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

1. The SURVEY USA polling organization interviewed 1200 adults from the United States to mea-sure public opinion on the Ebola virus. The results are at `https://infogr.am/poll-ebola-concern/`. Of the respondents, 19% said that they are "very" concerned they will catch Ebola.

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

   > **Solution:** The Ebola opinions of all U.S. adults.

   (b) What is the population parameter of interest?

   > **Solution:** p, the proportion of all U.S. adults that are "very" concerned they will catch Ebola.

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

   > **Solution:** The sample proportion is $\hat{p} = 0.19$. The sample size is $n = 1200$. 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
   >
   > $$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.19 \pm 1.96\sqrt{\frac{(0.19)(1-0.19)}{1200}}$$
   > $$= 0.19 \pm 0.02$$
   > $$= (0.17, 0.21)$$

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

   > **Solution:** We need a simple random sample, and we need to have expected at least 10 successes and 10 failures ($np \geq 10$ and $n(1-p) \geq 0$). The latter condition is almost certainly satisfied since we had $n\hat{p} = 228$ successes and $n(1 - \hat{p}) = 972$ failures in the sample. For the former condition, we need the sample to be unbiased.

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

   (a) Gender: 13 Female, 23 Male.

   > **Solution:** If we let p be the proportion of Female in the population, then we have $\hat{p} = \frac{13}{13+23} = 0.36$ and $n = 13 + 23 = 36$. The 95% confidence interval for p is
   >
   > $$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.36 \pm 1.96\sqrt{\frac{(0.36)(1-0.36)}{36}}$$
   > $$= 0.36 \pm 0.16$$
   > $$= (0.20, 0.52)$$

   (b) Drinks at least one cup of coffee on a typical day: 26 Yes, 11 No.

   > **Solution:** If we let p be the proportion of coffee drinkers in the population, then we have $\hat{p} = \frac{26}{26+11} = 0.70$ and $n = 26 + 11 = 37$. The 95% confidence interval for p is
   >
   > $$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.70 \pm 1.96\sqrt{\frac{(0.70)(1-0.70)}{37}}$$
   > $$= 0.70 \pm 0.15$$
   > $$= (0.55, 0.85)$$

   (c) Industry: 13 Finance, 24 Other.

   > **Solution:** If we let p be the proportion of the population working in Finance, then we have $\hat{p} = \frac{13}{13+24} = 0.35$ and $n = 13 + 24 = 37$. The 95% confidence interval for p is
   >
   > $$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.35 \pm 1.96\sqrt{\frac{(0.35)(1-0.35)}{37}}$$
   > $$= 0.35 \pm 0.15$$
   > $$= (0.20, 0.50)$$

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

   > **Solution:** All first-year Langone students: their genders, whether they drink coffee, and whether they work in Finance.

4. In Problem 2, 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 10 successes and 10 failures in each sample.