# MATH 1530 CAPSTONE TECHNOLOGY PROJECT SPRING 2019

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. Assume the first observation was randomly chosen as the starting point. 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 C18-C19 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 19.839. Discuss (two or more complete sentences) the **differences and similarities** between 19.839 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.)**

Question 8 of the SPRING 2019 survey asked students, “What is the fastest you have ever driven an automobile on the highway or interstate (in mph)?”

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

**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.)**

Question 4 of the SPRING 2019 survey asked students, “There are many health issues associated with Adverse Childhood Experiences (ACE). Go to the following link and take the short ACE quiz: <https://www.npr.org/sections/health-shots/2015/03/02/387007941/take-the-ace-quiz-and-learn-what-it-does-and-doesnt-mean>. What was your ACE quiz score?” Note, you do not need to take this quiz.

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

**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 ACE**.** Question 2 of the survey asked students, “What gender do you identify with? (Female, Male, Other)” Question 4 of the SPRING 2019 survey asked students, “What was your ACE quiz score?”

**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 **ACE** 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 ACE score? **\_\_\_\_\_\_\_\_\_\_\_\_**

**e.** Which gender group has the highest median ACE score? **\_\_\_\_\_\_\_\_\_\_\_\_**

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

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

Problem 4: REGRESSION On the SPRING 2019 Math 1530 survey, question 9 asked students, “How many wrecks have you been in where you were the driver?” and question 10 asked students, “How many speeding tickets have you received?” We are interested in seeing whether we can use the number of speeding tickets to predict the number of wrecks a student has been in where they were the driver.

**a.** Create an appropriate graph to display the relationship between **Tickets** and **Wrecks**. 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 **Tickets** and **Wrecks.**

**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 **Tickets** on **Wrecks**. 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 wrecks for a student that has had 3 speeding tickets. (Show your math.)

**Predicted number of wrecks =**

**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.)**

**Wall and Abortion:** Question 14 from the SPRING 2019 Math 1530 survey asked students, “Do you support the idea of building a border wall between the U.S. and Mexico? (Yes, No)” and Question 15 of the survey asked students, “Would you identify yourself as pro-choice or pro-life on abortion? (Pro-choice, Pro-life)” We want to check if there is a relationship between Wall and Abortion among ETSU students. Assume the students who took the (SPRING 2019 Math 1530) class survey are from an SRS of ETSU students.

**a.** Create an appropriate **graph** to display the relationship between **Wall** and **Abortion**. 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 supports the idea of building a border wall between the U.S. and Mexico *and* is pro-life?

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

**d.** What is the probability that this student supports the idea of building a border wall between the U.S. and Mexico *or* is pro-life?

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

**e.** What is the probability that this student does not support the idea of building a border wall between the U.S. and Mexico *given* the student is pro-life?

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

**f.** **BONUS:** Carry out a test for the hypothesis that there is no relationship between **Wall** and **Abortion**. 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.)**

**Gender and Textbook:** Question 2 from the SPRING 2019 Math 1530 survey asked students, “What gender do you identify with? (Female, Male, Other)” and Question 5 of the survey asked students, “Do you prefer to read textbooks in print or on an electronic device (such as a smart phone, tablet, computer, or e-reader)? (In print, On an electronic device)” We want to check if there is a relationship between **Gender and Textbook** among ETSU students. Assume the students who took the (SPRING 2019 Math 1530) class survey are from an SRS of ETSU students.

**a.** Create an appropriate **graph** to display the relationship between **Gender** and **Textbook**. 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 identifies as a female *and* says they prefer to read a textbook in print?

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

**d.** What is the probability that this student identifies as a female *or* says they prefer to read a textbook in print?

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

**e.** What is the probability that this student identifies as other *given* the student prefers to read a textbook on an electronic device?

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

**f.** **BONUS:** Carry out a test for the hypothesis that there is no relationship between **Gender** and **Textbook**. 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: In 2018, Gallup conducted a survey on age of retirement (<https://news.gallup.com/poll/234302/snapshot-americans-project-average-retirement-age.aspx>). One question asked was the same as question 13 from the SPRING 2019 MATH 1530 survey. Question 13 of the survey asked students, “At what age do you expect to retire?” Gallup asked this question for three different age groups. It was found that average age one expects to retire (for age group 18-29) was 63. Since the majority of students taking MATH 1530 fall in the age group 18-29, the variable **Retire\_Age\_18\_29** was created. This variable contains the values of what age a student expects to retire for those students aged 18 to 29. On average, is the age that ETSU students (aged 18-29) expect to retire the same as reported in the survey, 63?

**a.** Create a suitable graph to display the distribution of **Retire\_Age\_18\_29** 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, the age that ETSU students (aged 18-29) expect to retire the same as reported in the survey, 63.

**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 12 from the SPRING 2019 Math 1530 survey asked students “If you had to lose one of your five senses, which would you choose? (Hearing, Sight, Touch, Smell, Taste).” Many surveys have been conducted on which of the five senses would you want to lose if you had to lose one. One studied stated that 57% of people said smell would be the sense they would choose to lose (<https://www.quibblo.com/poll/1JoM1SK/If-you-HAD-to-give-up-a-sense-which-would-it-be>). Is the same true for the population of all U.S. college/university students?

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

**b.** How many of the students surveyed said “Smell” was the sense they would choose to lose?

**c.** What proportion of our sample said “Smell” was the sense they would choose to lose?

**d.** Assume (for the purpose of this problem) that we may treat the SPRING 2019 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 “Smell” 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.