

Math 58 (only) - Introduction to Statistics

your name here

due Feb 27, 2020

Homework 5

Assignment Summary (Goals)

- Binomial probabilities
 - review of p-values (conclusions) and power
1. Consider any binary situation we've discussed (including one of the examples below). Provide one advantage for using binomial probabilities, and one advantage for using normal probabilities.
 2. **Rock-Paper-Scissors** (Problem 1.7, ISCAM).

For informal sports events, players often play “rock-paper-scissors” to decide who serves first or who is the home team etc. Two players simultaneously show one of the three objects. The player showing the dominant object (e.g., rock beats scissors) wins. The optimal strategy is to alternate among the three objects. Play the game rock/paper/scissors against the computer using the website: <https://archive.nytimes.com/www.nytimes.com/interactive/science/rock-paper-scissors.html> (you may have to unblock flash, or you can find a rock-paper-scissors game elsewhere, any version is fine). Select the novice version of the computer to play against. Play for at least 30 rounds, but keep going for as long as you'd like. **Keep track of which option you choose (rock or paper or scissors)** for every round that you play (the computer will record this for you but that information will soon scroll off the screen, so make your own notes). Try to recreate how you would play against a person and don't view your prior results when making your next selection.

Seriously, go run the game at least 30 times before you do the rest of the assignment. Keep track of which option you play, not whether or not you beat the computer.

- (a) Identify the observational units in this study.

Do you have a sample of size at least 30 yet?

- (b) Identify the variable of interest.

How about now, what is your data from the computer simulation?

An article published in College Mathematics Journal (Eyler, Shalla, Doumaux, & McDevitt, 2009) found that players tend to not prefer scissors, choosing it less than 1/3 of the time. We will investigate whether your data suggest that you tend to choose scissors less than one-third of the time.

- (c) Calculate the statistic in this study and create a bar graph (use `ggplot`) of your results (all three options). Are your results in the direction conjectured by these researchers (choosing scissors less than 1/3 of the time)?

These are my results. I chose scissors 8 times, paper 10 times, and rock 12 times. You should input your own results. They don't need to add up to 30, but they should add up to something 30 or greater.

```
rochambo <- data.frame(choice = rep(c("scissors", "paper", "rock"),
                                   times=c(8,10,12)))
```

- (d) Define the parameter of interest in this study.

- (e) State appropriate null and alternative hypotheses about this parameter according to the theory suggested in the CMJ article.

- (f) Explain how you could use an ordinary six-sided die to simulate a distribution of $\hat{p}$ under this null hypothesis. Be sure to indicate what each possible outcome of the die (1, 2, 3, 4, 5, 6) would represent.
- (g) Based on your sample results (each student will be different), are you convinced that you choose scissors less than one-third of the time in the long run? Clearly explain your reasoning.

Use the built in R function (in the `library(mosaic)` package) called `xpbinom`.

- (h) Just for fun... let's do a little bit of this by hand (in case that type of question comes up on the exam). Assume that you only choose scissors twice out of 30 trials. What is your p-value? (Perform the calculation by hand. Feel free to use R as a calculator, but don't use a binomial function. Note that in R, `*` means multiply and `^` means "to the power of".)

Here, your computations should be done in R but as if you are using R as a calculator. Feel free to *check* your answers with `xpbinom`.

3. Campus Legend, Part I (Problem 1.8, ISCAM).

A statistics class at Cal Poly collected data on a well-known campus legend. Each student was asked to specify one of the four tires to answer in a situation where you have to make up which tire had recently been flat on your car. The prior conjecture is that a higher number than would be expected due to chance alone would pick the right front tire. In this class, 24 of 54 students in class chose the right front tire (a tire identified in advance as being one that people tend to pick out of the four). You will conduct a test of whether these data provide evidence that Cal Poly students tend to choose the right front tire more often than would be expected if the four tire choices were equally likely. [Complete data: 15 = left front, 8 = left rear, 24 = right front, 7 = right rear]

- (a) Identify the observational units and variable in this study. Also classify the variable as categorical or quantitative. If the variable is categorical, also indicate whether it is binary.
- (b) State the appropriate null and alternative hypothesis, in symbols and in words.
- (c) Use `ggplot` to produce a bar graph of the student responses. Submit this graph, and comment on what it reveals.
- (d) Use `xpbinom` to determine the (exact binomial) p-value for the test of your hypotheses in (b). (Look very carefully at the graph, you may need to adjust the input to `xpbinom`.)
- (e) Write a sentence describing what this p-value is the probability of.
- (f) Write a couple of sentences summarizing the conclusion that you would draw from this analysis and also explaining the reasoning process that underlies your conclusion.
- (g) Suppose that another statistics class conducts this same study in their own class, which has exactly half as many students. Suppose further that this class obtains the same proportion of students choosing the right front tire. Determine the exact p-value in this case. Describe how the p-value and your conclusion would differ for this class of 27 students compared to the first class of 54 students, and comment on why this makes intuitive sense.

4. Campus Legend, Part II (Problem 1.9, ISCAM).

Reconsider the previous exercise, where you tested whether sample data suggest that more than 25% of a population would answer "right front" when asked to name a tire that had gone flat. Suppose that you read of a study in which 30% of a random sample answered "right front."

- (a) What further information would you require to assess whether this sample result constitutes strong evidence that more than 25% of the population would answer "right front"? Also explain why this information is needed.
- (b) Determine the p-value of the test for the following sample sizes, in each case supposing that the sample proportion answering "right front" is 0.3: $n = 10, 50, 250, 500$. (Feel free to use technology, but explain what you ask the technology to do in each case.)

What do the probability calculations tell you about the relationship with the sample size?