36-220 Lab #5
Normal Distribution, Central Limit Theorem

Please write your name below, tear off this front page and give it to a teaching assistant as you leave the lab. It will be a record of your participation in the lab. Please remember to include whether you are in Section A or B. Keep the rest of your lab write-up as a reference for doing homework and studying for exams.

Name:

Section:

- The symbol ♣ at the beginning of a question means that, after you answer that question, you should raise your hand and have either the TA or lab assistant review your answer. Once they have reviewed your work they will place a check in the appropriate space in the table below. The purpose of this check is to be sure you have answered the question correctly.
- **You should try to complete as much of the lab exercise as possible. We understand that students work at different paces and have tried to structure the exercise so that it can be completed in the allotted time. If you work systematically through the handout and still don't complete every question don't worry. The important thing is that you understand what you are doing. Nonetheless, you are encouraged to complete the lab on your own.**

Check-Problem ♣	Instructor's Initials
Question 3	
Question 12	

36-220 Lab #5
Gaussian Distribution, Central Limit Theorem

1 Gaussian Distribution

To generate random data from a Normal distribution, select **Calc** > **Random Data** > **Normal**. Generate 25 rows of data and store the results in column C1. Use a mean of 4 and a variance of 3. Label column C1 as "Norm1".
Make a histogram or stem-and-leaf plot of Norm1, using either **Graph** > **Histogram** or **Graph** > **Character Graphs** > **Stem-and-Leaf**.

Question #1: Describe the distribution of Norm1. Does Norm1 have the nice symmetric bellcurve of the Gaussian distribution?

Now, generate 300 rows of data from the normal distribution and store the result in column C2, using the same mean and variance. Label this column "Norm2". Make a histogram or stem-and-leaf plot of Norm2. If you make a histogram, set the number of intervals (bins) to be about the same number as used for Norm1 (click on the **Options** button and then enter the number of intervals).
Question #2: Describe the distribution of Norm2. Does this variable look like it was drawn from a normal distribution more so than Norm1? Why is that?

2 Probability Plots

A **probability plot** is a graphical way of comparing a sample of data to a particular distribution (often normal). The goal is to answer the following question: Is it reasonable to assume that this sample of data came from this distribution? The probability plot comapres the sample data, ordered from smallest to largest, to the corresponding quantiles of the theoretical distribution. If the plot is roughly a straight line, then it is reasonable to assume that the data come from that distribution.
The MINITAB command for making a probability plot is **Graph** > **Probability Plot**. You specify the theoretical distribution being tested by clicking the

button **Distribution...** and choosing from the list. Also, when you press **Distribution...** there is a second tab titled **Data Display**. Under that uncheck the box **Show confidence interval**.
Make probability plots for the data "Norm1" and "Norm2".
♣ **Question #3: How does the probability plot for Norm1 compare to the plot for Norm2? How well do the observations fall along a straight line in each of the two plots?**

Create in column C3 a data set of 300 data values simulated from the exponential distribution, $\lambda = 0.25$. Do this by using the **Calc** > **Random Data** > **Exponential** function.
Make a probability plot for the data set in column C3, to check if it is reasonable to assume that it comes from a Gaussian distribution. (Forget the fact that you know what distribution was used to create the data.)
Question #4 : What do you conclude? Is it reasonable to assume that the third sample came from a Gaussian distribution?

3 Central Limit Theorem

We are going to generate data from a Uniform(0,1) distribution. A Uniform(0,1) distribution has the following density function:

$$f(x) = \begin{cases} 1 & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Question #5: Make a sketch of the density function (*Note: You can't use Minitab for this*)

To generate random data from a Uniform(0,1) distribution, select **Calc** > **Random Data** > **Uniform**. Indicate that you want 500 rows of data stored in columns C1-C30. Make sure that the lower endpoint is 0.0 and the upper endpoint is 1.0. Click **OK**. Each row of your worksheet is a random sample of 30 observations from a Uniform(0,1) distribution. You have 500 rows, so you have 500 such samples.
Make a histogram or stem-and-leaf plot of the data in column C1.
Question #6: Does the plot match the density you drew in the previous question?

Make a probability plot of the data in column C1. *Note: Keep the "Distribution" selection as "Normal"*. (This may take a few seconds.)

Question #7: Sketch a copy of the probability plot shown via MINITAB.

Select **Stat** > **Basic Statistics** > **Display Descriptive Statistics**. Select C1 and click **OK**.
Question #8: How do the sample mean and sample standard deviation of the data in column C1 compare to $E(X) = \frac{1}{2}$ and $\sqrt{V(X)} = \sqrt{\frac{1}{12}} = 0.289$?

Select **Calc** > **Row Statistics**. Under "Statistic," select "Mean." Under "Input Variables" type "C1-C10." In "Store result in," type C31. As a result, the first entry in column C31 will be the average of the first entries in columns C1-C10, the second entry in column C31 will be the average of the second entries in columns C1-C10, etc. Label column C31, *XBAR10*.

Now select **Calc** > **Row Statistics**. Under "Statistic," select "Mean." Under "Input Variables" type "C1-C30." In "Store result in," type C32. This will be

similar to *XBAR10*, except that C32 will be element-wise averages of columns C1-C30.
Label column C32, *XBAR30*.
The net result is that each observation of *XBAR10* is the mean of 10 independent Uniform(0,1) observations, and each observation of *XBAR30* is the mean of 30 independent Uniform(0,1) observations. So you have 500 samples of *XBAR10* and 500 samples of *XBAR30*.
We'd like to compare the distributions of *XBAR10* and *XBAR30*. To help us do so, we'll tell MINITAB to use the same number of intervals for each histogram. When you create the histogram, select **Options**. Then specify that you want 25 intervals. Make such a histogram for the data in *XBAR10* and another one for the data in *XBAR30*.
Question #9: Describe the distributions of *XBAR10* and *XBAR30*. How do they compare to the distribution of the Uniform data from column C1?

Question #10: Where are the distributions centered? Does this make sense? Why?

Question #11: Compare the spreads of the two distributions. Which variable has the smaller variance, *XBAR10* or *XBAR30*?

Make probability plots of *XBAR10* and *XBAR30*.

♣ **Question #12: Do the distributions of *XBAR10* and *XBAR30* look close to normal? Should they? Why?**

Question #13: We have learned that if $X_1, X_2, ..X_n$ are independent random variables drawn from the same distribution and $Var(X) = \sigma_x^2$,

then:

$$Var(\overline{X}) = Var(\frac{\sum_{i=1}^{n} X_i}{n}) = \frac{\sigma_x^2}{n}$$

Recall that this is a *population* variance (Note: the standard error is the standard deviation of a statistic). It is the variance of the distribution of $\overline{X}$. Using this formula, what is population variance of *XBAR10*? What is the population variance of *XBAR30*? What are their population standard errors (i.e., their standard deviations)?

Use **Statistics** > **Basic Statistics** > **Descriptive Statistics** to find the *sample* standard deviations of *XBAR10* and *XBAR30*.

Question #14: How do the sample standard deviations compare to the population values?