Problem set 7

S520

Upload your answers through the Assignments tab on Canvas by 11:59 pm, Friday 1st November. Draw graphs in R and include code.

1. (5 points. From the Spring 2017 final.) I downloaded data on the number of citations for a random sample of 1000 journal articles published in 1981. (The data is from the ISI Citation Indexes.) I ran some analysis on the data in R, and produced the following output:

```
> citations = scan("citations.txt")
Read 1000 items
> summary(citations)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   0.00    0.00    1.00    9.06    7.25  300.00
> var(citations)
[1] 565.2476
> # Number of articles with no citations
> sum(citations == 0)
[1] 460
```

   (a) Is the distribution of the number of citations (i) exactly normal, (ii) approximately normal, or (iii) not close to normal? How do you know?

   (b) Find an approximate 95% confidence interval for the mean number of citations.

   (c) Find an approximate 95% confidence interval for the *proportion* of journal articles with no citations.

2. (5 points.) Trosset exercise 9.6 exercise 9. Note: The length of a confidence interval is the distance from its upper bound to its lower bound.

3. (5 points.) In a May 2019 Gallup poll,[1] 63% of a sample of 1009 U.S. adults supported same-sex marriage.

   (a) Treating the data as a simple random sample, find a 95% confidence interval for the percentage of all U.S. adults who support same-sex marriage.

   (b) Suppose we wanted to have a 95% confidence interval for the percentage of all U.S. adults who support same-sex marriage with total length 2% (i.e. 0.02.) How large a simple random sample would we need?

[1] `https://news.gallup.com/poll/257705/support-gay-marriage-stable.aspx` . Note: The margin of error Gallup states is different from what we would calculate, because they adjust for the fact that they are not taking a true simple random sample.

4. (10 points.) Every semester, I give my undergraduate statistics classes a test for psychic powers. (I don't give to the test to my graduate statistics classes, because grad classes are not allowed to be fun.) The test consists of guessing which side of a screen a picture will appear on: left or right. In one trial of the test, a student has to guess "Left" or "Right"; then R's random number generator will randomly choose one side of the screen to display a picture of a star. The student repeats the process for a total of 20 trials.

   (a) In words, the null hypothesis for a particular student is that they don't have psychic powers and they're just randomly guessing. The alternative hypothesis is that they do have psychic powers and in the long run, they can do better than randomly guessing.

   Let p be the probability the student guesses correctly on any particular trial. Write down mathematical null and alternative hypotheses in terms of p.

   (b) Suppose the null hypothesis is true for a particular student. Let Y be the number of times the students guesses correctly. Then if the null is true, Y has a Binomial$(20, 0.5)$ distribution.

   Which of the following is true?

      i. If the student guesses 13 right out of 20, the significance probability (P-value) is the probability under the null of guessing 13 or more out of 20.

      ii. If the student guesses 13 right out of 20, the significance probability (P-value) is the probability under the null of guessing 13 or fewer out of 20.

   Hint: Remember that the smaller the P-value, the stronger the evidence for the *alternative* hypothesis.

   (c) Suppose a student guesses 13 right out of 20. What is the P-value? Do you think 13 out of 20 is intriguing evidence that the student has psychic powers?

   (d) Now suppose a student guesses 19 right out of 20. What is the P-value? Do you think 19 out of 20 is intriguing evidence that the student has psychic powers?

5. (5 points.) Use the binomial distribution to find the P-value for each of the following situations.

   (a) I roll a six-sided die repeatedly. Let p be the probability of getting a six. I wish to test $H_0 : p = 1/6$ vs. $H_1 : p < 1/6$. I end up getting 15 sixes in 100 rolls.

   (b) Let p be the probability of getting a sum of twenty-one in two cards at blackjack. I wish to test $H_0 : p = 32/663$ vs. $H_1 : p < 32/663$. I deal two cards 1,000 times (reshuffling each time) and get twenty-one 59 times.

   (c) I want to know if there's bias in coin tosses before Test cricket matches. My null hypothesis is that the home team and the away team both have a 50-50 chance of winning the tosses. I decide to perform a two-tailed test. I find that the home team has won 1150 times and the away team has won 1065 times.

   (d) I wonder if my intro stats students are just guessing on their multichoice tests, or if they really know anything. I ask a total of 720 questions, each with four options. My null hypothesis is that all the students are randomly guessing on all questions; I want a small P-value to be evidence that they're doing better than that. I find they end up getting 237 out of 720 right.