# MATH 1530 CAPSTONE TECHNOLOGY PROJECT SPRING 2018

Problem 1: Sampling. In the survey data, the variable “**BIRTHDAY**”is the days left until a student’s next birthday.

**a.** “**BIRTHDAY.SUBSET**”is a variable that contains the first 50 observations from the variable “**BIRTHDAY.**”Calculate the mean number of days until a student’s next birthday.

The mean number of days until a student’s next birthday of the first 50 students is **\_\_\_\_\_\_**. (Type the value into the space provided.)

**b.** What type of sampling method was used in part (a)? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**c.** Next, select a random sample of size n = 50 (Go to Calc > Random Data > Sample from Columns). Type the number 50 in the “Number of rows to Sample” slot. Enter the variable “**BIRTHDAY**” into the “From columns” slot. Enter C19 into the “Store samples in” slot. Calculate and report the mean age for your random sample of 50 students.

The sample mean number of days until a student’s next birthday is **\_\_\_\_\_\_\_**. (Type the value into the space provided.)

**d.** What type of sampling method was used in part (c)? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**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* number of days until a student’s next birthday is 175.11. Discuss (two or more complete sentences) the **differences and similarities** between 175.11 and the answers you got in (a) and (c).

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 5 of the SPRING 2018 survey asked students, “What is your U.S. shoe size?”

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

**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 3 of the SPRING 2018 survey asked students, “How much cash do you have in your wallet or purse right now?”

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

**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: RELIGION versus BIRTHDAY**.** Question 1 of the survey asked students, “How many days left until your next birthday?” Question 11 of the survey asked students, “What is your religious preference? (Jewish, Mormon, Muslim, No Religion, Other Religion, Protestant, Roman Catholic)?”

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

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

 Jewish Mormon Muslim No Religion Other Religion Protestant Roman Catholic

**c.** Create a side-by-side boxplot to display the age of students for the different levels of **RELIGION**. (Go to Graph > Boxplot > One Y with Groups > OK. Select **BIRTHDAY** for the “Graph variables” slot and **RELIGION** 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 excluding Mormon as possible answer since there are only 2 observations for this religion group.

**d.** Which religion group has the smallest variability in terms of the number of days until their next birthday? **\_\_\_\_\_\_\_\_\_\_\_\_**

**e.** Which religion group has the highest median number of days until their next birthday? **\_\_\_\_\_\_\_\_\_\_\_\_**

**f**. Which religion group has the largest IQR in terms of the number of days until their next birthday? **\_\_\_\_\_\_\_\_\_\_\_\_**

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 (excluding **ID**, **BIRTHDAY.SUBSET, GENDER.SUBSET, HAIR.SUBSET**)?

**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.)**

**SHOOTING and ABORTION:** Question 9 from the SPRING 2018 Math 1530 survey asked students “How worried are you that you or someone in your family will become a victim of a mass shooting? (Very worried, Somewhat worried, Not too worried, Not worried at all)” and Question 12 of the survey asked students, “Do you think abortions should be legal under any circumstances, legal only under certain circumstances, or illegal in all circumstances? (Legal under any circumstances, Legal only under certain circumstances, Illegal in all circumstances).” We want to check if there is a relationship between **SHOOTING and ABORTION** among ETSU students. Assume the students who took the (SPRING 2018 Math 1530) class survey are from an SRS of ETSU students.

**a.** Create an appropriate **graph** to display the relationship between **SHOOTING** 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 is very worried about being a victim of a mass shooting *and* feels abortion is legal under any circumstances?

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

**d.** What is the probability that this student is very worried about being a victim of a mass shooting *or* feels abortion is legal under any circumstances?

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

**e.** What is the probability that this student is very worried about being a victim of a mass shooting *given* that the student feels abortion is legal under any circumstances?

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

**f.** **BONUS:** Carry out a test for the hypothesis that there is no relationship between **SHOOTING** 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 RESOLUTION:** Question 2 from the SPRING 2018 Math 1530 survey asked students “What gender do you identify with? (Female, Male, Other)” and Question 13 from the SPRING 2018 Math 1530 survey asked students “Did you make a New Year's Resolution for 2018? (Yes, No).” We want to check if there is a relationship between **GENDER** and **RESOLUTION** among ETSU students. Assume the students who took the SPRING 2018 Math 1530 class survey are from an SRS of ETSU students.

**a.** Create an appropriate **graph** to display the relationship between **GENDER** and **RESOLUTION**. 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 *and* made a New Year’s resolution for 2018?

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

**d.** What is the probability that this student is female *or* made a New Year’s resolution for 2018?

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

**e.** What is the probability that this student is female *given* the student made a New Year’s resolution for 2018?

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

**f.** **BONUS:** Carry out a test for the hypothesis that there is no relationship between **GENDER** and **RESOLUTION**. 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 7 of the survey asked students, “How many times a week do you wash your hair?” and Question 2 of the survey asked students, “What gender do you identify with?” The variables **GENDER.SUBSET** and **HAIR.SUBSET** were created from a subset of corresponding values. The subset data is to be used to answer this question: On average, is there a difference in the number of times a female student washes their hair in a week versus a male student?

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

Perform a test of significance to see if, on average, there is a difference in the number of times a female student washes their hair in a week versus a male student using *α* = 0.05.

**b.** Write the correct null and alternative hypotheses 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.** 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 SPRING 2018 Math 1530 survey asked students “How worried are you that you or someone in your family will become a victim of a mass shooting? (Very worried, Somewhat worried, Not too worried, Not worried at all).” Gallup conducted a study and found that 29% of Americans are somewhat worried that they or their family will be a victim of a mass shooting (<http://news.gallup.com/poll/220634/four-americans-fear-victim-mass-shooting.aspx?g_source=link_newsv9&g_campaign=item_226202&g_medium=copy>). Is the same true for the population of all U.S. college/university students?

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

**b.** How many of the students surveyed said they were “Somewhat worried”?

**c.** What proportion of our sample said they were “Somewhat worried”?

**d.** Assume (for the purpose of this problem) that we may treat the SPRING 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 95% confidence interval for the proportion of students in the population who chose “Somewhat worried” 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 Gallup or do they appear to agree with it? EXPLAIN.