# MATH 1530 CAPSTONE TECHNOLOGY PROJECT FALL 2017

Problem 1: Identify Variable Type.Which of these questions from the class survey produced variables that are categorical and which are quantitative? Use your word processor to underline/highlight the best option.

### **a.** **AGE** Categorical Quantitative Neither

Produce an appropriate plot for this variable.

[Insert plot here]

**b.** **SHOPS** Categorical Quantitative Neither

Produce an appropriate plot for this variable.

[Insert plot here]

**c.** **OFFERED** Categorical Quantitative Neither

Produce an appropriate plot for this variable.

[Insert plot here]

Problem 2: Sampling. In the survey data, the variable “**AGE**”is the age in years for each student.

**a.** Type the first 10 observations from the column representing the variable **AGE** into 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** |  |  |  |  |  |  |  |  |  |  |

**b.** The mean age of the first 10 students is **\_\_\_\_\_\_**. (Type the value into the space provided.)

**c.** 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** |  |  |  |  |  |  |  |  |  |  |

**d.** Calculate and report the mean age for your random sample of 10 students. The sample mean age is **\_\_\_\_\_\_\_**.

e. 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 20.001. Discuss (two or more complete sentences) the **differences and similarities** between 20.001 and the answers you got in (b) and (d).

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.)**

1. Question 3 of the FALL 2017 survey asked students, “Approximately how much did you spend on textbooks this semester (Round to the nearest whole dollar)?”

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

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

**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.

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 15 of the FALL 2017 survey asked students, “Flip a fair coin until you get three heads in a row. How many flips did it take before you got three heads in a row?”

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

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

**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.

Problem 4: SMOKE versus AGE**.** Question 1 of the survey asked students, “What is your age (in years)?” Question 12 of the survey asked students, “What is your current smoking status? (Current smoker, Former smoker, Non-smoker)”

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

**b.** What is the mode of this distribution? Underline (or highlight) one option.

 Current Smoker Former Smoker Non-smoker

**c.** Create a side-by-side boxplot to display the age of students for the different levels of **SMOKE**. (Go to Graph > Boxplot > One Y with Groups > OK. Select **AGE** for the “Graph variables” slot and **SMOKE** 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.

**d.** Which smoking status has the oldest student? **\_\_\_\_\_\_\_\_\_\_\_\_**

**e.** Which smoking status has the highest median age? **\_\_\_\_\_\_\_\_\_\_\_\_**

**f**. Which smoking status has the smallest IQR? **\_\_\_\_\_\_\_\_\_\_\_\_**

Problem 5: STATES vs. COUNTRIES. On the FALL 2017 Math 1530 survey, question 8 asked students, “How many U.S. states have you visited, not including the state you were born? (Enter zero if you have never traveled outside the state you were born.)” and question 9 asked students, “How many countries have you visited, not including the country you were born? (Enter zero if you have never traveled outside the country you were born.)” We are interested in seeing whether we can use the number of states visited to predict the number of countries visited.

**a.** Create an appropriate graph to display the relationship between **STATES** and **COUNTRIES**. 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 **STATES** and **COUNTRIES.**

**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 **STATES** on **COUNTRIES**. 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 countries visited for a student that has visited 8 U.S. states. (Show your math.)

**Predicted number of countries visited =**

**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.)**

**OFFERED AND SMOKE\_MARIJUANA:** Question 13 from the FALL 2017 Math 1530 survey asked students “Have you ever been offered to smoke marijuana? (Yes, No)” and Question 14 of the survey asked students, “Have you ever smoked marijuana? (Yes, No).” We want to check if there is a relationship between **OFFERED AND SMOKE\_MARIJUANA** among ETSU students. Assume the students who took the (FALL 2017 Math 1530) class survey are from an SRS of ETSU students.

**a.** Create an appropriate **graph** to display the relationship between **OFFERED** and **SMOKE\_MARIJUANA**. 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 has been offered marijuana *and* has smoked marijuana?

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

**d.** What is the probability that this student has been offered marijuana *or* has smoked marijuana?

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

**e.** What is the probability that this student has not smoked marijuana given that the student has been offered marijuana?

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

**f.** What is the probability that this student has been offered marijuana given that the student has not smoked marijuana?

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

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.)**

**READ\_BOOKS AND SHOPS:** Question 5 from the FALL 2017 Math 1530 survey asked students “Do you prefer to read books (excluding textbooks) in print or on an electronic device (such as a smart phone, tablet, computer, or e-reader)? (In print, On an electronic device)” and Question 7 from the FALL 2017 Math 1530 survey asked students “Do you prefer to shop for non-grocery items online or in a physical store? (Online, In a physical store).” We want to check if there is a relationship between **READ\_BOOKS** and **SHOPS** among ETSU students. Assume the students who took the (FALL 2017 Math 1530) class survey are from an SRS of ETSU students.

**a.** Create an appropriate **graph** to display the relationship between **READ\_BOOKS** and **SHOPS**. 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 prefers to read books on an electronic device *and* the student prefers to shop for non-grocery items online?

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

**d.** What is the probability that this student prefers to read books on an electronic *or* the student prefers to shop for non-grocery items online?

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

**e.** What is the probability that this student prefers to read books in print given that the student prefers to shop for non-grocery items in a physical store?

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

**f.** What is the probability that this student prefers to shop for non-grocery items in a physical store given that this student prefers to read books in print?

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

Problem 7: In 2017, Business Insider did a story on using a mathematical theory to determine the best age to get married. This story reported the best age to get married was 26. ([**http://www.businessinsider.com/best-age-to-get-married-is-26-2017-2**](http://www.businessinsider.com/best-age-to-get-married-is-26-2017-2)). Question 10 of the survey asked students, “At what age do you think is the best to get married in order to have a successful marriage?” On average, is the age that ETSU students believe is ideal to get married in order to have a successful marriage the same as reported in the story, 26?

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

Perform a test of significance to see if, on average, the age that ETSU students believe is ideal to get married in order to have a successful marriage the same as reported in the story, 26, using *α* = 0.05.

**b.** Write the correct null and alternative hypothesis for the test: **\_\_\_\_\_\_\_**

**c.** Use Minitab to 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?

Bonus Problem:Question 11 on the FALL 2017 Math 1530 asked, “Which comes closest to your opinion about Confederate monuments in public spaces: “They should be removed.”, “They should remain.”, or “I do not know.”? (They should be removed, They should remain, I do not know)” An online survey conducted by the software company icitizen reported that 43% of adults said the monuments should be remain (<https://hyperallergic.com/397792/polls-americans-confederate-statues-removal/> ). Is the same true for the population of all U.S. college/university students?

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

**b.** How many of the students surveyed said “They should remain?”

**c.** What proportion of our sample said “They should remain?”

**d.** Assume (for the purpose of this problem) that we may treat the FALL 2017 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 “They should remain” 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 icitizen or do they appear to agree with it? EXPLAIN.