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

Confidence Interval for Mean

1. 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.

2. Complete Problem 1, 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 $\bar{x} \pm t_{\alpha/2,n-1} \frac{\sigma}{\sqrt{n}}$. For a 99% confidence interval, we have $\alpha = .01$ and $t_{\alpha/2,n-1} \approx 2.728$. Thus, our confidence interval for μ is $12 \pm 2.728 \frac{18}{\sqrt{36}}$ i.e., 12 ± 8.214.

3. Complete Problem 1, 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 $\bar{x} \pm t_{\alpha/2,n-1} \frac{\sigma}{\sqrt{n}}$. For a 80% confidence interval, we have $\alpha = .20$ and $t_{\alpha/2,n-1} \approx 1.307$. Thus, our confidence interval for μ is $12 \pm 1.307 \frac{18}{\sqrt{36}}$ i.e., 12 ± 3.921.

4. 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 47 Yelp reviews (4 1-star; 1 2-star; 6 3-star; 16 4-star; and 20 5-star). The average star rating is 4.0 and the sample standard deviation of the star ratings is 1.2. How much should we trust the number "4.0"? 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 $n = 47$. There are $n - 1 = 46$ degrees of freedom. Thus, using the t table, we have

   $$t_{\alpha/2,n-1} = t_{0.025,46} \approx 2.021.$$

   The 95% confidence interval for the population mean, μ, is

   $$\begin{aligned} \bar{x} \pm 2.021 \frac{s}{\sqrt{n}} &= 4.0 \pm 2.021 \frac{1.2}{\sqrt{47}} \\ &= 4.0 \pm 0.35 \\ &= (3.65, 4.35). \end{aligned}$$

   (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.

5. La Colombe at Lafayette and 4th St (`http://www.yelp.com/biz/la-colombe-new-york-2/`) currently has 559 Yelp reviews (11 1-star; 19 2-star; 41 3-star; 172 4-star; and 316 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

$$\begin{aligned}\bar{x} \pm 2\frac{s}{\sqrt{n}} &= 4.36 \pm 2\frac{0.91}{\sqrt{559}} \\ &= 4.36 \pm 0.08 \\ &= (4.26, 4.44)\end{aligned}$$

Confidence Interval for Proportion

6. A CNN/ORC post-debate poll surveyed 547 voters who watched the third presidential debate on October 19, 2016. The results are at `http://www.cnn.com/2016/10/19/politics/hillary-clinton-wins-third-presidential-debate-according-to-cnn-orc-poll`. Of the respondents, 52% thought that Clinton did the best job, while 39% thought that Trump did.

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

   > **Solution:** The opinions of all voters who watched the debate.

   (b) There are a few population parameters of interest. Choose one.

   > **Solution:** p, the proportion of all debate-watching voters who thought that Clinton won. (Alternatively, the proportion of all debate-watching voters who voters who thought the candidates tied, or that Trump won.)

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

   > **Solution:** The sample proportion is $\hat{p} = 0.52$. The sample size is $n = 547$. For a 95% confidence interval, we have $\alpha = 0.05$ and $\alpha/2 = 0.025$. Thus,
   >
   > $$z_{\alpha/2} = z_{0.050} = 1.960$$
   >
   > (use the df $= \infty$ section of the t-table.) Thus, the 95% confidence interval for p is
   >
   > $$\begin{aligned} \hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} &= 0.52 \pm 1.96\sqrt{\frac{(0.52)(1-0.52)}{547}} \\ &= 0.52 \pm 0.04 \\ &= (0.48, 0.56) \end{aligned}$$
   >
   > For the proportion of debate-watching voters who thought that Trump won, the confidence interval is
   >
   > $$\begin{aligned} \hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} &= 0.39 \pm 1.96\sqrt{\frac{(0.39)(1-0.39)}{547}} \\ &= 0.39 \pm 0.04 \\ &= (0.35, 0.43) \end{aligned}$$

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

   > **Solution:** We need a simple random sample, and we need to have expected at least 15 successes and 15 failures ($np \geq 15$ and $n(1-p) \geq 15$). The latter condition is almost certainly satisfied since we had $n\hat{p} = 284$ successes and $n(1-\hat{p}) = 263$ failures in the sample. For the former condition, we need the sample to be unbiased.

7. Use the following data from the class survey to estimate the relevant population proportions. Give 95% confidence intervals for these proportions.

   (a) Gender: 17 Female, 30 Male.

   Solution: If we let p be the proportion of Female in the population, then we have $\hat{p} = \frac{17}{17+30} = 0.36$ and $n = 17 + 30 = 47$. The 95% confidence interval for p is

$$\begin{aligned} \hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} &= 0.36 \pm 1.96\sqrt{\frac{(0.36)(1-0.36)}{47}} \\ &= 0.36 \pm 0.14 \\ &= (0.22, 0.50) \end{aligned}$$

   (b) Drinks at least one cup of coffee on a typical day: 37 Yes, 10 No.

   Solution: If we let p be the proportion of coffee drinkers in the population, then we have $\hat{p} = \frac{37}{37+10} = 0.79$ and $n = 37 + 10 = 47$. The 95% confidence interval for p is

$$\begin{aligned} \hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} &= 0.79 \pm 1.96\sqrt{\frac{(0.79)(1-0.79)}{47}} \\ &= 0.79 \pm 0.12 \\ &= (0.67, 0.91) \end{aligned}$$

   (c) Political affiliation: 36 Democrat, 6 Republican, 5 Other. (For this problem there are three different choices for the population parameter; choose one of them.)

   Solution: If we let p be the proportion of the population that are Democrats. Then, we have $\hat{p} = \frac{36}{36+6+5} = 0.77$ and $n = 36 + 6 + 5 = 47$. The 95% confidence interval for p is

$$\begin{aligned} \hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} &= 0.77 \pm 1.96\sqrt{\frac{(0.77)(1-0.77)}{47}} \\ &= 0.77 \pm 0.12 \\ &= (0.65, 0.99) \end{aligned}$$

   If instead we let p be the proportion of the population that are Republicans, then we have $\hat{p} = \frac{6}{36+6+5} = 0.13$ and $n = 36 + 6 + 5 = 47$. The 95% confidence interval for p is

$$\begin{aligned} \hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} &= 0.13 \pm 1.96\sqrt{\frac{(0.13)(1-0.13)}{47}} \\ &= 0.13 \pm 0.10 \\ &= (0.03, 0.23) \end{aligned}$$

8. In Problem 7, what are the relevant populations?

Solution: All first-year Stern MBA students: their genders, whether they drink coffee, and their political affiliations.

9. In Problem 7, what assumptions do we need for the confidence intervals to be valid?

Solution: We need a simple random sample, and we need to have expected at least 15 successes and 15 failures in each sample. Since we do not have at least 15 successes and 15 failures in examples (b) and (c), these intervals are only approximate, and the true confidence level may be below 95%.