): how would your results have differed if you had the same proportional data but half of the observations (that is, data were 40 right kissers out of 62). Repeat the analysis in (c).

(Feel free to delete everything from here out, but you might find the information below on grading interesting.)

Resources for learning R and working in RStudio

That was a short introduction to R and RStudio, but we will provide you with more functions and a more complete sense of the language as the course progresses.

In this course we will be using R packages called `dplyr` for data wrangling and `ggplot2` for data visualization. If you are googling for R code, make sure to also include these package names in your search query. For example, instead of googling "scatterplot in R", google "scatterplot in R with ggplot2".

The following cheatsheets may come in handy throughout the semester:

- RMarkdown cheatsheet
- Data wrangling cheatsheet
- Data visualization cheatsheet

Chester Ismay has put together a resource for new users of R, RStudio, and R Markdown here. It includes examples showing working with R Markdown files in RStudio recorded as GIFs.

Note that some of the code on these cheatsheets may be too advanced for this course, however the majority of the information will become useful throughout the semester.

HW & Lab assignments will be graded out of 5 points, which are based on a combination of accuracy and effort. Below are rough guidelines for grading.

Score & Description

5 points: All problems completed with detailed solutions provided and 75% or more of the problems are fully correct.

4 points: All problems completed with detailed solutions and 50-75% correct; OR close to all problems completed and 75%-100% correct. An assignment will earn a 4 if there is superfluous information printed out on the assignment.

3 points: Close to all problems completed with less than 75% correct

2 points: More than half but fewer than all problems completed and > 75% correct

1 point: More than half but fewer than all problems completed and < 75% correct; OR less than half of problems completed

0 points: No work submitted, OR half or less than half of the problems submitted and without any detail/work shown to explain the solutions.

General notes on homework assignments (also see syllabus for policies and suggestions):

- please be neat and organized, this will help me, the grader, and you (in the future) to follow your work.

- be sure to include your name on the assignment, and staple the pages together *prior* to class

- please include at least the number of the problem, or a summary of this question (this will also be helpful to you in the future to prepare for exams).

- it is strongly recommended that you write out the questions as soon as you get the assignment. This will help you to start thinking how to solve them!

- for R problems, it is required to use R Markdown

- please do not print errors, messages, warnings, or anything else that makes your homework unwieldy. You will be graded down for superfluous printouts.

- in case of questions, or if you get stuck please don't hesitate to email me (though I'm much less sympathetic to such questions if I receive emails within 24 hours of the due date for the assignment).