

Name: _____

Statistics Competency Fall 2017

Problem 1:

R and/or Minitab along with a calculator may be used for this entire section. Once you have completed this section, e-mail Part II to Dr. Price at pricejr@etsu.edu, Dr. Lewis at lewiscn2@etsu.edu, and Dr. Hendrickson at hendricksonj@etsu.edu. Please make sure you e-mail it to all three professors. The data file file (<http://faculty.etsu.edu/pricejr/1530/Fall17data.csv>)

A survey asked students, "Approximately, how many phone calls do you make per day?"

a. Create an appropriate graph to display the *distribution* of the variable called **CALLS** and insert it here.

b. Which of the following best describes the shape of the distribution? Underline (or highlight) your answer.

Skewed left

Uniform

Skewed right

Bimodal

Symmetric

c. Calculate the basic statistics for the data collected on **CALLS**. Copy and paste the output here.

Choose statistics that are appropriate for the shape of the distribution to describe the center and spread of **CALLS**.

d. Which statistic will you use to describe the center of the distribution? _____

e. In one or two sentences, describe why this statistic was chosen.

f. What is the value of that statistic? _____

g. Which statistic(s) will you use to describe the spread of the distribution? _____

h. What is (are) the value(s) of that (those) statistic(s)? _____

i. Are there any outliers in this distribution? Justify your answer using the IQR rule or an appropriate plot.

Name: _____

Problem 2: CLASS versus AGE. A survey asked students, “What is your age (in years)?” and “What is your classification in college? (Freshman/first-year, Sophomore, Junior, Senior)”

a. Create a suitable graph to display the *distribution* of **CLASS** and insert it here.

b. What is the mode of this distribution? (Please underline one option.)

Freshman/first-year Sophomore Junior Senior

c. Create a graph to display the age of students for the different levels of **CLASS**. Insert your graph here.

Use your graph found in part (c) to answer the following questions.

d. Which class has the oldest student? _____

e. Which class has the youngest student? _____

f. Which age-group has the largest IQR? _____

Name: _____

Problem 3: MILES vs. GAS. A survey asked students, “Approximately, how many miles do you live from campus? (Enter 0 if you live on campus)” and asked students, “Approximately, how much do you spend on gas (in U.S. dollars) a week?” We are interested in seeing whether we can use the number of miles to predict the amount of dollars spent on gas for a week.

- a. Create an appropriate graph to display the relationship between **MILES** and **GAS**. Insert it here.
- b. Does the plot show a positive association, a negative association, or no association between these two variables? EXPLAIN what this means with respect to the variables being studied.
- c. Describe the *form* of the relationship between **MILES** and **GAS**.
- d. Report the value of the correlation between this pair of variables? $r =$ _____
- e. Based on the information displayed in the graph and the correlation you just reported, how would you describe the *strength* of the association?
- f. Obtain the equation for the least squares regression of **GAS** on **MILES**. Copy & paste the output here.
- g. Interpret the value of the slope in the least squares regression equation you found in part (f).
- h. Use the regression equation in part (f) to predict amount of dollars spent on gas for a week for a student that lives 5 miles from campus.

Predicted amount of dollars =

- i. How well does the regression equation fit the data? Explain. Justify your answer with appropriate plot(s) and summary statistics.

Name: _____

Problem 4 LGBT AND AGE GROUP A survey asked students “In the U.S., more Americans are identifying as LGBT. Do you, personally, identify as lesbian, gay, bisexual, or transgender? (Yes, No)” and the survey asked students, “What is your age (in years)?” This variable was divided into three age groups: Ages “12 to 20”, “21 to 25”, and “Over 25”. We named this variable **AGE GROUP**. We want to check if there is a relationship between **LGBT AND AGE GROUP** among ETSU students. Assume the students who took the survey are from an SRS of ETSU students.

- a. Create an appropriate **graph** to display the relationship between **LGBT** and **AGE GROUP**. Insert your graph here.
- b. Create an appropriate two-way table to summarize the data. Insert your table here.

SUPPOSE WE SELECT ONE STUDENT AT RANDOM: (Calculate the following probabilities.)

- c. What is the probability that this student identifies as LGBT *and* is aged 12 to 20?

P = _____

- d. What is the probability that this student identifies as LGBT *or* is aged 12 to 20?

P = _____

- e. What is the probability that this student does not identify as LGBT given that the student is aged over 25?

P = _____

- f. What is the probability that this student is aged over 25 given that the student does not identify as LGBT?

P = _____

Name: _____

Problem 5: In 2013, Gallop found that the ideal number of children Americans want is 2.6 children per family. (<http://www.gallup.com/poll/164618/desire-children-norm.aspx>). A survey asked students, "What is your ideal number of children?" A professor feels that this number may be lower for college students. Is ETSU student's ideal number of children, on average, less than 2.6 children?

a. Create a suitable graph to display the distribution of **CHILDREN** reported by our sample of college students and insert it here.

Perform a test of significance to see if ETSU college student's ideal number of children, on average, is lower than the 2.6 children reported by Gallop using $\alpha = 0.05$.

b. Write the correct null and alternative hypothesis for the test: _____

c. Perform the appropriate test. Copy and paste the output for the test here.

d. What is the name of your test statistic and what is its value? _____

e. What is the P-value for the test? **P** = _____

f. State your decision regarding the hypotheses being tested.

g. State your conclusion. USE COMPLETE SENTENCES.

h. Is the P-value valid in this case? _____

i. What assumptions are you making in order to carry out this test?

Name: _____

Problem 6: A survey asked, "What is your religious identification? (Protestant/Other Christian, Catholic, Mormon, Jewish, Muslim, Other Non-Christian religion, None/Atheist/Agnostic)" The Gallup took a survey of U.S. adults in December 2016 and reported that and reported that 18.2% of U.S. adults said their religion identification was None/Atheist/Agnostic (http://www.gallup.com/poll/200186/five-key-findings-religion.aspx?g_source=Religion&g_medium=newsfeed&g_campaign=tiles). Is the same true for the population of all U.S. college/university students?

- a. Create an appropriate graph to display the distribution of **RELIGION** and insert it here.
- b. How many of the students surveyed said "None/Atheist/Agnostic?"
- c. What proportion of our sample said "None/Atheist/Agnostic?"
- d. Assume that we treat the sample of students as a simple random sample drawn from the population of all U.S. college/university students. Use Minitab to calculate a 95% confidence interval for the proportion of students in the population who chose "None/Atheist/Agnostic" to the survey question (based on our sample data). Copy and paste the Minitab output here.
- e. Interpret the confidence interval you reported in part (d).
- f. What do you think? Do our results contradict the results obtained from survey by Gallup or do they appear to agree with it? EXPLAIN.