

STAT 2507
Assignment # 4
Due: Wednesday, November 18, 2015, 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. **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 $1000 \times 80 = 80,000$ 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? _____

[4] **b.** Use *desc* command to find sample mean _____ and median _____ of c3. Do they confirm your diagnostic for the shape above? _____

[2] **c.** Generate the histogram for the data in column c1. What can you conclude about the shape of data in c1? _____

[3] **d.** Use *desc* to find sample mean _____ and sample standard deviation _____ of c1. Are they close to 9 and $9/\sqrt{1000}$? _____ Why? _____

[3] **e.** Do the values of the mean and median of c1 confirm your conclusion in (c) about the shape of data in c1? _____ Why? _____

2. **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 $\mu = 7$? _____

[3] b. How do you compare this number to the confidence level 95%? _____

Part II. Long-answer questions

1. 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. Suppose this population has a normal distribution. If a random sample of 64 stores is selected.

[5] a. Find the mean and the standard deviation of the sample mean \bar{X} . What is the sampling distribution of the sample mean?

[5] b. Find the probability that the sample mean will be between 4,980 and 5,075 customers per week?

[5] c. Find the probability that the sample mean deviates from the population mean by less than 30.

2. 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 professor 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.

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.

[5] **a.** Construct a 99% confidence interval for the average cost of a vacation trip for all single adults in the region.

[5] **b.** How large a sample would have to be taken to estimate the true average cost to within \$60 of the unknown population average cost with 99% confidence?

7. 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?