MATH 1530 CAPSTONE TECHNOLOGY PROJECT FALL 2016

**Problem 1: Identify Variable Type.** One of these is a variable that is categorical and one is quantitative. Consider the different graphs that correspond to each variable type. Use Minitab to createa graph appropriate for **each** variable’s type. Interpret each graph.

 **STATES ANTHEM**

**Problem 2: Sampling.** In the survey data, the variable “AGE”is the current age reported by each student.

a. Type the first 10 observations from the column representing the variable AGEinto the table below,and use this as your sample data for part (b). Then calculate the mean age of these first 10 observations and report the value below.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| N | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| **AGE (yrs)** |  |  |  |  |  |  |  |  |  |  |

b. The mean age of the first 10 students isyears. (Type the value into the space provided.)

c. Identify the type of sampling method you have just used: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

d. Next, select a random sample of size n = 10 (Go to Calc > Random Data > Sample from Columns). Type the number 10 in the “Number of rows to Sample” slot. Enter the variable “ID” and “AGE” into the “From columns” slot. Enter C17-C18 into the “Store samples in” slot. Record the data for your sample in the table below.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| N | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| **ID** |  |  |  |  |  |  |  |  |  |  |
| **AGE (yrs)** |  |  |  |  |  |  |  |  |  |  |

e. Calculate and report the mean age for your random sample of 10 students. The sample mean age is

years.

f. Identify the type of sampling method you have just used:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

g. Suppose we think of *all* the students who responded to the survey as a *population* for the purposes of this problem. In that case, the *population mean* age is 20.35 Discuss (two or more complete sentences) the **differences and similarities** between 20.35 and the answers you got in (b) and (e).

**Problem 3(e): If your E number ends in an even number (0, 2, 4, 6, or 8) then do this question. (Omit this page/problem if your E# ends with an odd number.)**

Question 10 of the FALL 2016 survey asked students, “How many schools have you attended (including elementary, middle school, high school, trade school, college, etc…)?”

a. Create an appropriate graph to display the *distribution* of the variable called **SCHOOLS** 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. Using Minitab, calculate the basic statistics for the data collected on **SCHOOLS**. Copy and paste all of the Minitab output here.

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

Which statistic will you use to describe the center of the distribution? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

e. What is the value of that statistic? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

f. Which statistic(s) will you use to describe the spread of the distribution? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

g. What is (are) the value(s) of that (those) statistic(s)? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

h. Look up the IQR rule on page 55 in our textbook. Are there any outliers in this distribution? If so, what are their values? How many are there? Justify your answer.

**Problem 3(o): If your E number ends in an odd number (1, 3, 5, 7, or 9) then do this question. (Omit this page/problem if your E# ends with an even number.)**

Question 8 of the FALL 2016 survey asked students, “If a student is caught on campus with marijuana, how much should that student be fined? (in US dollars)”

a. Create an appropriate graph to display the *distribution* of the variable called **MARIJUANA** 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. Using Minitab, calculate the basic statistics for the data collected on **MARIJUANA**. Copy and paste all of the Minitab output here.

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

Which statistic will you use to describe the center of the distribution? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

e. What is the value of that statistic? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

f. Which statistic(s) will you use to describe the spread of the distribution? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

g. What is (are) the value(s) of that (those) statistic(s)? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

h. Look up the IQR rule on page 55 in our textbook. Are there any outliers in this distribution? If so, what are their values? How many are there? Justify your answer.

**Problem 4: AGE\_GROUP versus STATES.** Question 2 of the survey asked students, “What is your age (in years)?” This variable was divided into four age groups: Ages “16 to 20”, “21 to 25”, “26 to 30”, and “Over 30”. We named this variable **AGE\_GROUP**. Question 9 of the survey asked students, “How many U.S. states have you visited?”

a. Create a suitable graph to display the *distribution* of **AGE\_GROUP** and insert it here.

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

 Ages 16 to 20 Ages 21 to 25 Ages 26 to 30 Over 30

c. Create a side-by-side boxplot to display the number of states students have visited for the different levels of **AGE\_GROUP**. (Go to Graph > Boxplot > One Y with Groups > OK. Select **STATES** for the “Graph variables” slot and **AGE\_GROUP** for the “Categorical variables for grouping” slot.) Insert your graph here.

Use the side-by-side boxplot found in part (c) to answer the following questions.

e. Which age-group has visited the most U.S. states? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

f. Which age-group has the smallest median? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

g. Which age-group has the largest IQR? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Problem 5: WORDS vs. LETTERS.** On the FALL 2016 Math 1530 survey, questions 12 and 13 asked students to either write or generate a random sentence. Question 12 asked the students to state the number of words in that sentence and question 13 asked the students to state the number of letters in that sentence. We are interested in seeing whether we can use the number of words to predict the number of letters.

a. Create an appropriate graph to display the relationship between **WORDS** and **LETTERS**. 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 **WORDS** and **LETTERS.**

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. Using Minitab, obtain the equation for the least squares regression of **LETTERS** on **WORDS**. 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 number of letters for a sentence that has 8 words. (Show your math.)

**Predicted number of letters =**

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

**Problem 6 (T): Flip a fair coin. If it lands on tails do this problem (Omit this page/problem AND DO PROBLEM 6(H) if it lands on heads.)**

**GENDER AND ELECTION** Question 1 from the FALL 2016 Math 1530 survey asked students “What gender do you identify with? (Female, Male, Other)” and Question 5 from the FALL 2016 Math 1530 survey asked students “In the upcoming 2016 U.S. presidential election, which presidential candidate do you plan to vote for? (Hillary Clinton, Donald Trump, Other)” We want to check if there is a relationship between **GENDER** and **ELECTION** among ETSU students. Assume the students who took the (FALL 2016 Math 1530) class survey are from an SRS of ETSU students.

a. Create an appropriate **graph** to display the relationship between **GENDER** and **ELECTION**. Insert your graph here.

b. Create an appropriate two-way table to summarize the data. Insert your table here. (**IN MINITAB: STAT 🡪 TABLES 🡪 CROSS TABULATION AND CHI-SQUARE. Make sure to select “Options” and click “No variables” under the *Display missing values for*”)**

**SUPPOSE WE SELECT ONE STUDENT AT RANDOM:** (Calculate the following probabilities and show your work.)

c. What is the probability that this student is female *or* will vote for Hillary Clinton?

**P = \_\_\_\_\_\_\_\_\_\_\_\_**

d. What is the probability that this student is male given that the student will vote for Donald Trump?

**P = \_\_\_\_\_\_\_\_\_\_\_\_**

e. What is the probability that this student will vote for Donald Trump given that the student is male?

**P = \_\_\_\_\_\_\_\_\_\_\_\_**

f. Do you think there may be an association between **GENDER** and **ELECTION**? Why or why not? Explain your reasoning based on what you see in your graph.

**Problem 6 (H): Flip a fair coin. If it lands on heads do this problem (Omit this page/problem AND DO PROBLEM 6(T) if it lands on tails.)**

**FACULTY\_WEAPONS AND STUDENTS\_WEAPONS** Question 6 from the FALL 2016 Math 1530 survey asked students “Do you agree with ETSU faculty and staff being able to carry concealed weapons on campus? (Agree, Disagree)” and Question 7 from the FALL 2016 Math 1530 survey asked students “Should students be able to carry concealed weapons on campus? (Yes, No)” We want to check if there is a relationship between **FACULTY\_WEAPONS** and **STUDENTS\_WEAPONS** among ETSU students. Assume the students who took the (FALL 2016 Math 1530) class survey are from an SRS of ETSU students.

a. Create an appropriate **graph** to display the relationship between **FACULTY\_WEAPONS** and **STUDENTS\_WEAPONS**. Insert your graph here.

b. Create an appropriate two-way table to summarize the data. Insert your table here. (**IN MINITAB: STAT 🡪 TABLES 🡪 CROSS TABULATION AND CHI-SQUARE. Make sure to select “Options” and click “No variables” under the *Display missing values for*”)**

**SUPPOSE WE SELECT ONE STUDENT AT RANDOM:** (Calculate the following probabilities and show your work.)

c. What is the probability that this student is agrees faculty should be able to carry concealed weapons on campus *or* students should be able to carry concealed weapons on campus?

**P = \_\_\_\_\_\_\_\_\_\_\_\_**

d. What is the probability that this student does not agree faculty should be able to carry concealed weapons on campus given that the student thinks students should not be able to carry concealed weapons on campus?

**P = \_\_\_\_\_\_\_\_\_\_\_\_**

e. What is the probability that this student thinks students should not be able to carry concealed weapons on campus given that this student does not agree faculty should be able to carry concealed weapons on campus?

**P = \_\_\_\_\_\_\_\_\_\_\_\_**

f. Do you think there may be an association between **FACULTY\_WEAPONS** and **STUDENTS\_WEAPONS**? Why or why not? Explain your reasoning based on what you see in your graph.

**Problem 7:** In 2015, the National Association of Colleges and Employers found that the average starting salary for a bachelor’s degree graduate from the Class of 2015 is $50,651 (<http://www.naceweb.org/s11182015/starting-salary-class-2015.aspx>). Question 11 of the survey asked students, “What is your ideal starting salary (yearly not hourly) that you wish to make after graduating college? (in US dollars)” Is ETSU student’s ideal starting salary, on average, $50,651 per year?

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

b. Perform a test of significance to see if ETSU college student’s ideal starting salary, on average, is the same amount as the starting salary for a bachelor’s degree graduate from the Class of 2015. If this is true, then the average **SALARY** reported by ETSU students should be $50,651. Thus,

**Ho**: μ = 50651 dollars

Write the correct alternative hypothesis for the test: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

d. Use Minitab to perform the appropriate test. Copy and paste the output for the test here.

e. What is the name of your test statistic and what is its value? \_\_\_\_\_\_\_\_\_\_\_\_\_

f. What is the P-value for the test? **P = \_\_\_\_\_\_\_\_\_\_\_\_**

g. State your decision regarding the hypotheses being tested.

h. State your conclusion. USE COMPLETE SENTENCES.

i. Is the P-value valid in this case? What assumptions are you making in order to carry out this test?

**Bonus Problem:** Question 3 on the FALL 2016 Math 1530 asked, “Should collegiate athletes be paid to play? (Yes, No)” The YouGov/Huffington Post took a survey of 1000 U.S. adults in October 2015 and reported that and reported that 56% of college athletes should not be paid (<https://today.yougov.com/news/2015/10/28/poll-results-paying-college-athletes/> ). Is the same true for the population of all U.S. college/university students?

a. Create an appropriate graph to display the distribution of **ATHLETES\_PAID** and insert it here.

b. How many of the students surveyed said “yes?”

c. What proportion of our sample said “yes?”

d. Assume (for the purpose of this problem) that we may treat the FALL 2016 sample of Math-1530 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 “yes” 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 the YouGov/Huffington Post or do they appear to agree with it? EXPLAIN.