Introduction to Confidence Intervals – Solutions

STAT-UB.0103 – Statistics for Business Control and Regression Models

The Central Limit Theorem (Review)

1. You draw a random sample of size $n = 64$ from a population with mean $\mu = 50$ and standard deviation $\sigma = 16$. From this, you compute the sample mean, $\bar{X}$.

   (a) What are the expectation and standard deviation of $\bar{X}$?

   > **Solution:**
   >
   > $$\begin{aligned} \mathrm{E}[\bar{X}] &= \mu = 50, \\ \mathrm{sd}[\bar{X}] &= \frac{\sigma}{\sqrt{n}} = \frac{16}{\sqrt{64}} = 2. \end{aligned}$$

   (b) Approximately what is the probability that the sample mean is above 54?

   > **Solution:** The sample mean has expectation 50 and standard deviation 2. By the central limit theorem, the sample mean is approximately normally distributed. Thus, by the empirical rule, there is roughly a 2.5% chance of being above 54 (2 standard deviations above the mean).

   (c) Do you need any additional assumptions for part (c) to be true?

   > **Solution:** No. Since the sample size is large ($n \geq 30$), the central limit theorem applies.

2. You draw a random sample of size $n = 16$ from a population with mean $\mu = 100$ and standard deviation $\sigma = 20$. From this, you compute the sample mean, $\bar{X}$.

   (a) What are the expectation and standard deviation of $\bar{X}$?

   > **Solution:**
   >
   > $$\begin{aligned} \mathrm{E}[\bar{X}] &= \mu = 100, \\ \mathrm{sd}[\bar{X}] &= \frac{\sigma}{\sqrt{n}} = \frac{20}{\sqrt{16}} = 5. \end{aligned}$$

   (b) Approximately what is the probability that the sample mean is between 95 and 105?

   > **Solution:** The sample mean has expectation 100 and standard deviation 5. If it is approximately normal, then we can use the empirical rule to say that there is a 68% of being between 95 and 105 (within one standard deviation of its expecation).

(c) Do you need any additional assumptions for part (c) to be true?

Solution: Yes, we need to assume that the population is normal. The sample size is small ($n < 30$), so the central limit theorem may not be in force.

Introduction to Confidence Intervals

3. Consider the following game. Population with mean μ and and known standard deviation $\sigma = 7$. I know μ, but you don't. You sample $n = 49$ observations from the population and compute the sample mean $\bar{X}$. Your goal is to guess the value of μ. Suppose you observe the sample mean $\bar{x} = 4.110$.

   (a) If μ were equal to 4, would $\bar{x} = 4.110$ be typical? Take "typical" to mean "we would observe a value like this about 95% of the time."

   Solution: The expectation of $\bar{X}$ is μ; the standard deviation of $\bar{X}$ is

   $$\mathrm{sd}(\bar{X}) = \frac{\sigma}{\sqrt{n}} = \frac{7}{\sqrt{49}} = 1.$$

   The value $\bar{x} = 4.110$ is within 1 standard deviation of the mean.

   (b) If μ were equal to 5, would $\bar{x} = 4.110$ be typical?

   Solution: Yes, $\bar{x}$ is within 2 standard deviation of the mean.

   (c) If μ were equal to 10, would $\bar{x} = 4.110$ be typical?

   Solution: No, X is over 5 standard deviations away from the mean.

   (d) What is the largest value of μ for which a sample of $\bar{x} = 4.110$ would be considered typical?

   Solution: 6.110

   (e) What is the smallest value of μ for which a sample of $\bar{x} = 4.110$ would be considered typical?

   Solution: 2.110

   (f) What can you say about the random interval $(\bar{X} - 2, \bar{X} + 2)$?

   Solution: There is a 95% that μ is in this interval.

   (g) What can you say about the observed interval $(\bar{x} - 2, \bar{x} + 2)$, where $x = 4.110$?

   Solution: We cannot use probability to talk about this interval, since there is no randomness involved. All we can say is that we have 95% *confidence* that μ is in this interval. (If we repeated the sampling processing many times, constructing a new confidence interval each time, then 95% of the confidence intervals would contain μ.)

Confidence Intervals for a Population Mean (Known Variance)

4. A random sample of n measurements was selected from a population with unknown mean μ and known standard deviation σ. Calculate a 95% confidence interval for μ for each of the following situations:

   (a) $n = 49$, $\bar{x} = 28$, $\sigma = 28$

   Solution: The formula for the 95% confidence interval is

   $$\bar{x} \pm 2\frac{\sigma}{\sqrt{n}}.$$

   If we substitute the values of the problem, we get $28 \pm 2\frac{28}{\sqrt{49}}$, i.e.

   $$28 \pm 8,$$

   or $(20, 36)$.

   (b) $n = 36$, $\bar{x} = 12$, $\sigma = 18$

   Solution: $12 \pm 2\frac{18}{\sqrt{36}}$, i.e.

   $$12 \pm 6.$$

   (c) $n = 100$, $\bar{x} = 125$, $\sigma = 50$

   Solution: $125 \pm 2\frac{50}{\sqrt{100}}$, i.e.

   $$125 \pm 10.$$

   (d) Is the assumption that the underlying population of measurements is normally distributed necessary to ensure the validity of the confidence intervals in parts (a)–(c)?

   Solution: No. Since the sample sizes are large ($n \geq 30$), the central limit theorem guarantees that $\bar{x}$ is approximately normal, so the confidence intervals are valid.

5. Complete the previous problem, with 99% confidence intervals instead of 95% confidence intervals.

Solution: Instead of using $\bar{x} \pm 2\frac{\sigma}{\sqrt{n}}$, the confidence interval is

$$\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}},$$

where $\alpha = .01$. In this case, $z_{\alpha/2} = 2.576$. The confidence intervals are:

(a) $28 \pm 2.576\frac{28}{\sqrt{49}}$, i.e.

$$28 \pm 10.304.$$

(b) $12 \pm 2.576\frac{18}{\sqrt{36}}$, i.e.

$$12 \pm 7.728.$$

(c) $125 \pm 2.576\frac{50}{\sqrt{100}}$,

$$125 \pm 12.88.$$

6. Find the values of α and $z_{\alpha/2}$ for computing 99.9% confidence intervals. (If you don't have a z table, draw a bell curve with a shaded region showing the relationship between α and $z_{\alpha/2}$).

Solution: For a 99.9% confidence interval, $\alpha = .001$, so $z_{\alpha/2} = z_{.0005} = 3.291$.

7. Find the values of α and $z_{\alpha/2}$ for computing 80% confidence intervals.

Solution: For an 80% confidence interval, $\alpha = .20$, so $z_{\alpha/2} = z_{.10} = 1.282$.