

Name: _____

Statistics Proficiency Fall 2021

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. Godbole at godbolea@etsu.edu, and Dr. Hendrickson at hendricksonj@etsu.edu. Please make sure you e-mail it to all three professors.

Problem 1: DISTANCE A survey asked students, "Using <https://www.google.com/maps>, how many miles is your permanent address from 1276 Gilbreath Dr, Johnson City, TN 37614? (If there are multiple routes, enter the miles for the shortest distance.)"

a. Create an appropriate graph to display the *distribution* of the variable called **DISTANCE** 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 **DISTANCE**. Copy and paste the output here.

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

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: HANDS versus VACCINE. A survey asked students, “In a typical day, about how many times do you wash your hands?” and “Do you agree disagree, or neither agree nor disagree with the following statement? Vaccines are safe. (Agree, Disagree, Neither agree nor disagree)”

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

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

Agree Disagree Neither agree nor disagree

c. Create a graph to display the number of times students wash their hands in a typical day (**HANDS**) for the different levels of **VACCINE**. Insert your graph here.

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

d. Which vaccine group has the maximum number of times they wash their hands in a typical day?

e. Which vaccine group has the largest IQR in terms of number of times they wash their hands in a typical day?

Name: _____

Problem 3: TIME vs. DISTANCE A survey asked students, “Using <https://www.google.com/maps>, how many miles is your permanent address from 1276 Gilbreath Dr, Johnson City, TN 37614? (If there are multiple routes, enter the miles for the shortest distance.)” and asked students, “Using <https://www.google.com/maps>, how long does it take to drive from your permanent address to 1276 Gilbreath Dr, Johnson City, TN 37614? (If there are multiple routes, enter the time for the shortest distance NOT the fastest time.)” We are interested in seeing whether we can use the distance to predict the time. **DISTANCE** is measured in miles and **TIME** is measured in minutes.

- a. Create an appropriate graph to display the relationship between **DISTANCE** and **TIME**. 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 **DISTANCE** and **TIME**.
- 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 **TIME** versus **DISTANCE**. 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 the time (in minutes) for a student who lives 25 miles away.
Predicted time =
- i. How well does the regression equation fit the data? Explain. Justify your answer with appropriate plot(s) and summary statistics.

Name: _____

Statistics Proficiency Fall 2021

Problem 4: SMOKERS and ELECTION A survey asked students “Do you feel it is okay for insurance companies to charge higher rates to smokers? (Yes, No)” and the survey asked students, “Would you prefer to amend the Constitution so the candidate who receives the most total votes nationwide wins the election, or to keep the current system, in which the candidate who wins the most votes in the Electoral College wins the election? (Amend the Constitution, Keep the Current System).” We want to check if there is a relationship between **SMOKERS** and **ELECTION** 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 **SMOKERS** and **ELECTION**. 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 says it is okay for insurance companies to charge higher rates to smokers *and* says they should amend the constitution?

P =

d. What is the probability that this student says it is okay for insurance companies to charge higher rates to smokers *or* says they should amend the constitution?

P =

e. What is the probability that this student says it is okay for insurance companies to charge higher rates to smokers given this student says they should amend the constitution.

P =

f. Carry out a test for the hypothesis that there is no relationship between **SMOKERS** and **ELECTION**. Use a significance level of $\alpha = 0.05$.

i. State the null and alternative hypotheses.

ii. Perform the test and include and output here.

iii. Which test statistic are you using and what is its value?

iv. State your decision and conclusion for the test.

Name: _____

Statistics Proficiency Fall 2021

v. Examine the data. Are the condition for inference in part (ii) violated. Explain.

Name: _____

Statistics Proficiency Fall 2021

Problem 5: PARENTS In 2019, TD Ameritrade conducted a survey on age it would be considered embarrassing to still be living at home with your parents (https://s2.q4cdn.com/437609071/files/doc_news/research/2019/Boomerang-Generation-Returning-to-the-Nest.pdf). A survey asked students, "At what age would you consider it embarrassing for an adult child to still be living at home with their parent(s)?" TD Ameritrade reported the average age that would be considered embarrassing to still be living at home with your parents was 28. On average, is the age that ETSU students consider it embarrassing for an adult child to still be living at home with their parent(s) the same as reported in the survey, 28?

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

Using $\alpha = 0.05$, perform a test of significance to see if, on average, the age that ETSU students consider it embarrassing for an adult child to still be living at home with their parent(s) the same as reported in the survey, 28.

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: WEEKDAY A survey asked, “Go to this website <https://www.mathsisfun.com/games/dayofweek.html>. Enter your birthday. What day of the week were you born on? (Sunday, Monday, Tuesday, Wednesday, Thursday, Friday, Saturday)” According to a survey by spacefem.com, the most common day of the week for babies to be born on is Thursday, with 17.3% of babies being born on that day.

- a. Create an appropriate graph to display the distribution of **WEEKDAY** and insert it here.

- b. 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 were born on a “Thursday” (based on our sample data). Copy and paste the Minitab output here.

- c. Interpret the confidence interval you reported in part (d).

- d. What do you think? Do our results contradict the results obtained from survey by spacefem.com or do they appear to agree with it? EXPLAIN.