STAT 2507
Assignment # 4 (Solutions)
Due: Thursday, November 20, 2014, in class

Last Name ______________________________ **First** ______________________________

Student Number ____________________ **Lab Section (Important)** _______________

Total of marks=100. Marks for each question are given in [ ]

Note: Use spaces left to answer all questions. No need to submit plots

Part I. Minitab Questions

1. **[14]Central limit theorem (CLT)** (You can open a new Minitab worksheet, simply by typing *new*). Generate 80 horizontal samples, each of size $n = 1000$ from exponential distribution with mean $\mu = 9$ and store in columns c3-c1002 as follows:
   random 80 c3-c1002;
   expo 9.
   Note: This may take a few moments as you are generating 1000×80=80,000 values!
   Create and store in column c1 the 80 values of $\bar{x}$ based on the 80 horizontal samples, each of the same size $n = 1000$ as follows:
   rmean c3-c1002 c1
   [2] a. Generate the boxplot of the first sample c3. According to the median position and/or the outliers, what can you conclude about the shape of this data set? **Skewed right**
   [4] **b.** Use *desc* command to find sample mean **9.62** and median **5.19** of c3. Do they confirm your diagnostic for the shape above? **Yes, mean is larger than median**
   [2] c. Generate the histogram for the data in column c1. What can you conclude about the shape of data in c1? **fairly symmetric**
   [3] d. Use *desc* to find sample mean **9.0355** and sample standard deviation **0.282** of c1. Are they close to 9 and $9/\sqrt{1000}$? **Yes** Why? **Expected value of sample mean is equal to $\mu = 9$ and standard error of $\bar{X}$ is equal to** $\frac{\sigma}{\sqrt{n}} = \frac{9}{\sqrt{1000}} = 0.285$
   [3] e. Do the values of the mean and median of c1 confirm your conclusion in (**c**) about the shape of data in c1? **median = 9.0533** Why? **As median ≈ mean, the shape of the histogram must be symmetric. Hence, this confirms our conclusion in part (c)**

2. **[6]Confidence interval (CI) for a population mean:** We want to build 120 confidence intervals (CIs) with confidence level $(1 - \alpha)100\% = 95\%$ for the mean μ of a Poisson distribution via the following steps:
   Step 1. Open a new worksheet. Generate and store in columns c6-c505 120 samples of size

500 each, from Poisson with parameter $\mu = 7$ as follows:

random 120 c6-c505;

poisson 7.

Step 2. Use columns c4 and c5 to store respectively the means and the standard deviations of the 120 horizontal samples you generated in step 1, as follows:

rmean c6-c505 c4

rstd c6-c505 c5

Step 3. Store the lower bound and the upper bound of each of your 95% confidence interval in c2 and c3 respectively by typing successively:

*let c2=c4-1.96*c5/sqrt(500)*

*let c3=c4+1.96*c5/sqrt(500)*

Step 4. Then create a column c1 containing 1 or 0 according to whether the corresponding interval [c2 , c3] covers μ or not by using the following Minitab command:

let c1=(c2 <= 7 and c3 >= 7)

Finally sum up the entries of column c1 to find out how many confidence intervals did cover the value $\mu = 7$ by typing:

tally c1

[3] a. How many confidence intervals that did contain the true value $\mu = 7$? **113 out of 120**

[3] b. How do you compare this number to the confidence level 95%? **113/120 = 0.942 It is very close to 0.95 as expected.**

Part II. Long-answer Questions

1. **[15]**Suppose a random sample of $n = 40$ observations is selected from a population that has normal distribution with mean 105 and standard deviation 10.

   [5] a. Find the mean and the standard deviation of the sample mean $\bar{X}$.

   [5] b. Find the probability that $\bar{X}$ exceeds 108.

   [5] c. Find the probability that the sample mean deviates from the population mean by less than 3.

   Solution:

   a. $E(\bar{X}) = 105$ and $SE\,(\bar{X}) = \frac{10}{\sqrt{40}} = 1.581$

   b. X is already normally distributed. Hence, regardless of the sample size by CLT $\bar{X}$ is normally distributed with mean of 105 and standard deviation (or standard error of $\bar{X}$) of 1.581.

   $$P(\bar{X} > 108) = P(Z > \frac{108 - 105}{1.581}) = P(Z > 1.90) = 1 - P(Z \leq 1.90) = 1 - 0.9713 = 0.0287$$

   c. Standardize by dividing by $SE(\bar{X})$

   $$\begin{aligned} P(|\bar{X} - \mu| < 3) &= P(|Z| < \frac{3}{1.581}) \\ &= P(-1.9 < Z < 1.9) \\ &= P(Z < 1.9) - P(Z < -1.9) = 0.9713 - 0.0287 = 0.9426 \quad or \\ &= P(Z < 1.9) - (1 - P(Z < 1.9)) = 2 * 0.9713 - 1 = 0.9426 \end{aligned}$$

2. **[15]**By statistics, faculty with rank of assistant professor finishing their second year of employment at a higher education institution in Ontario earn an average of \$65,500 per year with a standard deviation of \$3500. In an attempt to verify this salary level, a random sample of 64 assistant professors with two years of experience was selected from a personnel database for all higher education institutions in Ontario.

   [5] a. Describe the sampling distribution of the sample mean, $\bar{X}$, of the average salary of these 64 assistant professors.

   [5] b. Within what limit would you expect the sample mean to fall with probability 0.95

   [5] c. Obtain the probability that $\bar{X}$ is greater than 66,000.

   Solution:

   (a) As $n = 64 > 30$, by the CLT, $\bar{X}$ is approximately normally distributed with mean of 65,500 and standard deviation of $\bar{X}$, $SE\,(\bar{X})= \sigma/\sqrt{n} = 3500/\sqrt{64} = 437.5$ (i.e. $\bar{X} \sim N(65500, 437.5^2)$).

   (b) Since $n = 64 > 30$, $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} = \frac{\bar{X}-65500}{400}$ has approximately standard normal distribution. The limit that we would expect sample mean to fall with probability 0.95, is

   $$\mu \pm 1.96\sigma/\sqrt{n} = 65500 \pm 1.96 \times 437.5 = 65500 \pm 857.5 = [64642.5, 66357.5]$$

   (c) $P(\bar{X} > 66,000) \approx P(Z > \frac{66,000-65500}{437.5}) = P(Z > 1.14) = 1 - P(Z \le 1.14) = 1 - 0.8729 = 0.1271$

3. **[10]**In a report on why e-shoppers abandon their on-line sale transactions, a study found that "pages took me too much time to load" and "site was too confusing to me so that I could not find the product" were the two main complaints heard most often. Based on 50 customers' responses, the average time to complete on-line order was 5.3 minutes and the standard deviation was 2.5 minutes. Construct an 85% confidence interval for μ, the average completion time for an online order.

   Solution:

   Sample mean, $\bar{X} = 5.3$ and $s = 2.5$. As the sample size $n = 50 > 30$, We can use large sample confidence interval. $\alpha = 1 - 0.85 = 0.15 \Rightarrow (1 - \alpha/2) = 0.925 \Rightarrow Z_{\alpha/2} = 1.44$

   85% large sample confidence interval for mean μ: $\bar{X} \pm Z_{\alpha/2}\left(\frac{S}{\sqrt{n}}\right) = 5.3 \pm 1.44\left(\frac{2.5}{\sqrt{50}}\right) = 5.3 \pm 0.509 = (4.791, 5.809)$

4. **[10]**In a poll of 900 randomly selected adults, 250 indicated that movies are getting better. Construct a 95% confidence interval for the overall proportion of adults who say that movies are getting better.

   Solution:

   Point estimator for proportion of adults who say that movies are getting better, $\widehat{p} = \frac{250}{900} = 0.278$

   $n\widehat{p} = 250$ and $n\widehat{q} = 650 > 5$. So, large sample 95% confidence interval for population proportion is as follows:

   $$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{900}} = 0.278 \pm 1.96\sqrt{\frac{0.278(1-0.278)}{900}} = 0.278 \pm 0.029 = [0.249, 0.307]$$

5. **[10]**(choosing sample size) Suppose you wish to estimate the mean pH of rainfalls in an area that suffers heavy pollution due to the discharge of smoke from a power plant. Previous studies showed that the standard deviation is in the neighborhood of 0.4 pH, and you wish your estimate to lie within 0.1 of the unknown mean μ, with probability 0.90. Approximately how many rainfalls must be included in your sample?

   Solution:

   Margin of Error $= B = Z_{\alpha/2} \times \left(\frac{\sigma}{\sqrt{n}}\right) \leq 0.1 \Rightarrow n \geq \left(Z_{\alpha/2} \times \left(\frac{\sigma}{0.1}\right)\right)^2 = \left(1.645 \times \frac{0.4}{0.1}\right)^2 = 43.29$

   $\Rightarrow n > 43.29$

   Therefore, sample size can be 44.

6. **[10]**Assume that you have a 95% confidence interval for a population mean based on a sample size of $n_1 = 40$. If you wish to have a confidence interval of the same confidence level but with a length which is one fourth of the one you already have, then what would the sample size of n_2 be?

   Solution:

   $n_1 = 40 > 30 \Rightarrow$ length of large sample 95% confidence interval for mean, $l_1 = 2 \times Z_{\alpha/2}\frac{\sigma}{\sqrt{n_1}}$

   Length of large sample 95% confidence interval for mean when sample size is $n_2 \Rightarrow$

   $l_2 = 2 \times Z_{\alpha/2}\frac{\sigma}{\sqrt{n_2}}$

   But $4 \times l_2 = l_1 \quad \Rightarrow \quad \frac{4}{\sqrt{n_2}} = \frac{1}{\sqrt{n_1}} \quad \Rightarrow \quad n_2 = 16 \times n_1 = 640$

   We need to have sample size of 640 to get one fourth the length of the confidence interval of same confidence level.

7. **[10]** An experiment was conducted to test the effect of a new drug on a viral infection. The infection was induced in 100 mice, and the mice were randomly split into two groups of 50. The first group, the *control group*, received no treatment for the infection. The second group received the drug. After a 30-day period, the proportions of survivors, $\hat{p}_1$ and $\hat{p}_2$, in the two groups were found to be 0.34 and 0.64 respectively.

   (a)[6] Use a 95% confidence interval to estimate the actual difference in the cure rates, i.e. $p_1 - p_2$, for the treatment versus the control groups.

   (b)[4] Based on this confidence interval can you conclude that the drug is effective? Why?

   Solution:

   (a) $n_1\widehat{p}_1 = 50 * 0.34 = 17 > 5$, $n_1\widehat{q}_1 = 50 \times (1 - 0.34) = 33 > 5$,

   $n_2\widehat{p}_2 = 32 > 5$ and $n_2\widehat{q}_2 = 18 > 5 \Rightarrow$

   95% large sample confidence interval for $\widehat{p}_1 - \widehat{p}_2$ is as follows:

   $\left(\hat{p}_1 - \hat{p}_2 \pm 1.96\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}\right) = \left(0.34 - 0.64 \pm 1.96\sqrt{\frac{(0.34)(0.66)}{50} + \frac{(0.64)(0.36)}{50}}\right) = -0.30 \pm 0.187 =$ $(-0.487, -0.113)$

   (b) Since the confidence interval does not contain zero, we can say that there is a significant difference in treatment and control group proportions.