Stat 2507 **Assignment 4 Solution** **Summer 2016**

Due: Wednesday, July 27, 2016 in the class @ 6:05pm-7:30pm

You should pick up your marked assignment from the TA during your lab time
Assignment 4 Solution has 8 questions, for a total of 70 marks

The marking scheme is as follows:

Question:	1	2	3	4	5	6	7	8	Total
Marks	15	15	5	6	10	5	5	9	70
Score:									

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

Question 1) *Confidence interval(CI) for μ when σ is known* .... *15 marks*

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 $\mu = 80$ inches and $\sigma = 3$ inches. Simulate the results of this selection 20 times, and in each case find a 90% CI for μ. The following commands may be used:

```
MTB > Base 5.
MTB > random 9 C1-C20;
SUBC> normal 80, 3.
MTB > zinterval 0.90 3 C1-C20
```

(a) (1 mark) How many of your CIs contain μ? **18(Could be different depend on sample)**

(b) (2 marks) Would you expect all 20 of the CIs to contain μ? **No**. Why? $np = 20(0.9) = 18$.

(c) (2 marks) Do all the intervals have the same width? **Yes**. Why (what is the theoretical width)? **CI** $\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ **has width** $w = 2(1.645)\frac{3}{3} = 3.29$

(d) (1 mark) How many of your intervals contained the value 77? **3**

(e) (1 mark) Suppose you constructed 89% intervals instead of 90%. Would they be narrower or wider? **Narrower**.

(f) (1 mark) Suppose you constructed 95% intervals instead of 90%. Would they be narrower or wider? **Wider**.

(g) (3 marks) Suppose you took samples of size $n = 4$ instead of $n = 9$. Would you expect more or fewer intervals to contain 77? **More**. What about 80? **Same**. What about the width of the intervals for $n = 4$, would they be narrower or wider than for $n = 9$? **Wider**. Why? $w = 2(1.645)\frac{3}{2} = 4.935$

(h) (4 marks) Suppose that n is as before (i.e., 9), but $\sigma = 7$. Would you expect more or fewer intervals to contain 77? **More**. What about 80? **Same**. What about the width of the intervals when $\sigma = 7$, would they be narrower or wider than for $\sigma = 3$? **Wider**. Why? $w = 2(1.645)\frac{7}{3} = 7.841$

Question 2) *CI for μ when σ is NOT known* .... *15 marks*

Repeat the simulation of Question 1 but now assume σ is unknown and use the "tintervals" command to get the 20 of the 90% intervals:

```
MTB > Base 5.
MTB > random 9 C1-C20;
SUBC> normal 80, 3.
MTB > tinterval 0.90 C1-C20
```

(a) (2 marks) What is the assumption to use t-interval? **Sample from normal population**.

(b) (2 marks) How many of your CIs contain μ? **14(any # >14 depending on sample)**

(c) (2 marks) Would you expect all 20 of the CIs to contain μ? **No**. Why? $np = 20(0.9) = 18$.

(d) (2 marks) Do all the intervals have the same width? **No**. Why (what is the theoretical width)? **CI $\bar{x} \pm t_{\alpha/2}\frac{s}{\sqrt{n}}$ has width $w = 2(1.833)\frac{s}{3}$ which depends on the sample**.

(e) (1 mark) How many of your intervals contained the value 77? **3**

(f) (1 mark) Suppose you constructed 89% intervals instead of 90%. Would they be narrower or wider? **Narrower**.

(g) (1 mark) Suppose you constructed 95% intervals instead of 90%. Would they be narrower or wider? **Wider**.

(h) (4 marks) Suppose you took samples of size $n = 100$ instead of $n = 9$. Would you expect more or fewer intervals to contain 77? **More**. What about 80? **Same**. What about the width of the intervals for $n = 100$, would they be narrower or wider than for $n = 9$? **Narrower**. Why? $w = 2(1.833)\frac{s}{10}$

Question 3) .... *5 marks*

The proportion of individuals with an Rh-positive blood type is 85%. You have a random sample of $n = 500$ individuals. What is the probability that the sample proportion lies between 83% and 88%?

Solution:

Since $np = 500(0.85) = 425 > 5$, $nq = 500(0.15) = 75 > 5 \Rightarrow \hat{p} \sim N(0.85, \sigma = \sqrt{\frac{pq}{n}} = 0.01597)$

$$P(0.83 < \hat{p} < 0.88) = P(-1.25 < Z < 1.88) = 0.9699 - 0.1056 = 0.8643$$

Question 4) .... *6 marks*

Suppose a random sample of $n = 36$ observations is selected from a population that has normal distribution $N(\mu = 106, \sigma = 12)$.

(a) (2 marks) Give the mean and the Std of the sample mean $\bar{X}$ and its distribution.

Solution:

mean $\mu_{\bar{x}} = \mu = 106$ and Std$= \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{12}{6} = 2$ then $\bar{X} \sim N(106, 2)$

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

Solution:

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

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

Solution:

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

Question 5) .... *10 marks*

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

(a) (2 marks) 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.

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

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

Solution:
$\bar{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \Rightarrow 4.56 \pm 2.33\frac{0.75}{4} = (4.124.99)$

(c) (2 marks) How large a sample size is necessary if the width of the 95% interval is to be 0.40?

Solution:
$n = \left(\frac{2z_{\alpha/2}\sigma}{e}\right)^2 = \left(\frac{2(1.96)(0.75)}{0.40}\right)^2 = 54.02 \approx 55$

(d) (4 marks) A dean wishes to estimate the average cost of the freshman year at a particular college correct to within \$500, with a probability of 0.95. If a random sample of freshmen is to be selected and each asked to keep financial data, how many must be included in the sample? Assume that the dean knows only that the range of expenditures will vary from approximately \$4800 to \$13000.

Solution:
Using the range approximation to obtain an estimate of σ, we have

$$\sigma \approx \frac{R}{4} = \frac{13,000 - 4800}{4} = 2050$$

and the desired value of n is obtained:

$$1.96\frac{\sigma}{\sqrt{n}} \leq B \Longleftrightarrow 1.96\frac{2050}{\sqrt{n}} \leq 500 \Rightarrow n \geq 64.58 \approx 65$$

Question 6) .... *5 marks*

The serum cholesterol levels of 20 subjects randomly selected from the L.A. Heart Data, with the following scores

$$71, 73, 84, 77, 93, 86, 89, 68, 91, 82, 67, 65, 86, 76, 62, 75, 75, 57, 72, 84$$

Find 95% CI for μ

Solution:
$\bar{x} = \frac{\sum x_i}{n} = \frac{1533}{20} = 76.65$ and
$s^2 = \frac{1}{n-1}\left[\sum x_i^2 - n\bar{x}^2\right] = \frac{1}{19}(119419 - 117504.5) = 100.7658 \Rightarrow s = 10.04$
Since $n < 30$, and we use the value of t with $df = 19$, the value of t will be $t_{0.025} = 2.093$

$$\bar{x} \pm t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}} \Rightarrow 76.65 \pm 2.093\frac{10.04}{4.472} = (71.952, 81.348)$$

Question 7) .... *5 marks*

How likely is someone to vote in the next presidential election? A random sample of 300 adults was taken, and 192 of them said that they always vote in presidential elections. Construct a 95% CI for the proportion of adult Americans who say they always vote in presidential elections.

Solution:

The point estimate of p is $\hat{p} = \frac{x}{n} = \frac{192}{300} = 0.64$, and the approximate 95% CI for p is

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} \Rightarrow 0.64 \pm 1.96\sqrt{\frac{(0.64)(0.36)}{300}} = 0.64 \pm 0.054 = (0.586, 0.694)$$

Question 8) .... *9 marks*

Check the right answer

(a) (2 marks) Assume that candy boxes of a certain type are claimed to have a weight of μ kilograms. A random sample of 35 of these candy boxes produced an average weight of 458 grams and a Std of 0.082 kilograms. A 99% CI for the average weight, μ, of all such candy boxes is:

○ $(0.222, 0.494)$ ○ $(0.222, 0.654)$ √ $(0.425, 0.490)$ ○ $(0.344, 0.564)$

(b) (2 marks) If you make 200 such intervals, based on 200 independent samples with a size of 35 each, the probability that 10 of these sample will not capture the true mean μ can best be approximated by

○ A $N(0,1)$ ○ A $Bin(n,p)$ has $\mu = 198$ √ **A** $Pois(\mu = 2)$ ○ None.

(c) (2½ marks) We want to compare the difference in mean daily intake of dairy products for adults in two different towns: A and B. Random samples of 30 adults from each town produced the following mean and Std of the daily intakes of dairy products: $\bar{x}_A = 167.1$, $\bar{X}_B 140.9$, $S_1 = 24.3$, and $S_2 = 17.6$. Find a 95% CI for the difference in the mean dairy product intakes.

○ $(17.463, 41.345)$ ○ $(0.222, 0.654)$ √ $(15.463, 36.937)$ ○ $(10.344, 48.564)$

(d) (2½ marks) Suppose that in a random sample of 1000 individuals from country A, 51 have a particular gene. The number is 56 for a random sample of size 1000 from country B. Construct a 99% CI for the difference in the proportion of the people who carry this gene for the 2 countries.

√ $(-0.0309, 0.0209)$ ○ $(-0.224, 0.345)$ ○ $(-0.016, 0.034)$ ○ $(0.012, 0.157)$