HW 9 – Math 58B

your name here

due Thursday, April 6, 2023

Assignment Summary (Goals)

practice creating & interpreting:

- t-tests for the population mean
- t-intervals for the population mean (Conf Int)
- t-intervals for predicting single response (Pred Int)

Q1. Learning Community Q Describe one thing you learned from someone in your learning community this week (it could be: content, logistical help, background material, R information, etc.) 1-3 sentences.

Q2. Guess the Instructor's Age[1] The file `AgeGuesses.txt` contains guesses of an instructor's age by her current students (not *your* instructor, but the instructor of the ISCAM textbook). Let μ represent the average guess of her age by all current students at the university and suppose the sample constitutes a representative sample of all students at this school on this issue.

```
AgeGuesses = read.table("http://www.rossmanchance.com/iscam2/data/AgeGuesses.txt",
                        sep="\t", header=TRUE, na.strings="*")
head(AgeGuesses)
```

```
##   guesses
## 1      26
## 2      28
## 3      30
## 4      31
## 5      33
## 6      35
```

(a) Produce numerical and graphical summaries of the distribution and describe what you learn (in context).

(b) Use a histogram to decide whether the data has strong deviations from the pattern of a normal distribution.

(c) Use `t.test` to determine a 90% one-sample t-interval for these data. Include your output and comment on the validity of this procedure. Provide a one-sentence interpretation of this interval.

The R code will look like this (note that the R command provides a t-test as well as a CI, but the null hypothesis R uses is $H_0 : \mu = 0$ which is a ridiculous null hypothesis for the research question at hand):

```
___ %>%
  select(___) %>%
  t.test(conf.level = ___)
```

(d) Count how many of the class guesses are inside the 90% confidence interval. Compute the percentage of the class guesses that are inside the interval. Is this close to 90%? Should it be?

[1] From ISCAM, HW 2.5

The R code to do this problem might use the `filter()` command. Something like this:

```
___ %>%
  filter(___ > ___ , ___ < ___) %>%
  summarise(count = ___)
```

(e) Calculate and interpret a 90% prediction interval. Include the details of your calculation and comment on the validity of this procedure. How does the prediction interval compare (midpoint, length) to the confidence interval? How many of the class guesses are inside the interval?

You'll need to create the prediction interval by hand (i.e., using R as a calculator). Use the prediction interval formula in the notes http://st47s.com/Math58/Notes/inference-for-numerical-data.html#prediction-interval-for-individual-healthy-body-temperatures.

Q3. What to use when?? Much of the material covered in the class so far has followed a consistent structure. That is, interest is in parameters (numerical descriptions of the population) – sometimes testing the values of the parameters, sometimes estimating (with confidence intervals) the values of the parameters. Taking a step back to reflect on how the methods we've covered are similar and how they are different will help you understand all of the material.

Below you will find a series of examples[2]. For each example,

Assigned:

- From the following list of examples, which procedure would you follow to address the research question?

Optional to complete (but excellent practice!):

- What units are being measured (or observed)?
- What variables are being measured? Are they categorical? binary? quantitative?
- Which is the response and which is the explanatory variable (if appropriate)?
- Is it possible that the data were randomly sampled? If no, why not? If yes, what would you then be able to conclude?
- Is it possible that the explanatory variable was randomly assigned? If not, why not? If yes, what would you then be able to conclude?
- If you choose a hypothesis test, state the null and alternative hypotheses, and define the relevant population parameter(s).
- If you choose a confidence interval, define the relevant population parameter(s).
- Check any needed technical conditions applicable to the test.

Different possible	methods
(1) One sample bootstrap interval for a proportion	(2) One sample z-interval for a proportion
(3) One sample z-test of a proportion	(4) One sample t-test of a mean
(5) Prediction interval for individual observation	(6) One sample t-interval for a mean
(7) z-interval for OR / RR	(8) Bootstrap interval for OR/ RR
(9) z-test for OR / RR	(10) Randomization test for OR/ RR
(11) Two sample bootstrap interval for proportions	(12) Two sample z-interval for proportions
(13) Two sample t-test of means	(14) Two sample t-interval for means
(15) Two sample randomization test, two categorical variables	(16) Two sample randomization test, one binary, one quantitative
(17) Two sample z-test of proportions	(18) Chi-square test
(19) none of the above are appropriate	

(a) A campus administrator wants to know if some campus groups are more likely to consume alcohol than others. They take a random sample of 1500 students and classify them as high risk or low risk drinkers,

[2] See the following handout for more inferential summaries: https://m58-intro-stats.netlify.app/slides/param_est_58.pdf

and whether they belong to a sorority, fraternity, or neither.

(b) A researcher wants to determine whether people with "positive attitudes" tend to live longer than those without positive attitudes. They collect data on those who were classified with a positive attitude and those who were not and record how long each person lived.

(c) A university is trying to determine whether parking is a problem on its campus. The student newspaper contacts a random sample of 200 students and asks whether or not they are frustrated with the parking situation. They want to estimate the proportion of students at the college who are frustrated with the parking situation.

(d) A student reporter wants to know whether Democrats were more likely to vote for Candidate NW than Republicans. They take random samples belonging to the Democratic and Republican parties and asks whether or not they voted for Candidate NW.

(e) Doctors are interested in understanding the range of healthy birth weights for full-term babies. A hospital collects data on all the births at the hospital over the year. They measure: birth weight, premature / not, amount of weight mother gained, mother's age, father's age.

(f) A new species of lizards is found, and a researcher wants to know their average flight speed. They capture 30 lizards and race them individually on a track in their laboratory.

(g) The Best Buy electronics store wants to estimate how much more men tend to spend at the store than women. They collect receipts for a random sample of 100 customers and examine the total amount of the bill.

(h) Every year, 900 people in the US die from bicycle accidents. Researchers are interested in understanding the effect of blood alcohol content (BAC) has on the odds of a fatal accidents. For a random sample of bicycle accidents (both fatal and non-fatal) in a given year, whether the BAC is above the legal limit or not is obtained.

```r
praise()
```

```
## [1] "You are bedazzling!"
```