For instructor use

Problem	Score
1	
2	
3	
4	
5	
6	
Total	

Student Name:

MATH 1342
Elementary Statistics
Test 2
Spring 2021

Problem 1 (2 points)

Please acknowledge awareness of the ACC Academic Integrity and Disciplinary Process by placing your initials beside each of the following statements.

_____ I understand that it is not permissible to collaborate with other students enrolled in this course, or to seek outside help on tests administered online.

_____ I understand that only the resources listed by the instructor are allowed for use on all tests administered online

_____ I understand that violation of any of these policies may result in disciplinary action, to include one or more of the following:

- A reduction in the grade on the work in question;
- A failing grade on the work in question;
- A failing grade or other denial of credit in the course;
- Recommending disciplinary sanctions to the Dean of Student Services.

Signature: _____

Problem 2 (8 points) Write True or False after each of the following statements.

1. For the two variables: depth of tire tread and number of miles driven on the tires, we expect the *association* to be negative.

2. We wish to determine if the weekly Grocery bill changes based on the Number of family members. The *explanatory variable* is Grocery bill, and the *response variable* is Number of family members.

3. Suppose that the original sample is: 12, 15, 15, 15, 28, 49.
   The following values constitute a possible *bootstrap sample*: 12, 12, 12, 15, 28, 49

4. The *margin of error* of the confidence interval CI = (10, 40) is equal to 15.

Problem 3 (12 points) Write the letter in front of the best answer after each of the following questions.

1. A study was designed to answer the following question: How much difference is there in the mean protein level between organic and non-organic eggs sold at HEB? Give the notation for the parameter of interest.
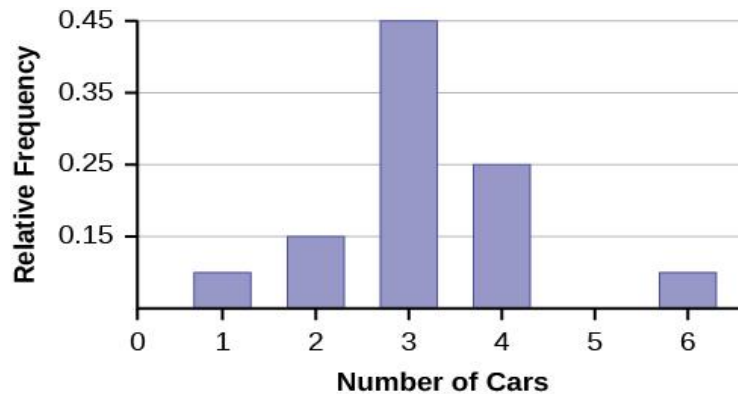
   A. $p_1 - p_2$ B. $\hat{p}_1 - \hat{p}_2$ C. $\mu_1 - \mu_2$ D. $\bar{x}_1 - \bar{x}_2$ E. $\rho_1 - \rho_2$ F. $r_1 - r_2$

2. The points on a scatterplot lie very close to the line whose equation is y = 4 − 3x.
   The *correlation* between x and y is close to:

   A. −1   B. 0   C. 1   D. −3

3. 68 randomly chosen households in South Austin were asked the number of cars they owned. The results are given in the graph below. Approximately how many households own 4 cars?

A. 0   B. 5   C. 10   D. 17   E. 30   F. 36

4. Which of the following statements is a proper interpretation of a 95% confidence interval.

A. I am 95% sure that this interval will contain the sample statistic.

B. 95% of the possible samples from this population will have sample statistics in this particular interval.

C. I am 95% sure that this interval will contain the population parameter.

D. 95% of the population values will fall within this interval.

Problem 4 (20 points)
Florida has over 7700 lakes. We wish to estimate the average concentration of Calcium carbonate (measured in mg/L) of all Florida lakes. Below are two boxplots. One represents the Calcium data for one random sample of size $n = 30$. The other represents the values in a *sampling distribution* of 1000 means of Calcium data from samples of size 30.
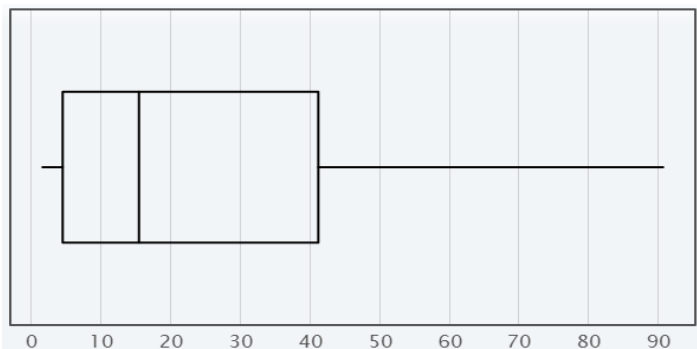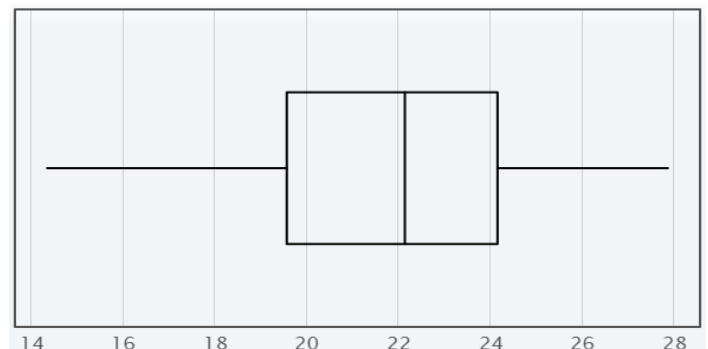
Boxplot A   Boxplot B

1. Give the notation and describe the parameter of interest in context.

   Parameter:

   Description:

2. Give the notation and describe in context the statistic we use to estimate the parameter of interest.

   Statistic:

   Description:

3. Which boxplot represents the sampling distribution? Justify your answer.
   (For full credit, please give at least two arguments in support of your conclusion.)

   Boxplot: A   B   (Choose one.)

   Explanation:

   1.

   2.

4. Consider the boxplot showing the data from a sampling distribution.

   4.1 What does one value in the sampling distribution represent?

   4.2 How many values are included in the data to make the boxplot?

   4.3 Use an <u>integer number</u> with correct units and appropriate notation to give a rough estimate
        of the value of the population parameter.

Problem 5 (25 points)
The residents in a small town are debating whether to build a new bridge across the river. The table
below summarizes the opinions of a random sample of 550 residents, broken down by gender.
We wish to estimate the difference in the proportions of all male and female residents who are in
favor of building the bridge.

Opinion	Male	Female	Total
In favor	131	185	**316**
Not in favor	85	149	**234**
Total	**216**	**334**	**550**

p_1- proportion of <u>all male residents</u> who are
      in favor of building the bridge

p_2- proportion of <u>all female residents</u> who are
      in favor of building the bridge

1. Give the notation for the parameter of interest and the statistic we can use to estimate it.

   Parameter:

   Statistic:

2. Calculate the value of the *best point estimate* for the difference in the proportions of all male and
   female residents who are in favor of building the bridge. <u>Show your calculations.</u>
   Round your answer to two decimal places.

   Best estimate =

3. The figure below gives a *bootstrap distribution* for the difference in proportions,
   based on 1000 simulated bootstrap samples.

samples = *1000*
mean = *0.051*
std. error = *0.045*
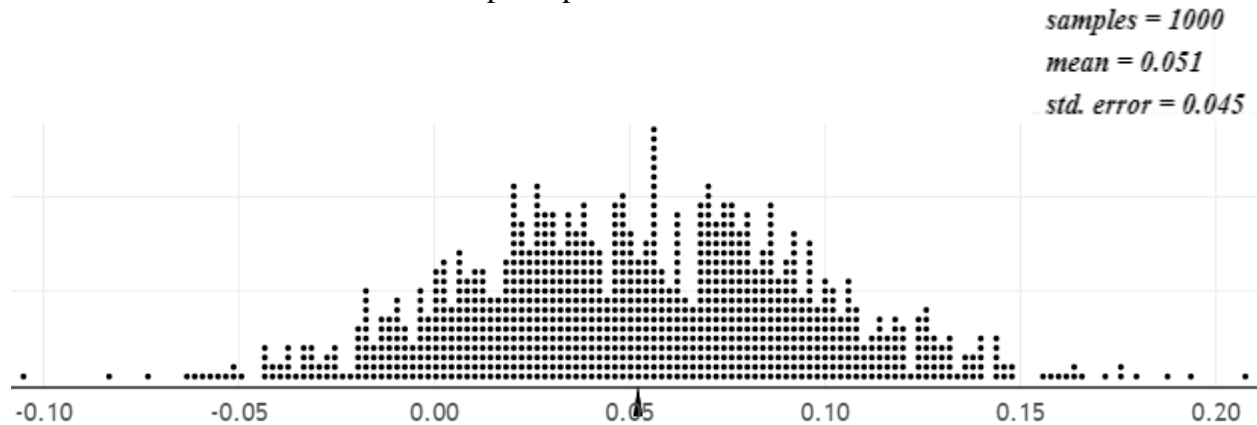
What shape does the distribution have? Choose one:

left-skewed   right-skewed   symmetric and bell-shaped   symmetric but not bell-shaped

4. Use the *percentile method* to estimate a 98% confidence interval.
   Round you answer to two decimal places.

   Number of dots you counted in each tail:

   98% CI:

5. Use the *standard error method* to estimate a 95% confidence interval. Show your calculations.
   Round you answer to two decimal places.

   95% CI:

6. Based on your 95% confidence interval, is it plausible that the proportions of male and
   female residents who are in favor of building the bridge are the same? Justify your answer.

   Plausible: Yes   No   (Choose one.)
   Explanation:

7. Which of the following would you expect to result in a narrower confidence interval
   than the 95% confidence interval you constructed above?

   A. Increasing the confidence level to 99%

   B. Generating more bootstrap samples from the existing data

   C. Gathering opinions from more residents for a larger sample

Problem 6 (40 points)
When water flows across farm land, some of the soil is washed away, resulting in erosion. The following data shows the amount of soil washed away for various flow rates.

Flow Rate (liters per second)	0.31	0.85	1.26	1.37	1.72	1.83	2.47	2.73	3.24	3.75
Eroded Soil (kilograms)	1.78	3.52	3.18	4.81	3.05	5.05	4.25	4.01	5.62	7.07

StatKey output from running the least squares procedure is shown below.
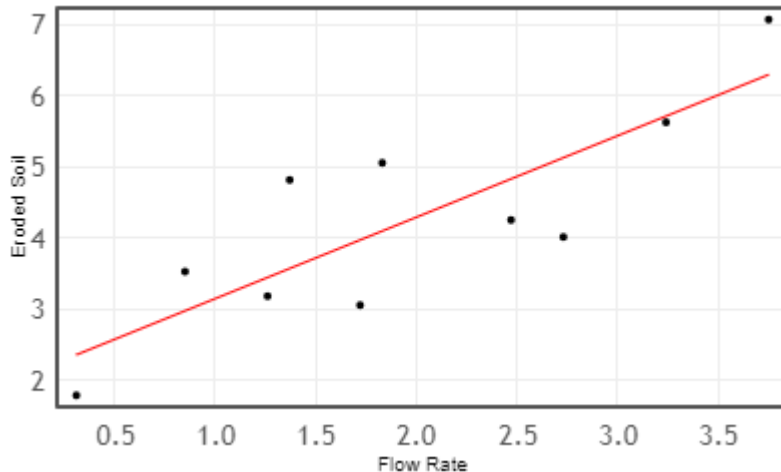
Statistic	Flow Rate	Eroded Soil
Mean	1.953	4.234
Standard Deviation	1.083	1.493
Sample Size		10
Correlation		0.832
Slope		1.146
Intercept		1.996

1. Indicate the name and type of the explanatory and response variables.

   Explanatory: _____ , Categorical   Quantitative   (Choose one.)

   Response: _____ , Categorical   Quantitative   (Choose one.)

2. What is the value of the correlation between flow rate and eroded soil, using correct notation?

   Correlation:

3. Give the value of the slope of the regression line using correct notation and interpret it in context.

   Slope:

   Interpretation:

4. Give the value of the intercept using correct notation and interpret it in context. Do you think it makes sense in this context?

   Intercept:

   Interpretation:

   Makes sense: Yes   No   (Choose one.)
   Explanation:

5. Predict the amount of soil washed away for a water flow rate of 5 liters per second. <u>Show your calculations</u>. Round your answer to two decimal places.

   Amount of soil =   kg

6. Describe reasons why you might not place much faith in that prediction.

7. What is the notation for the parameter we can estimate using the correlation from our sample?

   Parameter:

8. What sample size should be used to construct a *bootstrap distribution* to estimate this parameter?

   $n =$

9. A dotplot for one such bootstrap distribution from 5000 bootstrap samples is shown below.
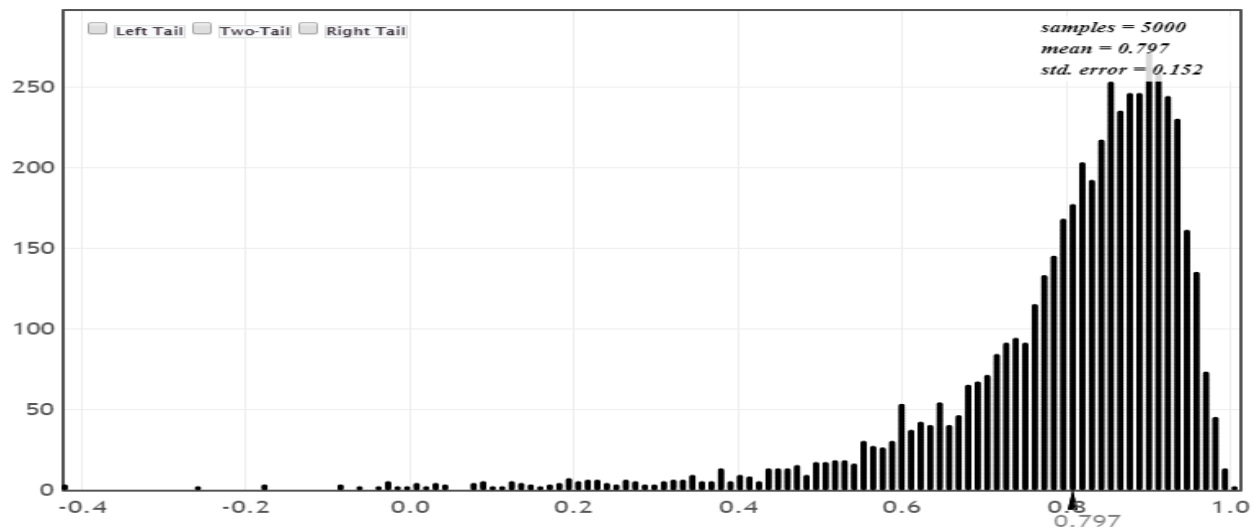
   Is it appropriate to construct a confidence interval from this distribution? Why or why not?

   Appropriate: Yes   No   (Choose one.)

   Explanation: