STAT 2507 Solution-Assignment # 4 Fall 2012

Part I: Lab Part

Use spaces left to answer lab questions.

1. Generate and store in column c4 10,000 values from the normal distribution with $\mu = 6.5$ and $\sigma = 3$ as follows:
   random 10000 c4;
   normal 6.5 3.
   (a) [1] use Minitab to draw a histogram for these data . What is the value on the horizontal axis around which the histogram seems to be symmetric?x =**about 6.3**
   Construct and store in column c5 the data set z_i ($i = 1, \ldots, 10,000$) obtained from the previously generated data set x_i by the standardization procedure $z_i = (x_i - \mu)/\sigma$ by typing:
   let c5=(c4-6.5)/3
   (b) [2] use Minitab to draw a histogram for the z_is . Around which value does it seem to be symmetric?**0** What are the sample mean and standard deviation $\bar{z}$ and s for this new data set?$\bar{z}$=-0.0056 and s=1.004 Why are they close to 0 and 1 respectively?**Since z_i are standardized**

2. Suppose that X has a *uniform* distribution on (0,1), that is, X is a continuous random variable taking values in the interval (0,1). [The probability density function of X is simply $f(x) = 1$, for all x in (0,1).] In this question, we are going to study the validity of the CLT (Central Limit Theorem), via simulation. The steps are: **step 1** generate 6 independent random samples of size 88 observations from the distribution of X and save them in the vectors c1, c2, c3, c4, c5, c6. Then, **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, ..., c16. This can be done in minitab by using the commands (the command 'rmean' computes the mean across rows rather than down columns):

   random 88 c1-c6;

   uniform 0, 1.

   rmean c1 c11

   rmean c1-c2 c12

rmean c1-c3 c13

rmean c1-c4 c14

rmean c1-c5 c15

rmean c1-c6 c16

Obtain the histograms of c11, c12,...,c16 (for example, typing 'hist c12' will produce the histogram of c12). Note that a vector like c14, contains 88 numbers, each of which is the average of 4 independent observations from a uniform distribution on the interval (0,1).

(a) [2] Which of the 6 histograms is very different from the normal distribution? Answer: The one corresponding to c11. (Choose one of c11, c12,...,c16.)

(b) [2] Which one looks most normal (i.e., normal distribution)? c16.

(c) [2] Suppose you have a vector, called c40, that contains 88 numbers- each of which is the average of 40 independent observations from a uniform (0,1) distribution. Which one would you expect to look more normal: the histogram of c15 or that of c40? c40. Why? Based on CLT, the distribution of sample mean approaches to normal by increasing sample size.

3. Suppose that X has a N(0,1) distribution. That is, X has a normal distribution with mean 0 and variance 1. Now repeat steps 1 and 2 of Question 2 and answer the following questions. **Note:** The only thing that changes in the above set of commands is 'uniform 0, 1.' which becomes 'normal 0, 1.'

(a) [2] Obtain the histograms of c11, c12,...,c16. Which one looks more normal? Any one is Okay. Which one do you expect to look more normal? Same. Why? They are all normal.

(b) [2] Let c40 be the vector that contains 88 numbers, each of which is the average of 40 independent observations from a N(0,1) distribution. Which one do you think has a smaller variance: the values in c12 or the values in c40? c40. Why? Since $\sigma^2/n = 1/40$ as compared to 1.

(c) [1] 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*$\surd$ (**circle one**).

4. [*Confidence interval for μ when σ is known*]

Suppose $n = 9$ people are selected at random from a large population. Assume the heights of the people in this population are normal, with mean $\mu = 68.71$ inches and $\sigma = 3$ inches. Simulate the results of this selection 20 times and in each case find a 90% confidence interval for μ. The following commands may be used:

random 9 c1-c20;

normal 68.71 3.

zinterval 0.90 3 c1-c20

(a) [1] How many of your intervals contain μ? 18 (could be different number.)

(b) [1] What is the probability that 100 (not 20) such intervals would contain μ? 0.90.

(c) [1] Do all the intervals have the same width? Yes. Why (what is the theoretical width)? $\bar{X} \pm z\frac{\sigma}{n}$, so width is $2(1.645\frac{3}{\sqrt{9}} = 3.29)$.

(d) [1] Suppose you constructed 89% intervals instead of 90%. Would they be narrower or wider? Narrower.

(e) [1] How many of your intervals contained the value 71? 3(it could be different number).

(f) [2] Suppose you took samples of size $n = 4$ instead of $n = 9$. Would you expect more or fewer intervals to contain 71? More. What about 68.71? Same. What about the width of the intervals for $n = 4$: Would they be narrower or wider than for $n = 9$? Wider $(2(1.645\frac{3}{\sqrt{4}} = 4.935.))$

Part II: Written Questions

5. Data collected over a long period of time showed that a particular genetic defect occurs in 1 of every 1000 children. Let X be the random variable "number of children with genetic defect in a sample of 50,000 children examined". The records of a medical clinic show $X = 60$ children.

   (a) [3] What is the probability of observing a value of X equal to 60 or more?

   Sol:

   X can be approximated by normal with $\mu = 50000(1/1000) = 50$ and $\sigma = \sqrt{50 \times .999} = 7.07$. $P(X \geq 60) = 1 - P(X \leq 59) = 1 - P(Z \leq \frac{59.5-50}{7.07}) = 1 - P(Z \leq 1.34) = 1 - .9099 = .0901$

   (b) [2] Would you say that the observation of $X = 60$ children with genetic defect was rather unlikely?

   Sol:

   The probability in a. is rather small. May be if we did observe 60 cases, this is an indication that the ratio is more than 1/1000.

6. Suppose a random sample of $n = 36$ observations is selected from a population that has normal distribution with mean 106 and standard deviation 12.

   (a) [1] Give the mean and the standard deviation of the sample mean $\bar{X}$.

   Sol:

   Mean is 106 and standard deviation is 12/6=2.

   (b) [2] Find the probability that $\bar{X}$ exceeds 110.

   Sol:

   $P(\bar{X} > 110) = 1 - P(\bar{X} \leq 110) = 1 - P(Z \leq \frac{110-106}{2}) = 1 - P(Z \leq 2) = 1 - 0.9772 = 0.0228$

   (c) [2] Find the probability that the sample mean deviates from the population mean by less than 4.

   Sol:

   $P(-4 \leq \bar{X} - \mu \leq 4) = P(-\frac{4}{2} \leq Z \leq \frac{4}{2}) = P(Z \leq 2) - P(Z \leq -2) = 0.9772 - 0.0228 = 0.9544$

7. The breaking strength of a river has a mean value of 10000 psi and a standard deviation of 500 psi.

(a) [2] What is the probability that the sample mean breaking strength for a random sample of 40 rivers is between 9900 and 10200?

Sol:

X: breaking strength of a river with $E(X) = 10000,\ \sigma = 500,\ n = 40$, so according to CLT

$\bar{X} \sim N(10000, \frac{500^2}{40})$

$P(9900 < \bar{X} < 10200) = P(-1.26 < Z < 2.53) = 0.9943 - 0.1038 = 0.8905$

(b) [1] If the sample size had been 15 rather than 40, could the probability requested in part a) be calculated from the given information?

Sol:

No, since sample size is less than 30 and we cannot apply CLT

8. The first assignment in a statistical computing class involves running a short program. If past experience indicates that 40% of all students will make no programming errors, compare the (approximate) probability that in class of 50 students

(a) [2] At least 25 will make no errors

Sol:

X : number of students with no programming errors is binomial random variable with $n = 50,\ p = 0.4$, since $np = 0.4(50) = 20,\ nq = 30$ so $X \sim N(20, 12)$

$P(X \geq 25) = P(Z \geq \frac{24.5-20}{\sqrt{12}}) = P(Z > 1.30) = 1 - 0.9032 = 0.0968$

(b) [2] Between 15 and 25 will make no errors.

Sol:

$P(15 < X < 25) = P(\frac{14.5-20}{\sqrt{12}} < Z < \frac{24.5-20}{\sqrt{12}} = P(-1.59 < Z < 1.30) = 0.9032 - 0.0559 =$ 0.8473

9. Assume that the helium porosity of coal samples taken from any particular seam is normally distributed with true standard deviation 0.75.

(a) [2] Compute a 95% CI for the true average porosity of a certain seam if the average porosity for 20 specimens from the seam was 4.85

Sol:

$\bar{X} \pm z_{\alpha/2}\frac{\sigma}{n} \to 4.85 \pm \frac{(1.96)(0.75)}{\sqrt{20}} = (4.52,\ 5.18)$

(b) [2] Compute a 98% CI for true average porosity of another seam based on 16 specimens with a sample average porosity of 4.56.

Sol:
$Z_{\alpha/2} = Z_{0.01} = 2.33$, so the interval is $4.85 \pm \frac{2.33(0.75)}{\sqrt{16}} = (4.41,\ 5.29)$

(c) [2] How large a sample size is necessary if the width of the 95% interval is to be 0.40?
Sol:
$n = (\frac{(2(1.96))(0.75)}{0.4})^2 = 54.02$ so $n = 55$

(d) [2] What sample size is necessary to estimate true average porosity to within 0.2 with 99% confidence?
Sol:
$n = (\frac{((2.58))(0.75)}{0.2})^2 = 93.61$ so $n = 94$

10. On the basis of extensive tests, the yield point of particular type of mild steel-reinforcing bar known to be normally distributed with $\sigma = 100$. The comparison of the bar has been slightly modified, but the modification is not believed to have affected either the normality or the value of σ.
(a) [2] Assuming this to be the case, if a sample of 25 modified bars resulted in a sample average yield point of 8439 Ib, compute a 90% CI for the true average yield point of the modified bar.
Sol:
$\bar{X} \pm z_{\alpha/2}\frac{\sigma}{n} \rightarrow 8439 \pm \frac{(1.645)(100)}{\sqrt{25}} = (8406.1,\ 8471.9)$

(b) [2] How would you modify the interval in part (a) to obtain a confidence level of 92%?
Sol:
$1 - \alpha = 0.92 \rightarrow \alpha/2 = 0.04 \rightarrow Z_{\alpha/2} = Z_{0.04} = 1.75$
$\bar{X} \pm z_{\alpha/2}\frac{\sigma}{n} \rightarrow 8439 \pm \frac{(1.75)(100)}{\sqrt{25}} = (8404,\ 8474)$