STAT 218 - Midterm 1

Dr. Theobold

October 20, 2022

Name: _____

Section (circle one):   2:10pm   4:10pm

Read and Sign the Following Statement:

I understand that giving 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 Dr. Robinson'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!

- You may use a calculator. You **may not** use your phone or any device that connects to the internet as a calculator.

- Show all work as clearly as possible. Point totals are shown in brackets next to each part. Formulas without values entered do not count as work.

- All answers should be reported in decimal form, rounded to three decimal places.

- For multiple choice and multi-select problems, completely fill in the provided circle (multiple choice) or square (multi-select) for your desired answer choice(s). If you change an answer, be sure to completely erase your initial selection.

- You have 50 minutes to complete this exam, so budget your time wisely.

Provided Formulas

$$R^2 = r^2$$

$$IQR = Q3 - Q1$$

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

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

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

$$SE(\bar{x}) = \frac{s}{\sqrt{n}}$$

general formula for a confidence interval: point estimate $\pm$ multiplier $\times$ SE(point estimate)

t-based confidence interval: $\bar{x} \pm t^*_{df} \times SE(\bar{x})$

Q8 [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) Point estimate (midpoint)

(b) Multiplier

(c) Standard error

Q6 [3 points] The purpose of creating a null distribution is to: (Select all that apply)

(a) Discover what statistics might have occurred if the null hypothesis was true.

(b) To determine if the null hypothesis is true.

(c) To determine if the observed statistic is unlikely if the null was true.

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 directly 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.

```
## # A tibble: 392 x 6
##    latitude site   size air_temp water_temp name
##       <dbl> <chr> <dbl>    <dbl>      <dbl> <chr>
## 1        30 GTM    12.4     21.8       24.5 Guana Tolomoto Matanzas NERR
## 2        30 GTM    14.2     21.8       24.5 Guana Tolomoto Matanzas NERR
## 3        30 GTM    14.5     21.8       24.5 Guana Tolomoto Matanzas NERR
## 4        30 GTM    12.9     21.8       24.5 Guana Tolomoto Matanzas NERR
## 5        30 GTM    12.4     21.8       24.5 Guana Tolomoto Matanzas NERR
## 6        30 GTM    13.0     21.8       24.5 Guana Tolomoto Matanzas NERR
## 7        30 GTM    10.3     21.8       24.5 Guana Tolomoto Matanzas NERR
## 8        30 GTM    11.2     21.8       24.5 Guana Tolomoto Matanzas NERR
## 9        30 GTM    12.7     21.8       24.5 Guana Tolomoto Matanzas NERR
## 10       30 GTM    14.6     21.8       24.5 Guana Tolomoto Matanzas NERR
## # ... with 382 more rows
```

- Identify the observational units in the data set.

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

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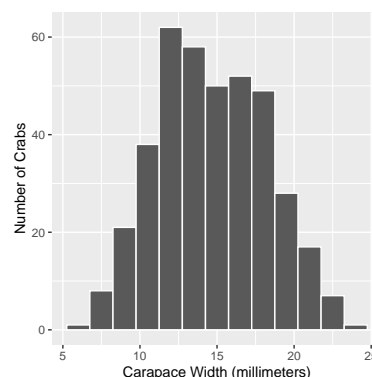
(c) [2 pts] When using a t-distribution to find a 95% confidence interval for μ, how many degrees of freedom should be used?

(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 the observed sample mean carapace width of the 392 Fiddler Crabs is between 14.31 and 15.01 millimeters.

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

Mistake 1: _____

Fix: _____

Mistake 2: _____

Fix: _____

Mistake 3: _____

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 Mcdonalds, Chick-Fil-A, Sonic, Arby's, Burger King, Dairy Queen, Subway, and Taco Bell. To obtain these data, for every restaurant I randomly sampled 64 items from their entire menu and recorded the nutritional content of each item selected (e.g., calories, saturated fat, calcium, protein, etc.).

(a) [2 pts] Describe the sampling method I used to obtain these 512 fast food 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 <u>words</u> and <u>notation</u>.

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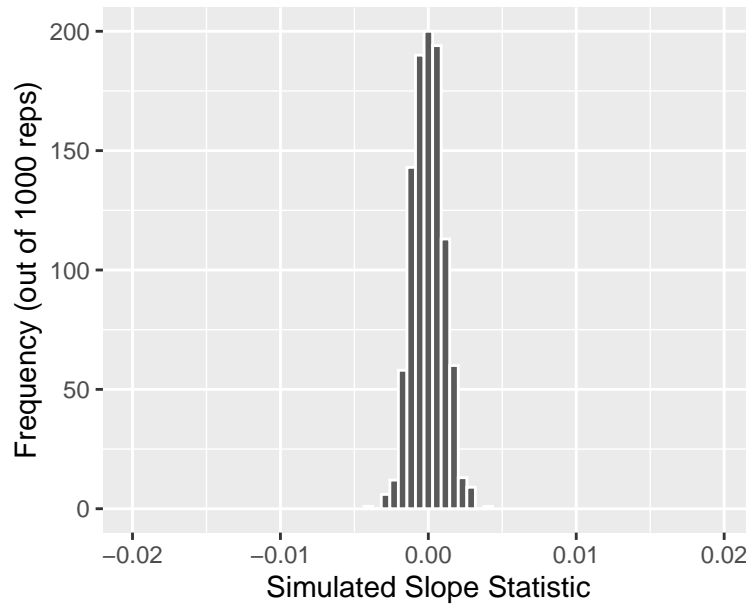
(e) [2 pts] Using the regression output below, draw a vertical line where the observed statistic falls on the null distribution.

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)

  (a) If there is no linear relationship between the total calories and the saturated fat of a fast food item, we would observe a sample slope of 0.017 or more extreme with a probability of less than 1 in 1000.

  (b) 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.

  (c) 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%.

  (d) 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!

Q11 The Konza Prairie Long-Term Ecological Research has collected data on bison on the Konza prairie since 1994, making it the longest continuous record of wild ungulate weight gain anywhere in the world. Researchers conduct a round-up once a year at the end of the grazing season wherein each bison is weighed, calves are vaccinated and receive unique IDs, and excess individuals are culled.

For this investigation, we are interested in assessing if, despite the effect of climate change on their habitat, the weight of yearling, male bison is what is described as "healthy" — a weight of approximately 750 pounds.

Below are summary statistics for the 48 of the yearling, male bison captured in 2020.

min	Q1	median	Q3	max	mean	sd	n	missing
490	595	620	662.5	770	629.5	63.17	47	1

(a) [3 points] Define the parameter of interest in words and use proper notation to assign a symbol.

(c) [2 points] If we decided to use inferential methods to assess if the true mean weight of yearling bison on the Konza Prairie was healthy, we must verify two conditions. What are these two conditions?

Condition 1:

Condition 2:

END OF EXAM

Points Earned: _____

Total Points: 46