STAT 2507
Assignment # 3
(Chapters 7 & 8)

DUE: Sections E, F **Monday, March 16, in class**
Section G **Tuesday, March 17, in class**
Section H **Wednesday, March 18, in class**

Last Name____________________________________ First Name_____________________

Student #______________________________ Your LAB Section_______________

Part I Lab Questions

1. [8] To illustrate that sample statistics such as $\overline{X}$ and S^2 really do vary from sample to sample: Use the following Minitab instructions to generate 10 random samples, each of size $n = 5$, from a normal distribution with $\mu = 20$ and $\sigma^2 = 16$. Then calculate the sample mean $\overline{x}$ and the sample variance s^2 for each of the 10 samples. **Print out your results to be handed in with the assignment,** and also record the values of $\overline{x}$ and s^2 (in c6 & c7) in the table shown after the Minitab instructions.

Minitab Instructions

calc – random data – normal Generate **10** rows of Data Store in Columns **c1-c5**
Mean: 20 **Standard Deviation: 4**

Calc – row statistics – mean, variance Input Variables: **c1 – c5** Store results in **c6,c7**

Table of $\overline{x}$ and s^2

$\overline{x}$										
s^2										

Questions

a) [1] Are the values for $\overline{x}$ and s^2 the same for all 10 samples? __________
b) [3] Why is this to be expected? ____:
____
c) [4] Would you expect to get the same values if you reran your Minitab program?___________
Explain: ____

2. [12] We know that if $X_1, X_2, \ldots, X_n$ is a random sample from any distribution with mean μ and variance σ^2, then the sampling distribution of the sample mean $\overline{X}$ has mean μ and variance σ^2/n. We further know that if the sample is from a normal distribution then the sampling distribution of $\overline{X}$ is also normally distributed.

The Central Limit Theorem states that if $X_1, X_2, ..., X_n$ is a random sample of size n from **any** distribution with mean μ and variance σ^2, the distribution of $\overline{X}$ gets closer and closer to a normal distribution as the sample size n gets larger and larger. That is, $Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$ approaches a standard normal distribution as the sample size n approaches ∞.

If the distribution of the population is roughly symmetric and either flat or mounded then $\overline{X}$ will be approximately normally distributed for sample sizes $n > 30$. However, for some very skewed distributions or U-shaped distributions n must be considerably larger than 30 for $\overline{X}$ to be even approximately normally distributed.

The Poisson is a discrete distribution depending on a parameter μ. For small values of μ it is very right skewed. (See histogram for Poisson = $\mu = 0.5$ in text p 207.)

In this lab question, you will use MINITAB to generate 500 random samples:
(A) of size 5 from a Poisson distribution with $\mu = 0.1$
(B) of size 30 from a Poisson distribution with $\mu = 0.1$
(C) of size 100 from a Poisson distribution with $\mu = 0.1$
For each case have MINITAB calculate the sample mean $\overline{X}$ of each of the 500 samples , and then construct a histogram of these values of $\overline{X}$. **Hand these 3 histograms in with the assignment**.

Minitab Instructions

(A)

1) **To generate 500 random samples of size 5 from a Poisson Distribution $\mu = 0.1$**

calc - Random Data - Poisson Generate **500** Rows of Data Store in Column(s): **c1–c5**

Note: There are 500 rows: the 5 values for sample 1 in row 1, the 5 values for sample 2 in row 2, ...

2) **To Calculate the Sample Mean $\overline{X}$ for each of the 500 samples**

calc - row statistics - mean Input Variables: **c1 – c5** Store Results in: **c6**
Label column c6 something like **Xbar5P**

3) **To Graph a Histogram of the 500 Sample Means**

Graph – Histogram- simple Graph Variables: **c6**

4) **Copy and paste the histogram into a Word Document**

(B)

1) **To Generate 500 random samples of size 30 from a Poisson distribution with $\mu = 0.1$**

calc – Random Data – Poisson Generate **500** rows of Data Store in Columns: **c1-c30**
Mean: **0.1**

2) Repeat steps 2 – 4 above, but label the **c31** column **Xbar30P**

(C)

Repeat steps 1 - 4 for 500 random samples of size **100** from a Poisson Distribution with $\mu = 0.1$, making appropriate changes in the instructions.

Questions

a) [1] Does the histogram of the sample means based on samples of size 5 from (A) look approximately bell-shaped? _______________

b) [3] Why is this expected? ____

c) [1] Does the histogram of the Poisson sample means based on samples of size 30 look approximately bell-shaped? _________________________

d) [3] Why might this be expected? ____

e) [1] Does the histogram of the Poisson sample means based on samples of size 100 look approximately bell-shaped? _____________

3. [5] Use Minitab to randomly select 30 samples of size n = 9 from a normal population with mean $\mu = 69$, and variance $\sigma^2 = 16$, and calculate the 90% confidence interval estimate of μ from each sample.

Minitab Instructions

Calc – Random Data - normal Generate **9** Rows of Data Store in Columns: **c1-c30**
Mean: **69** Standard Deviation: **4**
Note: The 30 samples of size n = 9 are in **columns** not rows.

Stat – Basic Statistics – 1-Sample Z Samples in Columns: **c1-c40**
Standard Deviation: **4**
Options – Confidence level: **90** Alternatives: **Not Equal**

Copy and paste these confidence intervals into a Word document **to be handed in with assignment.**

Questions

a) [1] The value of $z_{\alpha/2}$ used to compute the interval is: ______________

b) [1] The length of the interval is: ____________

c) [1] The number of intervals over the 30 replications that actually contain μ is: _____________

d) [3] Compare the percentage of the intervals that contain μ to the stated confidence level of 90%:
____

Part II Long Answer Questions

1. [17] The heights of North American women are normally distributed with a mean of 64 inches and a variance of 4 inches. What is the probability that

   a) [4] a randomly selected woman is taller than 66 inches?
   b) [4] the mean height of a random sample of 4 women is greater than 66 inches?
   c) [4] the mean height of a random sample of 100 women is greater than 66 inches?
   d) [5] If the population of women's heights is NOT normally distributed, which, if any, of (a), (b), (c) can you answer? Explain.

2. [13] The number of customers per week at each store of a supermarket chain has a population mean of 5000 and a standard deviation of 500. If a random sample of 64 stores is selected

   a) [5] What is the approximate probability that the sample mean will be between 4,980 and 5,075 customers per week?
   b) [4] What is the approximate probability that the sample mean will be below 4,075 customers per week?
   c) [4] What is the approximate probability that $\overline{X}$ lies within 75 customers of the true mean number of customers, μ?

3. [10] In a survey conducted to determine, among other things, the cost of vacations taken by single adults in a particular region, 144 individuals were randomly sampled. Each person was asked to assess the total cost of his or her most recent vacation. The average cost was $2,386 and the standard deviation was $400.

   a) [5] Estimate with a 99% confidence interval the average cost of a vacation trip for all single adults in the region.
   b) [5] How large a sample would have to be taken to estimate the true average cost to within $60 with 99% confidence? Use the sample standard deviation to estimate σ.

4. [3] *Consumer Research* reported information on the time required for caffeine from products such as coffee and soft drinks to leave the body after consumption. Suppose that the 95% confidence interval estimate of the population mean time for adults is from 5.6 hours to 6.4 hours. What is the <u>point estimate</u> of the mean time for caffeine to leave the body of adults after consumption?

5. [10] A new product was test marketed in Toronto and in Vancouver. Equal amounts of money were spent on advertising in the two areas, but different advertising media were used. Advertising in the Toronto area was done entirely on television, while in Vancouver it consisted of a mixture of television, radio and billboards. Two months after the advertising campaign began, surveys were taken to estimate consumer awareness of the product. In the Toronto area, 631 out of 1,000 randomly selected consumers were aware of the product, while in the Vancouver area, it was 798 out of 1,000 randomly selected consumers.

   a) [5] Estimate with 95% confidence the difference between the proportion of Toronto area consumers who were aware of the product and the proportion of Vancouver area consumers who were aware of it.
   b) [5] Interpret this interval.

6. [10] The farm-equipment manufacturer wants to compare the average daily downtime for two sheet-metal stamping machines located in two different factories. Investigation of company records for 100 randomly selected days on each of the two machines gave the following results:

machine I: average downtime of 12 minutes with a variance of 6
machine II: average downtime of 9 minutes with a variance of 4

a) [5] Estimate with approximate 90% confidence the difference between the average downtime for machine I and that for machine II.
b) [5] Interpret you results in part (a) and draw a conclusion about the difference between the true average downtimes for the two machines.

7. [12] In a survey of 250 voters prior to an election, 40% indicated that they would vote for the incumbent candidate.

a) [4] Estimate with a 90% confidence interval the population proportion of voters who support the incumbent.
b) [4] Using the sample estimate of p from the above study, how large a sample is necessary to estimate the proportion of voters who favour the incumbent to within 0.04 with 90% confidence?
c) [4] Repeat (b) assuming that no previous study has been conducted.