STAT 2507 C Assignment # 4 (Ch. 6,7,8) Fall 2017 Due: in class, Nov. 22.

Last Name ————————————————- , First ————————-

Student # ————————————————— Lab group————————

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

Part I. Lab questions. Use only the blanks left to answer lab questions. Include all graphs but DO NOT submit data that you are asked to generate.

1.**Continuous distributions:**

Generate and store in column c1 4,000 values from the exponential distribution with scale parameter=6 by typing the following (in the command session):

random 4000 c1;

exponential 6.

Note: The mean μ and the standard deviation σ of such a distribution are both equal to the scale parameter and this is the value you are asked to enter in the command above.

[3] a. Use desc command to find the sample mean $\bar{x}$=————,

the median———, and the sample standard deviation s =————— for these 4,000 data.

[3] b. What value(s) do $\bar{x}$ and s seem to be close to?————————–

Why so?----------

[3] c. Generate a histogram for the 4,000 values you generated from this exponential distribution. What is the shape of the distribution?---------- Is this confirmed by the values of the mean and the median you found in (a.)?

————Explain————————————————–.

2.**Normal distribution**: Generate and store in column c2 5,000 values from the normal distribution with mean 5 and standard deviation 3.5 as follows:

random 5000 c2

normal 5, 3.5.

[3] a. Generate the histogram for these data. What is the shape of this histogram?

----

[3] b. What is the value on the horizontal axis around which the histogram seems to be symmetric? X=---

[3] c. Use Minitab to find the sample mean $\bar{x}$=————— and the standard deviation s----------

for the data you generated.

[3] d. What values are $\bar{x}$ and s respectively close to?---------- and ----------. Why so?----------

3. **Central limit theorem (CLT) at work** (You can open a new Minitab worksheet, simply by typing new). Generate and store in columns c3-c1002 120 horizontal samples, each of size n = 1000, from exponential distribution with mean μ = 8 as follows:

random 120 c3-c1002;

expo 8.

Note This may (or may not!) take a few moments as you are generating 1000×120=120,000 values! Create and store in column c1 the 120 values of $\bar{x}$ based on the 120 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?----------

[3]b. Use desc command to find the sample mean---------and the median-------of c3. Do they confirm your diagnostic in part (a) for the shape above?----------Explain.----------

[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 command to find sample mean----------and sample median---------- of c1.

Do they confirm your findings in (c.) about the shape of data in c1?----------Explain?----------

[3] e. What is the sample standard deviation of c1?---------- What values do the sample mean and the sample standard deviation of c1 seem to be respectively close to?---------- and ----------. Why so?----------

4.**Confidence interval (CI) for a population mean**: We want to build 150 confidence intervals (CIs) with confidence level (1 − α)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 150 samples of size 500 each, from Poisson with parameter μ = 6.5 as follows:

random 150 c6-c505;

poisson 6.5.

Step 2. Use columns c4 and c5 to store respectively the means and the standard deviations of the 150 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% CIs 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 or not the corresponding interval [c2 ,c3] covers μ or not, by typing in the following logical function:

let c1=(c2 <6.5 and c3 >6.5)

Finally sum up the entries of column c1 to find out how many (and what percent of) CIs did cover the value μ = 6.5 by typing: tally c1;

counts;

percent.

[3] a. How many confidence intervals (and out of what) did contain the true value μ = 6.5?----------

and What is the percent of intervals that did cover μ = 6.5?----------

[3] b. What value should the percent of intervals you found in (a) be close to?----------Explain.----------

Was this the case?----------Explain----------

Part II. Long-answer questions

1. Data collected over a long period of time showed that a particular genetic defect occurs in 2 of every 1000 children. Let X be the random variable "number of children with genetic defect in a sample of 20,000 children examined". The records of a medical clinic show X = 25 children.

[5] a. What is the probability of observing a value of X more than 25?

[2] b. Would you say that the observation of X = 25 children with genetic defect was rather unlikely?

2. Suppose a random sample of n = 15 observations is selected from a population that has normal distribution with mean 50 and standard deviation 10.

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

[5] b. Find the probability that $\bar{X}$ exceeds 55.

[5] c. Find the probability that the sample mean deviates from the population mean by less than 4.

3. By statistics, faculty with rank of assistant professor (AP) finishing their 2nd year of employment at a higher education institution in Ontario earn an average of $ 75,360 per year with a standard deviation of $2,800. In an attempt to verify this salary level, a random sample of 40 AP with 2 years of experiment 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 of these 40 AP.

[8] b. Within what limit would you expect the sample mean to fall with probability .95?

[6] c. If the random sample actually produced a sample mean $\bar{X}$ = 72, 000, would you consider this rather unusual? What conclusion might you draw then?

4. [4] In a report on why e-shoppers abandon their online sale transactions, a study found that "pages took me too much time to load" and "site was too confusing to me so that I couldn't find the product" were the two main complaints heard most often. Based on 50 customers' responses, the average time to complete online order was 5.3 minutes and the standard deviation was 3.1 minutes. Construct a 95% confidence interval for μ the average completion time for an online order.

5.[4] In a poll of n = 700 randomly selected adults, 390 indicated that movies are getting better. Construct a 90% confidence interval for the overall proportion p of adults who say that movies are getting better.

6. Suppose the number of successes in n = 200 binomial trials is 25.

[4] a. Find an 80% confidence interval (CI) for p, the probability of success at each trial.

[4] b. Find a 90% CI for p. Why is this interval wider than the previous one?

7. [3] (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 .6 pH, and you wish your estimate to lie within .1 of the unknown mean μ, with probability .80. Approximately how many rainfalls must be included in your sample?