

Statistics Competency Exam Fall 2018 Part II

Name _____

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 lewisn2@etsu.edu, and Dr. Hendrickson at hendricksonj@etsu.edu. Please make sure you e-mail it to all three professors.

The data labeled F16 that I have emailed you is needed to answer questions 1–4.

The data in the csv file is based on a student survey (that had a total of 902 respondents) that asked the following questions:

- **GENDER:** What gender do you identify with? (Female, Male, Other)
- **AGE:** What is your age (in years)?
- **ATHLETES_PAID:** Should collegiate athletes be paid to play? (Yes, No)
- **ANTHEM:** Do you agree with Colin Kaepernick (NFL quarterback) protesting the U.S. national anthem? (Kaepernick is sitting or kneeling during the U.S. national anthem in protest for racial injustice.) (Yes, No)
- **ELECTION:** In the upcoming 2016 U.S. presidential election, which presidential candidate do you plan to vote for? (Hillary Clinton, Donald Trump, Other)
- **FACULTY_WEAPONS:** Do you agree with ETSU faculty and staff being able to carry concealed weapons on campus? (Agree, Disagree)
- **STUDENTS_WEAPONS:** Should students be able to carry concealed weapons on campus? (Yes, No)
- **MARIJUANA:** If a student is caught on campus with marijuana, how much should that student be fined? (in US dollars)
- **STATES:** How many U.S. states have you visited?
- **SCHOOLS:** How many schools have you attended (including elementary, middle school, high school, trade school, college, etc)?
- **SALARY:** What is your ideal starting salary (yearly not hourly) that you wish to make after graduating college? (in US dollars)
- **WORDS:** Count the number of words in a random sentence that you wrote or generated. Enter the number of words.
- **LETTERS:** Count the number of letters in the same random sentence that you wrote or generated. Enter the number of letters.

1. Who are the individuals?
2. Which variables are categorical?
3. **WORDS and LETTERS:** Questions 12 and 13 from the survey asked students to write or generate a random sentence and “Count the number of words...” and “Count the number of letters...” from the same random sentence. We are interested in predicting the number of words from the number of letters in random sentences written by students. Assume the respondents are an SRS of all ETSU students.
 - (A) Create an appropriate plot to display the relationship between **WORDS** and **LETTERS** and insert it here. 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.
 - (B) What is the correlation between the pair of variables?
 - (C) Obtain the least squares regression equation for the pair of variables. Insert it here.
 - (D) Interpret the value of the slope in the least squares regression equation you found from part (c).
 - (E) What percent of the variation in **WORDS** is accounted for by its linear relationship with **LETTERS**?
 - (F) What is the predicted word count in a randomly generated sentence that contains 75 letters?

4. **SALARY** Question 11 from the survey asked students “What is your ideal starting salary (yearly not hourly) that you wish to make after graduating college? (in US dollars)” According to Forbes (as of Nov. 2018), the starting annual salary average is \$62,428. Is there good evidence that the ideal starting salary, on average, for ETSU students is greater than \$62,428?
- (A) Create an appropriate graph to display the distribution of ideal starting annual salary. Insert it here.
 - (B) Calculate a 99% confidence interval for the ideal starting salary of ETSU students and interpret this interval.
 - (C) Perform an appropriate hypothesis test and include the output here.
 - (D) What is the P-value for this test?
 - (E) State your decision and conclusion for the test using a significance level of $\alpha = 0.01$.
 - (F) What assumptions are we making about the sample for our interpretation to be valid?

The data labeled SP16 (that I have emailed you) is needed to answer questions 5-7.

The data in the csv file is based on a student survey that asked the following questions:

- **GENDER:** What gender do you identify with? (Female, Male, Neither, Both)
- **AGE:** What is your age (in years)?
- **EVENING_TIME:** What is your favorite way of spending an evening? (Staying at home, Watching TV, Resting/Relaxing, Reading, Visiting with friends, Other)
- **SMOKING:** Have you, yourself, ever smoked cigarettes in the past week? (Yes, No)
- **PRAYER_CLASSROOM:** Do you favor or oppose daily prayer in the classroom? (Favor, Oppose)
- **POLITICAL_AFFILIATION:** What political party do you identify with? (Republican, Democrat, Independent, Other)
- **BOOKS:** How many books (not including textbooks or other books for class) did you read in 2015?
- **MIN_WAGE:** How much do you believe minimum wage should be? (in US dollars)
- **TEXT:** How many different people do you text on a normal day?
- **RELIGION_ID:** What is your religious identification? (Christian Religion, Non-Christian Religion, None)
- **FASTEST_SPEED:** What is the fastest you have ever driven a car? (in mph)
- **SPEED_TICKETS:** How many speeding tickets have you got since you started driving?
- **CAR_AGE:** How old is your car? (in years)
- **MPG:** How many miles per gallon (mpg) does your car get? (in mpg)

5. **TEXT:** Question 9 from the survey asked students, “How many different people do you text on a normal day?”
- (A) Create an appropriate display for this variable and insert it here.
 - (B) Describe the shape of the distribution.
 - (C) Calculate numerical measures appropriate for the shape of the distribution to describe the center and spread of **TEXT**.
 - i. Which statistic will you use to describe the center of the distribution?
 - ii. What is the value of that statistic?
 - iii. Which statistic(s) will you use to describe the spread of the distribution?
 - iv. What is(are) the value(s) of that(those) statistic(s)?
 - (D) Create a side-by-side boxplot to display the number of different people that students text for the different levels of **Gender**. Insert your graph here.
 - (E) Describe the distributions of **TEXT** for the different levels of **GENDER** and compare them.
6. **POLITICAL AFFILIATION AND RELIGIOUS IDENTIFICATION:** Question 6 from the survey asked students “What political party do you identify with? (Republican, Democrat, Independent, Other)” and Question 10 from the survey asked students “What is your religious identification? (Christian Religion, Non-Christian Religion, None)” We want to check if there is a relationship between political affiliation and religious identification. Assume the students who took the survey are from an SRS of ETSU students.
- (A) Create an appropriate two-way table to summarize the data between **POLITICAL AFFILIATION** and **RELIGIOUS IDENTIFICATION** and insert it here.
 - (B) Find the probability that a randomly selected student identifies with the Republican party and identifies with a Christian Religion.
 - (C) Find the probability that a randomly selected student identifies with a Non-Christian Religion or as an Independent for political affiliation.
 - (D) Find the probability that a randomly selected student identifies with the Democratic party given they identify with a Christian Religion.
 - (E) Carry out a test for the hypothesis that there is no relationship between political affiliation and religious identification for ETSU students. Use a significance level of $\alpha = 0.05$.
 - i. State the null and alternative hypothesis.
 - ii. Perform the test and include any output here.
 - iii. Which test statistic are you using and what is its value?

- iv. State your decision and conclusion for the test.
 - v. Examine the data. Are the conditions for inference in part (ii) violated? Explain.
7. **PRAYER_CLASSROOM:** Question 5 from the survey asked students “Do you favor or oppose daily prayer in the classroom? (Favor, Oppose)”
- (A) What proportion of our sample said they “Favor” daily prayer in the classroom?
 - (B) Assume (for the purpose of this problem) that we may treat the sample of Math 1530 students that responded to the survey as a simple random sample drawn from the population of all U.S. college/university students. Calculate a 95% confidence interval for the proportion of students in the population who “Favor” daily prayer in the classroom (based on our sample data).
 - (C) Interpret the confidence interval you reported in (B).
 - (D) Is it possible to use your results from (B) to answer the question “Is there evidence that the proportion of all U.S. college/university students that “Favor” daily prayer in the classroom differs from 0.5?” Explain why or why not.