**Final Exam \_\_\_\_ / 66 pts**

MAT 135 August 8th 2018

Introduction to Statistics

Instructions: The exam is divided in two sections. The first part is handwritten. You may use Excel for the calculations, but write everything down as indicated in the instructions of each problem. The second part will refer to a dataset containing the winning times of race horses in the Kentucky Derby. The dataset has been uploaded to *D2L > Course Content > Exams > Final Exam*. Save the Excel file and replace your name in the file name where appropriate. Save and label all of the answers to each question the second part in a spreadsheet. When you are finished, upload the spreadsheet to the dropbox on D2L.

Formulae:

*Binomial Distribution*

*Confidence Intervals*

*Test Statistic*

Part 1 : Handwritten Calculations

1. (*10 pts*) According to Wikipedia, 33% of Americans are overweight. In order to measure the validity of this statistic, you decide to stand outside of the Department of Motor Vehicles and ask every tenth person that enters the building their weight. Once you have a sample of 55 people ,you return home to calculate some probabilities.

a. Does this method of sampling satisfy the requirements of the binomial distribution? What might be some potential pitfalls of applying the binomial distribution to this sample of people? How could we improve our sampling method?

b. Assume the changes you proposed in part *a* have been made to the sample data and we now have a binomial distribution. Answer the following questions. You may use Excel to aid in the calculations, but set up the problem. Write down the probability statement and any steps you apply to arrive at the answer:

i. What is the expected number of overweight people in this sample of data?

ii. What is the probability exactly 19 people in this sample are overweight?

iii. What is the probability between 15 and 20 people in this sample are overweight?

iv. What is the probability at least 22 people in this sample are overweight?

A. Using a significance level of 0.05, is 22 a significant number of overweight people to observe in this sample?

v. What is the probability at least 26 people in this sample are overweight?

B. Using a significance level of 0.05, is 26 a significant number of overweight people to observe in this sample?

2. (*8 pts*) Calculate the following probabilities regarding the standard normal distribution by either referring to a table or using the appropriate Excel function. Show how the probability is calculated in terms of the cumulative distribution function, . Label and shade the provided graph appropriately for each problem. Shade in the area to the left or to the right as indicated by the problem. Label the mean of the standard normal distribution in each case.

a.

 b.



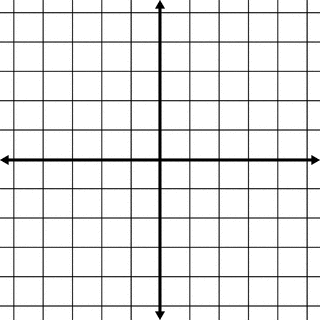
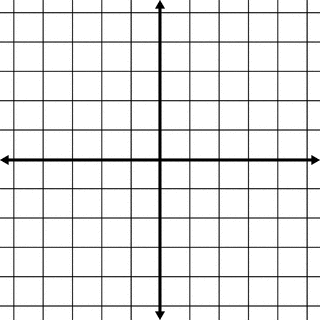
c.

 d.

3. (*12 pts*) Suppose the height of a population is normally distributed around the mean with a standard deviation of. You sample 45 people from this population and measure each of their heights. Answer the following questions regarding this sample of data:

a. How is the distribution of the sample mean related to the distribution of the population? Sketch the distribution of the population and the distribution of the sample mean on the provided graphs. Label the mean and standard deviation of each distribution:

Distribution of the Population Distribution of the Sample Mean



b. What is the probability of observing a member of this population taller than 5.9 *ft* tall?

c. What is the probability of observing a member of this population between 5.2 *ft* and 5.6 *ft* tall?

d. How tall does a person have to be in order to be taller than 95% of this population?

e. What is the probability of observing a sample mean greater than 5.9 *ft*?

f. Explain the difference between the answers in part *b* and *e*.

Part 2: Excel Calculations

Instructions: Navigate to *D2L > Course Content > Exams > Final Exam > Data Sets* and open the data set entitled, *KENTUCKY\_DERBY\_WINNING\_TIMES\_yourname.xls*. Save the spreadsheet, replacing your name where appropriate. Save and clearly label all of your answers.

4. (*10 pts*) Calculate the following sample statistics from this sample of data. Save your answers in a cell and label each one:

a. The minimum

b. The 25th percentile

c. The median

d. The 75th percentile

e. The maximum

f. The range

a. The sample mean

b. The sample standard deviation.

c. The number of observations.

5. (*4 pts*) Construct a histogram with 8 classes for this sample of data. Use a class width of 0.12. Label the cells for the lower and upper class limits.

a. Does the shape of this distribution appear to be normal, skewed right or skewed left? Refer to the values of the sample mean and median to justify your answer numerically as well as graphically.

b. If the underlying population distribution is not normal, is the sample mean normally distributed? What conditions are necessary for the sample mean to be normally distributed? Are the condition satisfied by this particular sample? Why or why not?

6. (*8 pts*) Assume the standard deviation has been estimated previously from historical data and found to be . Using a significance of construct a confidence interval for the true mean of the population by following the steps in this problem:

a. Calculate the standard error. Save and label your answer.

b. Calculate the critical Z-score. Save and label your answer.

c. Calculate the margin error. Save and label your answer.

d. Find the upper and lower bound for the interval containing the true mean. Save and label your answer.

7. (*8 pts*) The historical estimate of the standard deviation in the previous part was found to be flawed. Instead, you will estimate the standard deviation from the sample itself. Using the same significance as in #5, find the new bounds for the confidence interval. Save and label all calculations needed to find the bounds.

a. In a cell, explain the difference between #5 and #6. Save and label your answer.

8. (*10 pts)* You want to test the claim the true mean of the race times for Kentucky Derby winners is greater than 2.16 seconds.

a. State the null hypothesis. Save and label your answer in a cell.

b. State the alternate hypothesis. Save and label your answer in a cell.

c. Calculate the test statistic to test the validity of this hypothesis. Save and label your answer in a cell.

d. Make a conclusion about the validity of this hypothesis. Is there evidence to support the claim the mean race time is greater than 2.16 seconds? Save and label your answer.

**EXTRA CREDIT**

9. (*5 pts*) As a diligent statistician, you notice something peculiar about this dataset. The winning time appears to be decreasing with time! You decide to create a time series analysis of the dataset by following the steps in this problem:

a. Create a scatter plot of the dataset using the year as the independent variable and the winning time as the dependent variable.

I. Insert a trend line and superimpose the equation of the line of best fit on the graph.

II. Appropriately label and scale all the axes and title the graph.

b. Aha! There is indeed a trend! Let us be more precise! In a cell, calculate the correlation of the year and the winning time.

c. Using a significance of , determine if the correlation found in the previous part is significant by referring to the table of critical values for the Pearson correlation coefficient.

d. If the relationship is significant, use the line of best fit to make a prediction about the winning time for the Kentucky Derby this year.