# MATH 1530 CAPSTONE TECHNOLOGY PROJECT FALL 2018

Problem 1: Sampling In the survey data, the variable “**AGE**”is the age in years for each student.

**a.** Starting with the first observation, select every 20th observation until you have 10 observations. Type the 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 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 above 10 students is **\_\_\_\_\_\_**. (Type the value into the space provided.)

**c.** What type of sampling was used in part (a)?

**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** |  |  |  |  |  |  |  |  |  |  |

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

**f.** What type of sampling was used in part (d)?

**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 19.423. Discuss (two or more complete sentences) the **differences and similarities** between 19.423 and the answers you got in (b) and (e).

Problem 2(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 6 of the FALL 2018 survey asked students, *“The following link will allow you to roll two virtual six-sided dice. Make sure the Roll is set to 2 virtual dice. Click Roll Dice. Here is the link:* [*https://www.random.org/dice/*](https://www.random.org/dice/)*.* ***What is the sum of the dots on the dice?****”* Note: You do NOT need to go to the website and roll the dice.

**a.** Create an appropriate graph to display the *distribution* of the variable called **Dice** 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 **Dice**. 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 **Dice.**

**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 2(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.)**

1. Question 8 of the FALL 2018 survey asked students, “Approximately how many disposable plastic straws do you use during a typical week?”

**a.** Create an appropriate graph to display the *distribution* of the variable called **Straws** 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 **Straws**. 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 **Straws**.

**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: Gender versus Foot**.** Question 2 of the survey asked students, “What gender do you identify with? (Female, Male, Other) ” Question 3 of the survey asked students, “What is the length of your foot (in inches)?”

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

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

 Female Male Other

**c.** Create a side-by-side boxplot to display the age of students for the different levels of **Gender**. (Go to Graph > Boxplot > One Y with Groups > OK. Select **Foot** for the “Graph variables” slot and **Gender** 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 gender group has the largest variability in terms of their foot length? **\_\_\_\_\_\_\_\_\_\_\_\_**

**e.** Which gender group has the highest median foot length? **\_\_\_\_\_\_\_\_\_\_\_\_**

**f**. Which gender group has the largest IQR in terms of their foot length? **\_\_\_\_\_\_\_\_\_\_\_\_**

 **g**. Discuss (two or more complete sentences) what this plot tells you.

Problem 4: REGRESSION We are interested in creating a regression model between two numeric variables asked from the survey.

**a.**  Which two quantitative variables have the strongest correlation?

**b.** Report the value of the correlation between these two variables? ***r* = \_\_\_\_\_\_\_\_\_\_\_\_**

**c.** Create an appropriate graph to display the relationship between these two variables.

[Insert plot here]

**d.** 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.

**e.** Describe the *form* of the relationship between these two variables**.**

**f.** Based on the information displayed in the graph and the correlation you just reported, how would you describe the *strength* of the association?

**g.** Using Minitab, obtain the equation for the least squares regression line. Use your best judgement to determine which is the response variable and which is the predictor variable.

[Copy & paste the output here.]

**h.** Interpret the value of the slope in the least squares regression equation you found in part (f).

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

Problem 5 (T): **If the sum of the digits in your E number is an even number then do this question. (Omit this page/problem if the sum of the digits in your E number is an odd number.)**

**Party and Morals:** Question 10 from the FALL 2018 Math 1530 survey asked students, “In politics, as of today, do you consider yourself a Republican, a Democrat, or an Independent? (Republican, Democrat, Independent)” and Question 13 of the survey asked students, “Do you personally believe that in general pornography is morally acceptable or morally wrong? (Morally acceptable, Morally wrong)” We want to check if there is a relationship between Party and Morals among ETSU students. Assume the students who took the (FALL 2018 Math 1530) class survey are from an SRS of ETSU students.

**a.** Create an appropriate **graph** to display the relationship between **Party** and **Morals**. 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 a Republican *and* feels pornography is morally acceptable?

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

**d.** What is the probability that this student is a Republican *or* feels pornography is morally acceptable?

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

**e.** What is the probability that this student is a Democrat *given* that the student feels pornography is morally acceptable?

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

**f.** **BONUS:** Carry out a test for the hypothesis that there is no relationship between **Party** and **Morals**. Use a significance level of α = 0.05.

i. State the null and alternative hypotheses.

ii. Perform the test and include any output from Minitab 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.

Problem 5 (H): **If the sum of the digits in your E number is an odd number then do this question. (Omit this page/problem if the sum of the digits in your E number is an even number.)**

**Party and Disease:** Question 10 from the FALL 2018 Math 1530 survey asked students, “In politics, as of today, do you consider yourself a Republican, a Democrat, or an Independent? (Republican, Democrat, Independent)” and Question 14 of the survey asked students, “When a person has a disease that cannot be cured, do you think doctors should be allowed by law to end the patient’s life by some painless means if the patient and his or her family request it? (Yes, No)” We want to check if there is a relationship between Party and Disease among ETSU students. Assume the students who took the (FALL 2018 Math 1530) class survey are from an SRS of ETSU students.

**a.** Create an appropriate **graph** to display the relationship between **Party** and **Disease**. 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 an Independent *and* says yes that when a person has a disease that cannot be cured the doctors should be allowed by law to end the patient’s life by some painless means if the patient and his or her family request it?

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

**d.** What is the probability that this student is an Independent *or* says yes that when a person has a disease that cannot be cured the doctors should be allowed by law to end the patient’s life by some painless means if the patient and his or her family request it?

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

**e.** What is the probability that this student is a Republican *given* the student says no that when a person has a disease that cannot be cured the doctors should be allowed by law to end the patient’s life by some painless means if the patient and his or her family request it?

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

**f.** **BONUS:** Carry out a test for the hypothesis that there is no relationship between **Party** and **Disease**. Use a significance level of α = 0.05.

i. State the null and alternative hypotheses.

ii. Perform the test and include any output from Minitab 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.

Problem 6: Question 1 of the survey asked students, “What is your age?” and Question 11 of the survey asked students, “Do you think the use of marijuana should be made legal? (Yes, No)” On average, is there a difference in the mean age for those who said marijuana should be made legal and those who said marijuana should not be made legal?

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

Using *α* = 0.05, perform a test of significance to see if, on average, there is a difference in the mean age for those who said marijuana should be made legal and those who said marijuana should not be made legal.

**b.** Write the correct null and alternative hypotheses for the test: **H0: \_\_\_\_\_\_\_\_\_\_\_\_\_\_

Ha: \_\_\_\_\_\_\_\_\_\_\_\_\_\_**

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

Problem 7 Question 9 from the FALL 2018 Math 1530 survey asked students “What is your favorite color in a rainbow? (Red, Orange, Yellow, Green, Blue, Indigo, Violet).” Many surveys have been conducted on a person’s favorite color of the rainbow is blue. One studied stated that 44% of people, stated blue was their favorite color of the rainbow (<http://awp.diaart.org/km/usa/survey.html>). Is the same true for the population of all U.S. college/university students?

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

**b.** How many of the students surveyed said “Blue” was their favorite color in the rainbow?

**c.** What proportion of our sample said “Blue” was their favorite color in the rainbow?

**d.** Assume (for the purpose of this problem) that we may treat the FALL 2018 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 90% confidence interval for the proportion of students in the population who chose “Blue” 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 the survey or do they appear to agree with it? EXPLAIN.