Stat 500 **Lab 3** **Fall 2020**

1. For continuous distributions, probabilities are defined as areas under the density curve. Using the ideas from calculus, we calculate these probabilities by integrating under the density curve. For all three of the distributions above, the integral of the density curve does not have a closed-form solution. Historically, probabilities for these distributions have been determined using published tables. However, now we have many resources available in order to find these probabilities and to learn about the distributions. In this lab, we will rely on web applets to work with these distributions.

A. Normal Distribution

A normal distribution is used to model continuous data when its histogram follows a bell-shaped curve. The parameters for the normal distribution are the mean μ and the variance σ^2. The standard normal distribution has mean $\mu = 0$ and variance $\sigma^2 = 1$.

Below is a link for an applet for the normal distribution. This link allows you to specify the mean and standard deviation of the normal distribution, a value and which direction to calculate the probability (lower-tail, upper-tail, two-tailed).

http://homepage.stat.uiowa.edu/~mbognar/applets/normal.html

Use the applet to find the solutions to the following questions:

Scores on a particular achievement test (Y) are assumed to have a normal distribution with mean $\mu = 75$ and variance $\sigma^2 = 100$.

a. Find the probability a randomly selected student will score more than 90 points on this achievement test. 0.44038
b. Find the probability a randomly selected student will score less than 70 points on this achievement test. 0.48006
c. 30% of all students taking this achievement test will score better than what value? 127.44005
d. What is the median score on the achievement test? 75
e. What is the 15th percentile score on the achievement test? 178.64334

B. Chi-square (χ^2) Distribution

The chi-square (χ^2) distribution belongs to the larger gamma family of distributions. The parameter for the chi-square (χ^2) distribution is the degrees of freedom, which is usually denoted as ν. Use the applet below to investigate the properties of the chi-square (χ^2) distribution.

http://homepage.stat.uiowa.edu/~mbognar/applets/chisq.html

a. In the applet, enter 5 for the value of the degrees of freedom. The density curve will appear with information about the mean and variance of the distribution below. What are the mean and variance of the distribution? Mean = 5, variance = 10
b. Change the value of the degrees of freedom to several other numbers. Look first at the mean and variance of the distribution as you change the value of the degrees of freedom. What is the relationship between the mean and variance of the distribution and the degrees of freedom? Mean = degree of freedom, variance = 2 df
c. Now look at the shape of the density curve as you change the value of the degrees of freedom. What is the relationship between the shape of the distribution and the degrees of freedom? The higher the df is, the similar the distribution shape to normal distribution

C. t Distribution

The t distribution arises due to the need to estimate the population variance in statistical inference. Much of the inference we do in STAT 500 will involve this distribution. Like the chi-square distribution, the parameter for the t distribution is called the degrees of freedom, which is denoted as ν.

The t distribution has a lot of properties in common with the standard normal distribution. However, the t distribution has more probability in the tails of its distribution than does the standard normal distribution. To study this difference, use the normal applet above and the t applet below to complete the following table. Round all values to 5 decimal places.

http://homepage.stat.uiowa.edu/~mbognar/applets/t.html

Probability greater than	Standard Normal Distribution (Z)	t Distribution with $\nu = 10$	t Distribution with $\nu = 25$	t Distribution with $\nu = 50$
1.5	0.06681	0.08225	0.07307	0.06995
2	0.02275	0.03669	0.02824	0.02547
2.25	0.01222	0.02409	0.01675	0.01444
2.5	0.00621	0.01572	0.00967	0.00787
2.75	0.00298	0.01024	0.00546	0.00414
3.0	0.00135	0.00667	0.00302	0.0021

What do you notice about the values in the table? How do the probabilities for the t distribution compare to the standard normal distribution? How do these probabilities for the t distribution change as the degrees of freedom increases?

The greater the observed values are, the smaller the probabilities are. The higher df is, the closer to standard normal.

2. For a school statistics poster competition in 2006, students timed 15 randomly selected teenagers from the school and 15 randomly selected staff from the school over the age of 30 on how long it took each person to text the following sentence on their phone: “the quick brown fox jumps over the lazy dog.” Each subject had the sentence in front of them while they were typing. The text message had to be typed with no errors, no abbreviations, and no use of the phone directory. Time was measured using a stop watch to within 0.01 seconds. Participants were timed using two phones – their own phone and a “control” phone, which was the same for all participants. The data are located in the file **smsspeed.csv** and the SAS code for the analyses below can be found in the file **smsspeed.sas**.

We would like to determine if teenagers were faster “texters”, on average, than adults.

First we will explore the differences in mean times when participants used their own phone (parts a-e).

a. From the output, find the difference in the two sample means $\bar{Y}_1 - \bar{Y}_2$ and the estimate of the pooled standard deviation S_p. Difference in the two sample means = 44.0120
Estimated pooled standard deviation = 18.5107

b. Use the values from part (a) to calculate the test statistic T for the hypothesis test. Use the value from the SAS output to verify your calculation.

$\frac{\text{Y1\_bar - Y2\_bar - (mu1 - mu2)}}{\text{Sp sqrt(1/n1 + 1/n2)}}$ = 44.012/(18.5107*sqrt(2/15)) = 6.511468

c. Give the degrees of freedom for the test statistic T. How was this value obtained?

df = 28 because there are 15 observations, 2 groups, so we "give" 2 dfs to the two group means.

d. The output gives the p-value for the two-sided hypothesis test. What is the p-value for the one-sided hypothesis test to find out whether the staff needs longer time than the teens? p-value/2

e. Give the decision and conclusion for this one-sided hypothesis test. Make sure to give your answer in the context of the problem. We have one sided p-value < 0.0001. Hence, at significance level of 0.05, we reject the null hypothesis. Then, the staff needs longer time than teens.

f. You can explore the differences in mean times when participants used the control phone. Use the output from SAS to conduct a hypothesis test for the difference in the mean times to send the text message between the adults and teenagers. Make sure that you can (i) state the Null Hypothesis and the Alternative Hypothesis, (ii) calculate or obtain from SAS output the Test Statistic and the P-value, and (iii) provide your decision for the hypothesis testing and your conclusion.