Stat 2507 **Assignment 3 Solution** **Summer 2016**

Due: Monday, July 11, 2016 in the class @ 6:05pm-7:30pm

You should pick up your marked assignment from the TA during your lab time
Assignment 3 Solution has 8 questions, for a total of 65 marks

The marking scheme is as follows:

Question:	1	2	3	4	5	6	7	8	Total
Marks	5	12	5	5	13	5	5	15	65
Score:									

Activate "Enable command" from Editor on the toolbars menu when the session window is active.

Question 1) . . . . *5 marks*

To illustrate that sample statistics such as $\bar{X}$ and S^2 really do vary from sample to sample: Use Minitab to generate 15 random samples, each of size $n = 5$, from a$N(\mu = 10, \sigma^2 = 25)$. Calculate the sample mean, $\bar{X}$ and the sample variance S^2, for each of the 10 samples.

Type the following Minitab commands to generate numbers

```
MTB > Base 3.
MTB > Random 15 C1-C5;
SUBC> Normal 10 5.
MTB > RMean C1-C5 C6.
MTB > RStDev C1-C5 C7.
MTB > Formula C8=C7^2
```

Print out your results to be handed in with the assignment. Explain your results?

↓	C1	C2	C3	C4	C5	C6	C7	C8
	X1	X2	X3	X4	X5	Mean	Std	Var
1	9.8078	6.3579	15.6179	8.5161	9.5721	9.9744	3.43673	11.8111
2	7.8409	10.2101	10.2880	4.4359	8.8300	8.3210	2.39903	5.7554
3	2.9807	16.3477	8.0394	16.2037	8.2524	10.3648	5.79405	33.5710
4	9.2606	7.3223	10.8959	0.6359	4.6962	6.5622	4.03867	16.3108
5	11.2035	8.3201	-0.2367	0.3138	-0.1108	3.8980	5.45295	29.7347
6	11.8184	6.0989	12.8109	11.1345	14.6812	11.3088	3.20400	10.2656
7	9.9267	14.5971	13.2338	7.6199	12.1138	11.4982	2.76158	7.6263
8	-3.3582	1.5640	7.7116	18.7395	11.8647	7.3043	8.63666	74.5918
9	12.4237	4.8894	12.6478	18.2371	10.9215	11.8239	4.77153	22.7675
10	11.2468	4.4457	5.0139	2.4269	8.0365	6.2340	3.44811	11.8894
11	20.3165	0.2317	7.3538	5.5887	11.0738	8.9129	7.47587	55.8886
12	26.7570	10.9171	5.7963	9.1204	10.4845	12.6151	8.15668	66.5314
13	13.8579	18.2926	12.1840	9.6169	7.8393	12.3581	4.04450	16.3580
14	14.1819	9.7468	-2.8447	7.2988	6.9367	7.0639	6.24729	39.0286
15	13.4909	13.3158	15.2935	10.2089	7.3952	11.9409	3.13036	9.7992

$\bar{X}$ and s^2 are different from sample to sample because sample size is very small

Question 2) . . . . *12 marks*

We are going to study the validity of the CLT, via simulation.

Definition 0.1 (CLT):
A random sample $X_1, \cdots, X_n$ from any distribution with μ and σ^2, then the sampling distribution of the sample mean $\bar{X}$ has mean μ and variance $\frac{\sigma^2}{n}$. CLT states that:

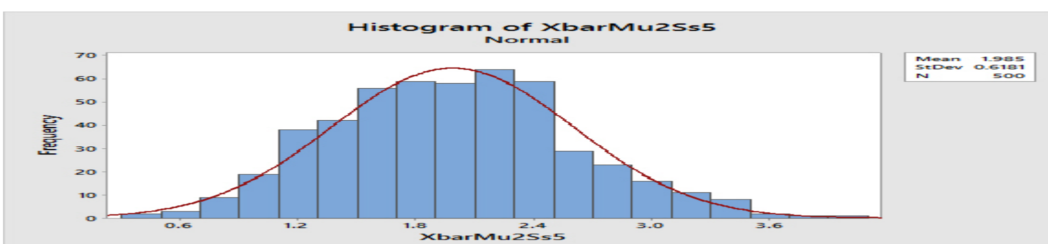
1. If the sample is from a normal distribution then the sampling distribution of $\bar{X}$ is also normally distributed regardless of the sample size.
2. If the distribution of the population is roughly symmetric or mound-shaped, then the distribution of $\bar{X}$ will be approximately normally distributed regardless of the sample size.
3. If the distribution of the sampled population is skewed, then the sample size of n must be at least 30 before the sampling distribution of $\bar{X}$ becomes approximately normal.
4. If the distribution of the sampled population is very skewed or U-shaped,then the sample size must be much larger than 30, before the sampling distribution of $\bar{X}$ becomes approximately normal.

In this lab question, Use MINITAB to generate 500 random samples and hand in your histogram for $\bar{X}$ with an explanation of distribution:

```
MTB > Random 500 C1-C(n);
SUBC> Poisson mu.
MTB > RMean C1-C(n) C(n+1).
MTB > name C(n+1) 'XbarMu()Ss(n)'
MTB > Histogram 'XbarMu()Ss(n)';
SUBC> Bar;
SUBC> Distribution;
SUBC> Normal.
```

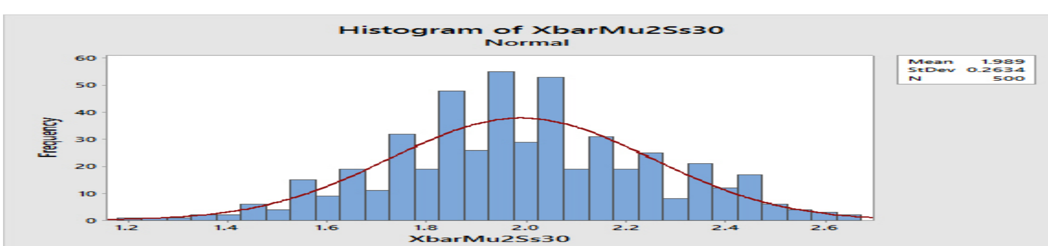
Using the code above, only change the value in the bracket to what is needed below (Make sure you remove the brackets)

(a) (2½ marks) of size 5 from a Poisson distribution with $\mu = 2$. Does the histogram look approximately bell-shaped? Explain.



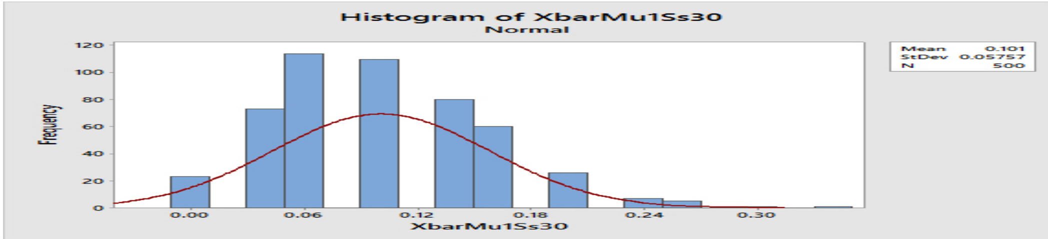
It is not too bell-shaped. If you look closely you will see it is skewed to the right because the sample size is very small $n = 5$

(b) (2½ marks) of size 30 from a Poisson distribution with $\mu = 2$. Does the histogram look approximately bell-shaped? Why is this expected?



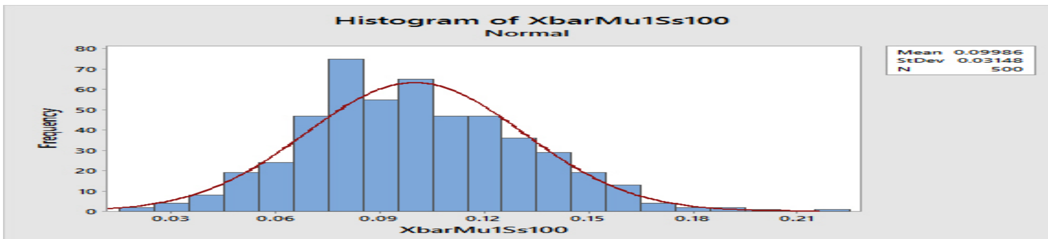
Yes, it is bell-shaped. This is expected because of CLT.

(c) (2½ marks) of size 30 from a Poisson distribution with $\mu = 0.1$ Does the histogram look approximately bell-shaped? Why is this expected?



No. it is not bell-shaped. The Poisson distribution with $\mu = 0.1$ is very skewed.

(d) (2½ marks) of size 100 from a Poisson distribution with $\mu = 0.1$ Does the histogram look approximately bell-shaped?



somewhat looks bell-shaped.

(e) (2 marks) Histogram from (b) and (c) are obtained using the same sample size. Provide the reason for the difference.

The Poisson distribution with $\mu = 0.1$ is a very skewed distribution so n has to be quite large before $\bar{X}$ becomes approximately normal by the CLT.

Question 3) .... *5 marks*

We are going to study the validity of the CLT, via simulation.

Suppose that X has a uniform distribution on $(0, 1)$, that is, X is a continuous r.v with pdf $f(x) = 1$, if $x \in [0, 1]$. The steps are:

1. Step 1 generate 6 independent random samples of size 200 observations from the distribution of X, and save them in the vectors $c1, c2, c3, c4, c5, c6$.
2. Step 2 Find the means (i.e. $\bar{x}$) for samples of size $n = 1, 2, 3, 4, 5, 6$ and save them in the vectors $c11, c12, \cdots, c16$.

Using Minitab code

```
MTB >random 200 c1-c6;
SUBC> uniform 0, 1.
MTB > rmean c1 c11
MTB > name C11 'C11(n=1)'
MTB > rmean c1-c2 c12
MTB > name C12 'C12(n=2)'
MTB > rmean c1-c3 c13
MTB > name C13 'C13(n=3)'
MTB > rmean c1-c4 c14
MTB > name C14 'C14(n=4)'
MTB > rmean c1-c5 c15
MTB > name C15 'C15(n=5)'
MTB > rmean c1-c6 c16
MTB > name C16 'C16(n=6)'
```

```
MTB > Histogram 'C11(n=1)' 'C12(n=2)' 'C13(n=3)' 'C14(n=4)' 'C15(n=5)' 'C16(n=6)';
SUBC> Bar;
SUBC> Distribution;
SUBC> Normal.
```

Submit your plots on one/two pages only.

(a) (1 mark) Which of the 6 histograms is very different from the normal distribution? **C11(n=1)** (Choose one of $C11, C12, \cdots, C16$.)

(b) (2 marks) Which one looks most normal distribution? **Any one other than C11 is ok**

(c) (2 marks) Suppose you have a vector, called C30, that contains 200 numbers, each of which is the average of 30 independent observations from a uniform $(0, 1)$ distribution. Which one would you expect to look more normal: the histogram of C16 or that of C30? **C30**. Why? **CLT**

Question 4) . . . . *5 marks*

Suppose that $X \sim N(5, \sigma^2 = 9)$. Now repeat steps 1 and 2 of Question 3 and answer the following questions. **Note:** The only thing that changes in the above set of commands is 'uniform 0, 1.' which becomes 'normal 5, 3.'

Submit the histograms on one/two page(s) only

(a) (1 mark) Which histogram looks more normal? **All**. Which one do you expect to look more normal? **All**. Why? **Sample from normal**

(b) (2 marks) Let C40 be the vector that contains 200 numbers, each of which is the average of 40 independent observations from a $N(5, 9)$ distribution. Which one do you think has a smaller variance: C11 or C40? **C40**. Why? $\sigma^2/n = 1/40$ **as compared to 1**.

(c) (2 marks) Let C40 be as in part (b) above. What can you say about the mean of the values in C11 as compared to the mean of the values in C40?

○ Larger, ○ smaller, √ **nearly the same**

Question 5) . . . . *13 marks*

Suppose that the waiting time for a pizza to be delivered to an individual's residence has been found to be normally distributed with a mean of 30 minutes and a standard deviation of 8 minutes. What is the probability that a randomly selected individual will have a waiting time:

(a) (3 marks) Between 15 and 45 minutes?

Solution:
$Pr(15 \leq X \leq 45) = P\left(\frac{15-30}{8} \leq Z \leq \frac{45-30}{8}\right) = P(Z \leq 1.88) - P(Z \leq -1.88) = 0.9699 - 0.0301 = 0.9398$

(b) (2 marks) At least 10 minutes?

Solution:
$Pr(X \geq 10) = P\left(Z \geq \frac{10-30}{8}\right) = P(Z \geq -2.5) = 1 - P(Z \leq -2.5) = 1 - 0.0062 = 0.9938$

(c) (2 marks) At most 10 minutes?

Solution:
$Pr(X \leq 10) = P\left(Z \leq \frac{10-30}{8}\right) = P(Z \leq -2.5) = 0.0062$

(d) (2 marks) Less than 10 minutes?

Solution:
$Pr(X < 10) = Pr(X \leq 10) = 0.0062$

(e) (2 marks) No more than 45 minutes?

Solution:
$Pr(X \leq 45) = P\left(Z \leq \frac{45-30}{8}\right) = P(Z \leq 1.88) = 0.9699$

(f) (2 marks) More than 45 minutes?

Solution:
$Pr(X \geq 45) = Pr(Z \geq 1.88) = P(Z \leq -1.88) = 0.0301$

Question 6) .... *5 marks*

The batting average of all players in a professional baseball league is normally distributed with the mean of 0.27 and standard deviation of 0.01. A players' contract comes up for renewal. In order to have a good bargaining position, he would like to have a batting in the top 1% of all players. What batting average must be attained?

Solution:
$X \sim N(\mu = 0.27, \sigma^2 = 0.01^2)$

we want to find x_0 such that $P(X > x_0) = 0.01$ or $P(X < x_0) = 0.99$, then

$$P(X < x_0) = P\left(Z \leq \frac{x_0 - 0.27}{0.01}\right) = 0.99$$

From $Z - Table \Rightarrow \frac{x_0 - 0.27}{0.01} = 2.33 \Rightarrow x_0 = 2.33(0.01) + 0.27 = 0.2933$

Question 7) .... *5 marks*

A company manufactures washers, about 5% of which are defective. If a random sample of 100 washers are inspected, what is the probability that fewer than 4 are defective?

Solution:

Let X be the number of defective washers, so $X \sim Bin(100, 0.05)$ with $\mu = np = 100(0.05) = 5 \geq 5$ and variance $= \sigma^2 = npq = 100(0.05)(0.95) = 4.75$. Since $np \geq 5$ and $nq = 100(0.95) = 95 > 5$, then X can approximate by normal distribution $N(5, 4.75)$ and using continuity correction

$$P(X < 4) = P\left(Z \leq \frac{4 - 0.5 - 5}{\sqrt{4.75}}\right) = P(Z < -0.69) = 0.2451$$

Question 8) .... *15 marks*

A computer supply house receives a large shipment of computer disks each week. Past experience has shown that the number of flaws per disk can be described by the following probability distribution, where the r.v. X = number of flaws per computer disk.

x	0	1	2	3
$P(X = x)$	0.65	0.2	0.1	0.05

(a) (2 marks) Calculate the mean and standard deviation of X

Solution:

$\mu_x = \mathrm{E}(X) = \sum_{x=0}^{3} xP(X = x) = 0.2 + 0.2 + 0.15 = 0.55$ and

$\sigma_x^2 = \sum_{x=0}^{3} x^2 P(X = x) - \mu_x^2 = 0.2 + 0.4 + 0.45 - (0.55)^2 = 0.7475; \sigma_x = \sqrt{0.7475} = 0.865$

(b) (1 mark) Describe the shape of the distribution of X. **skewed to the right**

(c) (2 marks) Suppose that we randomly select a sample of 100 disks. Describe the shape of the sampling distribution of the sample mean $\bar{X}$. Explain.

Solution:
Based on CLT, the distribution of $\bar{X}$ is normal, with $\mathrm{E}(\bar{X}) = \mu = 0.55$ and
$\mathrm{Var}(\bar{X}) = \frac{\sigma_x^2}{n} = \frac{0.7475}{100} = 0.007475$

(d) (2 marks) Find the probability that the sample mean, $\bar{X}$, of the sample of 100 disks is less than 0.4?

Solution:
$P(\bar{X} < 0.4) = P\left(Z \leq \frac{0.4-0.55}{\sqrt{0.007475}}\right) = P(Z < -1.73) = 0.0418$

(e) (2 marks) Find $P(X < 0.4) = \underline{P(X = 0) = 0.65}$