Confidence Interval for Population Mean – Solutions
COR1-GB.1305 – Statistics and Data Analysis

1. Of the students who filled out the online class survey, 49 reported their GMAT scores. The sample mean of the reported scores was 681, and the sample standard deviation was 42.

   (a) What is a reasonable population to associate with this sample?

   > **Solution:** The GMAT scores of all first-year Langone students.

   (b) What is the meaning of the "population mean"?

   > **Solution:** μ, the mean GMAT score of all first-year Langone students.

   (c) Find a 95% confidence interval for the population parameter.

   > **Solution:** We have $\bar{x} = 681$, $s = 42$, and $n = 49$. For a 95% confidence interval, $\alpha 0.050$ and $\alpha/2 = 0.025$; with $n = 49$, there are $n - 1 = 48$ degrees of freedom. Consulting the t table, we find
   >
   > $$t_{\alpha/2,n-1} = t_{0.025,48} \approx 2.021.$$
   >
   > (Note that $\text{df} = 48$ is not in the table, so we look at $\text{df} = 40$ instead.) The 95% confidence interval for μ is
   >
   > $$\begin{aligned} \bar{x} \pm t_{\alpha/2,n-1} \frac{s}{\sqrt{n}} &= 681 \pm 2.021 \cdot \frac{42}{\sqrt{49}} \\ &= 681 \pm 12.126 \\ &= (668.874, 693.126). \end{aligned}$$

   (d) Under what conditions is the confidence interval valid?

   > **Solution:** Since n is large ($n \geq 30$), the interval is valid if we have a simple random sample.

2. Use the following sample means and sample standard deviations from the class survey to form 95% confidence intervals for the population mean of each variable.

   (a) Dinners per month: $\bar{x} = 4.2$, $s = 7.1$, $n = 53$.

   Solution:

   $$t_{\alpha/2,n-1} = t_{0.025,52} \approx 2.009$$

   $$\begin{aligned}\bar{x} \pm t_{\alpha/2,n-1}\frac{s}{\sqrt{n}} &= 4.2 \pm 2.009 \cdot \frac{7.1}{\sqrt{53}} \\ &= 4.2 \pm 2.0 \\ &= (2.2, 6.2)\end{aligned}$$

   (b) Age (years): $\bar{x} = 27.6$, $s = 4.7$, $n = 54$.

   Solution:

   $$t_{\alpha/2,n-1} = t_{0.025,53} \approx 2.009$$

   $$\begin{aligned}\bar{x} \pm t_{\alpha/2,n-1}\frac{s}{\sqrt{n}} &= 27.6 \pm 2.009 \cdot \frac{4.7}{\sqrt{54}} \\ &= 27.6 \pm 1.3 \\ &= (26.3, 28.9)\end{aligned}$$

   (c) Commute Time (minutes): $\bar{x} = 37.5$, $s = 21.0$, $n = 53$.

   Solution:

   $$t_{\alpha/2,n-1} = t_{0.025,52} \approx 2.009$$

   $$\begin{aligned}\bar{x} \pm t_{\alpha/2,n-1}\frac{s}{\sqrt{n}} &= 37.5 \pm 2.009 \cdot \frac{21.0}{\sqrt{53}} \\ &= 37.5 \pm 5.8 \\ &= (31.7, 43.3)\end{aligned}$$

3. In Problem 2, what are the relevant populations?

> **Solution:** The number of dinners per month, age, and commute times of *all* first-year Langone students.

4. In Problem 2, what assumptions do we need for the confidence intervals to be valid? How could we check these assumptions?

> **Solution:** We need that the sample is a simple random sample. In this case, that is equivalent to the sample being unbiased: every member of the population has an equal chance of being selected. If the sample is *not* unbiased, then we need the bias to be unrelated to the quantities of interest. In particular, we need that "Dinners per Month," "Age," and "Commute Time" are unrelated to the students choice of statistics class.
>
> We do *not* need that the population is normal, because $n \geq 30$ in each of these examples.

5. In each of the following situations, find α and $t_{\alpha/2,n-1}$.

   (a) An 80% confidence interval with $n = 10$.

   > **Solution:**
   >
   > $$\alpha = .20, \quad n-1 = 9 \text{ degrees of freedom}, \quad t_{.100,9} = 1.383.$$

   (b) A 99% confidence interval with $n = 25$.

   > **Solution:**
   >
   > $$\alpha = .01, \quad n-1 = 24 \text{ degrees of freedom}, \quad t_{.005,24} = 2.797$$

   (c) A 90% confidence interval with $n = 30$.

   > **Solution:**
   >
   > $$\alpha = .10, \quad n-1 = 29 \text{ degrees of freedom}, \quad t_{.050,29} \approx 1.701.$$
   >
   > Note that df $= 29$ is not in the table, so we approximate the value of $t_{.050,29}$ by using either df $= 28$ or df $= 30$ (either answer is acceptable).

6. A random sample of 36 measurements was selected from a population with unknown mean μ. The sample mean is $\bar{x} = 12$ and the sample standard deviation is $s = 18$. Calculate an approximate 95% confidence interval for μ. Use the approximation $t_{\alpha/2,n-1} = t_{0.025,35} \approx 2$.

   > **Solution:** We compute a 95% confidence interval for μ via the formula $\bar{x} \pm t_{0.025,n-1}\frac{s}{\sqrt{n}}$. In this case, we get $12 \pm 2\frac{18}{\sqrt{36}}$ i.e., 12 ± 6.

7. With respect to the previous problem, which of the following statements are true:

   A. There is a 95% chance that μ is between 6 and 18.

   B. The population mean μ will be between 6 and 18 about 95% of the time.

   C. In 95% of all future samples, the sample mean will be between 6 and 18.

   D. The population mean μ is between 6 and 18.

   E. None of the above.

Solution: The correct answer is E. The numbers μ, 6, and 18 are all nonrandom, so it makes no sense to talk about probability. Instead, we can say that we have 95% *confidence* that μ is between 6 and 18. The term "confidence" denotes subjective belief, as opposed to "probability," which is concerned with randomness.

8. Complete Problem 6, with a 99% confidence interval instead of a 95% confidence interval.

Solution: For a $100(1-\alpha)\%$ confidence interval for μ, we use the formula $\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$. For a 99% confidence interval, we have $\alpha = .01$ and $z_{\alpha/2} = 2.576$. Thus, our confidence interval for μ is $12 \pm 2.576\frac{18}{\sqrt{36}}$ i.e., 12 ± 7.728.

9. Complete Problem 6, with an 80% confidence interval instead of a 95% confidence interval.

Solution: For a $100(1-\alpha)\%$ confidence interval for μ, we use the formula $\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$. For a 80% confidence interval, we have $\alpha = .20$ and $z_{\alpha/2} = 1.282$. Thus, our confidence interval for μ is $12 \pm 1.282\frac{18}{\sqrt{36}}$ i.e., 12 ± 3.846.

10. How reliable is the SoHo Halal Guy's Yelp rating? The SoHo Halal Guy at Broadway and Houston (`http://www.yelp.com/biz/soho-halal-guy-new-york`) currently has 36 Yelp reviews (1 1-star; 1 2-star; 5 3-star; 12 4-star; and 17 5-star). The average star rating is 4.19 and the sample standard deviation of the star ratings is 0.98. How much should we trust the number "4.19"? We will use a confidence interval to quantify the uncertainty associated with this number.

   (a) What is a reasonable population to associate with this sample?

   > **Solution:** All ratings of the Halal Cart (past and future).

   (b) What is the meaning of the population mean, μ?

   > **Solution:** The parameter of interest is μ, the mean start rating of all people who ever review the Halal Cart. Equivalently, the μ is equal to expected star rating of a random Halal Cart reviewer.

   (c) Find a 95% confidence interval for the population mean, μ.

   > **Solution:**
   > For a 95% confidence interval, we have $\alpha = 0.05$ and $\alpha/2 = 0.025$. The sample size is $n = 36$. There are $n - 1 = 35$ degrees of freedom. Thus, using the t table, we have
   >
   > $$t_{\alpha/2,n-1} = t_{0.025,35} \approx 2.032.$$
   >
   > The 95% confidence interval for the population mean, μ, is
   >
   > $$\begin{aligned} \bar{x} \pm 2.032\frac{s}{\sqrt{n}} &= 4.19 \pm 2.032\frac{0.98}{\sqrt{36}} \\ &= 4.19 \pm 0.33 \\ &= (3.86, 4.52). \end{aligned}$$

   (d) Under what conditions is the confidence interval valid?

   > **Solution:** For a confidence interval for a mean to be valid, we need that (i) the observed sample is a simple random sample from the population, and (ii) $n \geq 30$ or the population is normal. Clearly, assumption (ii) holds. Here, it is reasonable to assume (i) as long as the Halal Cart and its customer base do not change in the future.

11. La Colombe at Lafayette and 4th St (`http://www.yelp.com/biz/la-colombe-new-york-2/`) currently has 467 Yelp reviews (10 1-star; 15 2-star; 32 3-star; 148 4-star; and 262 5-star). The average star rating is 4.36 and the sample standard deviation of the star ratings is 0.91. Find a 95% confidence interval for the expected rating of a random La Colombe Yelp reviewer.

Solution: Since $n \geq 30$, we can approximate $t_{0.025,n-1} \approx 2$. (A more accurate approximation would be $t_{0.025,n-1} \approx 1.960$. An approximate 95% confidence interval for the population mean is

$$\begin{aligned}\bar{x} \pm 2\frac{s}{\sqrt{n}} &= 4.36 \pm 2\frac{0.91}{\sqrt{467}} \\ &= 4.36 \pm 0.08 \\ &= (4.26, 4.44)\end{aligned}$$