CONFIDENCE INTERVALS II

Our first $100(1-\alpha)\%$ CIs for the population mean μ come from the equation

$$\boxed{P\left(-z_{\alpha/2} < Z < z_{\alpha/2}\right) = 1-\alpha}$$

Because $Z = \dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ is approximately standard normal for large samples (or exactly standard normal for samples of any size from a normally distributed population), we can substitute this in and isolate μ to get the $100(1-\alpha)\%$ confidence interval: $\boxed{\overline{x} \pm z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}}$.

Video 1. Watch the first video for a more detailed recap.

Other substitutions are possible and each gives a different confidence interval. The following are all standard normal normal (or approximately standard normal for large n).

a) $Z = \dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ (known variance σ^2; approximate unless the population is normally distributed)

b) $Z = \dfrac{\overline{X}-\mu}{S/\sqrt{n}}$ (always approximate, use only for large samples)

c) $Z = \dfrac{\overline{X}_1-\overline{X}_2-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$ (independent samples from populations with known variances σ_1^2 and σ_2^2; approximate unless both populations are normally distributed)

d) $Z = \dfrac{\overline{X}_1-\overline{X}_2-(\mu_1-\mu_2)}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}$ (independent samples; always approximate, use only for large samples)

Isolating μ (or $\mu_1-\mu_2$) in the middle of the inequality gives a confidence interval for μ (or $\mu_1-\mu_2$).

1. Find the formula for a $100(1-\alpha)\%$ confidence interval for μ (or $\mu_1-\mu_2$) based on each possible substitution above:

a) $100(1-\alpha)\%$ CI for μ (known σ): $\overline{x} \pm z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$

b) $100(1-\alpha)\%$ CI for μ (large sample):

c) $100(1-\alpha)\%$ CI for $\mu_1-\mu_2$ (known σ_1 and σ_2, normal populations or large samples):

d) $100(1-\alpha)\%$ CI for $\mu_1-\mu_2$ (large samples):

Date: March 23, 2020.

2. This problem deals with the US Census bureau's 2017 American Community Survey (ACS). The survey reports mean income along with a standard error; the standard error in this case is the estimate $\frac{s}{\sqrt{n}}$.

a) The survey included 19,427 households in the Pacific West; these households had a mean income of \$101,716 with a standard error of \$1,584. Calculate a 99% confidence interval for the true mean income of a household in the Pacific West.

b) The survey also included 9,669 Mountain West households; these households had a mean income of \$88,739 with a standard error of \$1,746. Calculate a 99% confidence interval for the difference between the mean household incomes of these regions.

Video 2. Watch the video on confidence intervals for proportions.

A special case arises when dealing with estimating the proportion of a population with a certain property or characteristic. In this case the population has a Bernoulli distribution with parameter θ, the true proportion with the property. Now we can use $Z = \dfrac{\overline{X} - \theta}{\sqrt{\frac{\overline{X}(1-\overline{X})}{n}}}$, which is approximately standard normal (as long as both $n\overline{x}$ and $n(1-\overline{x})$ are both at least 10). This gives us the approximate $100(1-\alpha)\%$ CI:

$$\boxed{\overline{x} \pm z_{\alpha/2}\sqrt{\frac{\overline{x}(1-\overline{x})}{n}}}$$

3. As 2014 Pew Center study on religion in America surveyed a total of 35,071 Americans. Of those, 714 lived in Washington state. Of those living in Washington, 121 said they were Catholic.

a) Calculate a 98% confidence interval for the proportion of Americans who live in Washington.

b) Calculate a 98% confidence interval for the proportion of Washingtonians who are Catholic.

Video 3. Watch the video on confidence intervals based on Student's t distribution.

Let T be a random variable having Student's t distribution with ν degrees of freedom. As above, substitution into the expression

$$P\left(-t_{\alpha/2,\nu} < T < t_{\alpha/2,\nu}\right) = 1 - \alpha$$

leads to the confidence intervals for μ (or $\mu_1 - \mu_2$) **provided you have samples from normally distributed populations**:

a) $100(1-\alpha)\%$ CI for μ (normal population, all sample sizes): $\overline{x} \pm t_{\alpha/2,n-1}\dfrac{s}{\sqrt{n}}$

b) $100(1-\alpha)\%$ CI for $\mu_1 - \mu_2$ (normal populations with a common variance):

$$\overline{x} - \overline{y} \pm t_{\alpha/2,n_1+n_2-2}\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

where $s_p^2 = \dfrac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$ is the pooled estimator for the common variance.

c) $100(1-\alpha)\%$ CI for $\mu_1 - \mu_2$ (normal populations with difference variances): $\overline{x} - \overline{y} \pm t_{\alpha/2,\nu}\sqrt{\dfrac{s_1^2}{m} + \dfrac{s_1^2}{n}}$

where $\nu \approx \dfrac{\left(\frac{s_1^2}{m} + \frac{s_1^2}{n}\right)^2}{\frac{\left(\frac{s_1^2}{m}\right)^2}{m-1} + \frac{\left(\frac{s_1^2}{n}\right)^2}{n-1}}$ (round down to the nearest integer)

4. In a random sample of 16 games in 2016, the Gonzaga men's basketball team had an average score of $\overline{x} = 81.8750$ with a sample standard deviation of $s = 10.7881$. Calculate a 95% confidence interval for the mean score (assuming that scores are normally distributed).

5. In a random sample of 9 games in 2019, the men's basketball team had a mean score of 88.4444 with a sample standard deviation of 8.7050. Calculate a 95% confidence interval for the difference between the mean scores in 2016 and 2019. (Assume that scores for both years are normally distributed with the same variance).

6. Suppose that we want to predict Gonzaga's score in the next game (instead of producing confidence intervals for mean scores). This means that we should use a $100(1-\alpha)\%$ **prediction interval**:

$$\overline{x} \pm t_{\alpha/2,n-1}\sqrt{\frac{s^2(n+1)}{n}}$$

Use this formula to calculate a 95% prediction interval for the next score in 2019.