STAT 218 – Midterm 1

Dr. Allison Theobold

April 26, 2020

Name: _____

Section (circle one): 8:10am 9:10am

Read and Sign the Following Statement:

I understand that give or receiving help on this exam is a violation of academic regulations and is punishable by a grade of **F** in this course. This includes looking at other students' exams and / or allowing other students, actively or passively, to see answers on my exam. This also includes revealing, actively or passively, any information about the exam to any member of Professor Theobold's STAT 218 class who has not yet taken the exam. The use of cell phones is strictly prohibited.

I pledge not to do any of these things.

Signed: _____

Instructions.

- Read and sign the honesty pledge at the top of this page. Your exam will not be graded unless the honesty pledge is signed!
- Attempt all questions and write legibly.
- Show ALL the steps of your work clearly.
- You have 50 minutes to complete this exam, so budget your time wisely.

Provided Formulas

$$R^2 = r^2 = \frac{s_y^2 - s_{residuals}^2}{s_y^2}$$

$$IQR = Q3 - Q1$$

1.5 IQR Rule: above $Q3 + (1.5 \times IQR)$ or below $Q1 - (1.5 \times IQR)$

$$\hat{y} = b_0 + b_1 \times x$$

$$Residual = y - \hat{y}$$

t-based confidence interval: point estimate $\pm\, t^*_{df} \times SE$

$$SE(\mu) = \frac{\sigma}{\sqrt{n}}$$

Q7 [2 points] An article in the San Luis Tribune claims that the average age for people who receive food stamps in SLO is 40 years. A Cal Poly student believes the average age is less than that. The student obtains a random sample of 100 people in SLO who receive food stamps, and finds their average age to be 39.2 years. Performing a hypothesis test, the student finds their sample mean to be statistically significantly lower than the age of 40 stated in the article (p-value < 0.05). Indicate whether each of the following interpretations are valid or invalid.

(a) The statistically significant result indicates that the majority of people who receive food stamps are younger than 40. (Circle one.)

- Valid
- Invalid

(b) An error must have been made. This difference in means (39.2 vs. 40 years) is too small to be statistically significant. (Circle one.)

- Valid
- Invalid

Q9 [2 points] When you change from a 90% to a 95% confidence interval, which part(s) of the confidence interval change? (Select all that apply)

(a) Statistic (midpoint)

(b) Multiplier

(c) Standard error

Q5 [4 points] Indicate whether each statement about a bootstrap resample is **TRUE** or **FALSE**.

(a) The bootstrap resample and original sample **must** be the same size. _____

(b) The bootstrap resample and original sample are **both** taken from the population. _____

(c) The bootstrap resample can **only** use values that were in the original sample. _____

(d) The bootstrap resample uses **all** of the values that were in the original sample. _____

Q3 The Atlantic marsh Fiddler Crab, *Minuca pugnax*, lives in salt marshes throughout the eastern coast of the United States. Historically, M. pugnax were distributed from northern Florida to Cape Cod, Massachusetts, but like other species have expanded their range northward due to ocean warming.

The Plum Island Ecosystem Long Term Ecological Research site collected data on Fiddler Crabs from 13 marshes on the Atlantic coast of the United States in the summer of 2016. The marshes spanned from northeast Florida to northeast Massachusetts. Researchers were able to collect between 25 and 37 adult male Fiddler Crabs at each marsh.

(a) [3 pts] A preview of the dataset is provided below. Use this preview to address the following questions.

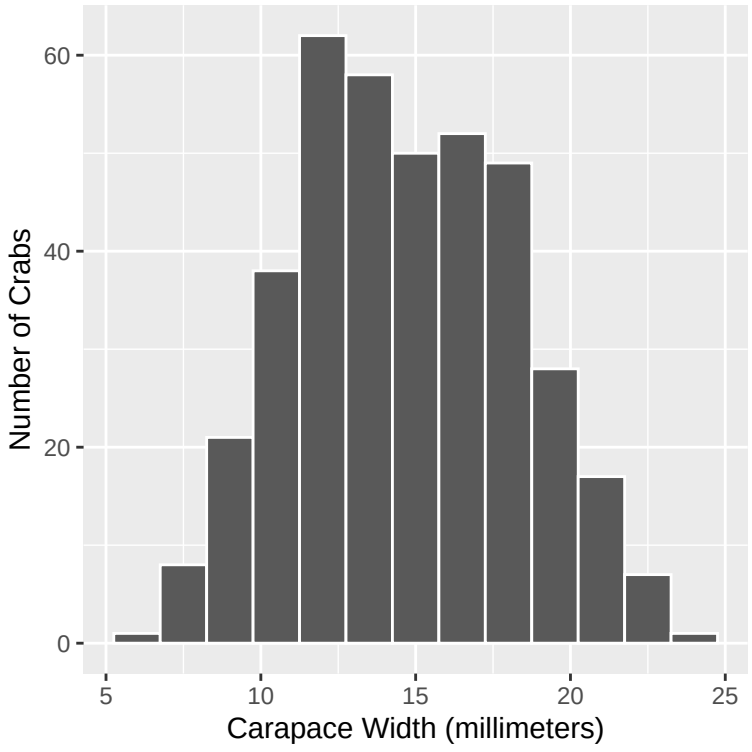
```
## Rows: 392
## Columns: 6
## $ latitude   <dbl> 30, 30, 30, 30, 30, 30, 30, 30, 30, 30, 30, 30, 30, 30, 30,~
## $ site       <chr> "GTM", "GTM", "GTM", "GTM", "GTM", "GTM", "GTM", "GTM", "GT~
## $ size       <dbl> 12.43, 14.18, 14.52, 12.94, 12.45, 12.99, 10.32, 11.19, 12.~
## $ air_temp   <dbl> 21.792, 21.792, 21.792, 21.792, 21.792, 21.792, 21.792, 21.~
## $ water_temp <dbl> 24.502, 24.502, 24.502, 24.502, 24.502, 24.502, 24.502, 24.~
## $ name       <chr> "Guana Tolomoto Matanzas NERR", "Guana Tolomoto Matanzas NE~
```

- Identify the cases in the data set.

- List the variables. Indicate whether each variable is categorical or quantitative.

- What would the dimensions of the data set be? (number of rows by number of columns)

(b) [3 pts] A histogram displaying the size of the sample of Fiddler Crabs is displayed below. Describe the shape of the distribution. Be sure to address the center, spread, shape, and outliers.



(c) [2 pts] When using a t-distribution to find a 95% confidence interval for μ, how many degrees of freedom should be used?

(d) [4 pts] A 95% confidence interval for the mean carapace width of Fiddler Crabs was found to be (14.31, 15.01). Below is the researchers' interpretation of this confidence interval:

There is a 95% chance that that every sample of Fiddler Crabs will have a mean carapace width between 14.31 and 15.01 millimeters.

Identify **two** mistakes committed and fix them. Be brief but clear in your description.

Mistake 1: _____

Fix: _____

Mistake 2: _____

Fix: _____

(e) [2 points] Can the researchers use their interval to make inferences about **all** Fiddler Crabs in the United States? Justify your answer!

Q2 I collected data on 512 different fast food items from McDonald's, Chick-Fil-A, Sonic, Arby's, Burger King, Dairy Queen, Subway, and Taco Bell. For each restaurant, I sampled 64 items from their menu and recorded the nutritional content of each item (e.g., calories, saturated fat, calcium, protein, etc.).

(a) [2 pts] Describe the sampling method I used to obtain these 512 fastfood items.

(b) [3 pts] I am interested in studying the linear relationship between the total calories of a food item and the amount of saturated fat that item contains.

Write the null hypothesis for my question of interest, using both words and notation.

(c) [2 pts] Is the alternative hypothesis one-sided or two-sided? (Circle one.)

- One-sided
- Two-sided

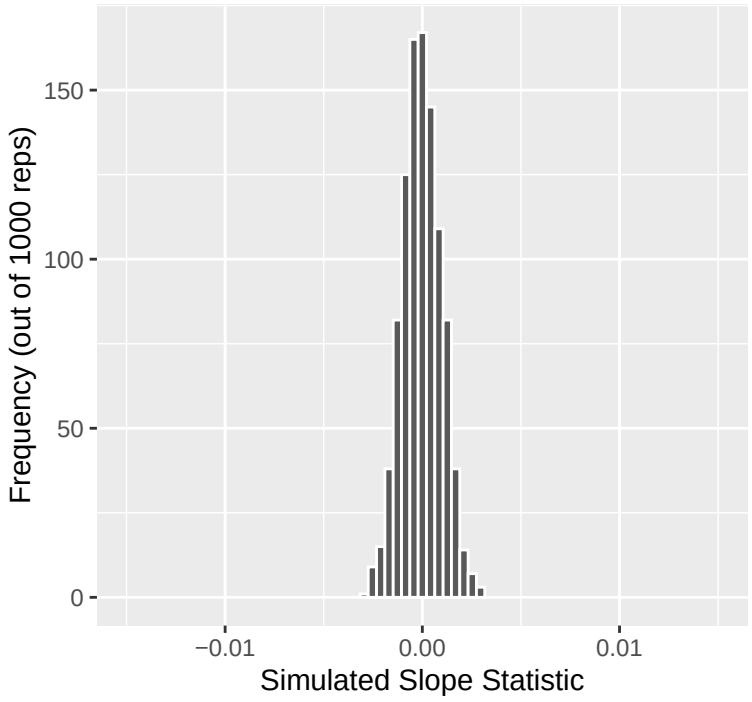
(d) [5 pts] Below is the plot of the simulated null distribution from `R`. Fill in the blanks below to correctly explain how one sample on the null distribution was created. Blanks preceded by (#) should be filled in with a number, all other blanks should be filled in with either the context of the study or the procedure that would need to be carried out to obtain one simulated sample.

On (#) _____ cards, write _____ on the cards.

Assume the null hypothesis is true and _____.

Generate a new sample of 512 ordered pairs by _____.

Calculate and plot the _____ from each simulated sample.



(e) [2 pts] Using the regression output below, draw a vertical line where the observed statistic falls on the null distribution.

term	estimate	std_error
intercept	-0.771	0.406
calories	0.017	0.001

(f) [2 pts] Shade the location of the plot you would use to calculate the p-value.

(g) [1 pts] Estimate the p-value associated with this hypothesis test.

(h) [3 pts] Which of the following is a correct interpretation of the p-value obtained? (Circle one)

- In less than 1 out of 1000 simulated samples, we would observe a sample slope of 0.017 or more extreme, if there is no linear relationship between the total calories and the saturated fat of a fast food item.
- If there is a linear relationship between the total calories and the saturated fat of a fast food, we would observe a sample slope of 0.017 or more extreme with a probability of less than 1 out of 1000.
- The probability of seeing a sample slope between the total calories and the saturated fat of a fast food item of 0.017 or more extreme is less than 0.1
- The probability that there is no linear relationship between the total calories and the saturated fat of a fast food item, is less than 0.1

(i) [2 points] Given the p-value for the hypothesis test, would the 95% confidence interval for β_1 contain 0? Be sure to justify your choice!

END OF EXAM

Points Earned: _____

Total Points: 46