CONFIDENCE INTERVALS II

A $100(1-\alpha)\%$ confidence interval for the population mean μ, or the difference between two population means $\mu_1 - \mu_2$, comes from the equation

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

The random variable Z can then be replaced with one of the following:

1. $Z = \dfrac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$ (normally distributed population with a known variance σ^2)
2. $Z \approx \dfrac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}}$ (large sample from any population)
3. $Z = \dfrac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ (independent samples from normally distributed populations with known variances σ_1^2 and σ_2^2)
4. $Z \approx \dfrac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$ (large independent samples from any populations)

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. $100(1-\alpha)\%$ CI for μ (known σ): $\overline{x} \pm z_{\alpha/2} \dfrac{\sigma}{\sqrt{n}}$
2. $100(1-\alpha)\%$ CI for μ (large sample): $\overline{x} \pm z_{\alpha/2} \dfrac{s}{\sqrt{n}}$
3. $100(1-\alpha)\%$ CI for $\mu_1 - \mu_2$ (known σ_1 and σ_2, normal populations): $\overline{x} - \overline{y} \pm z_{\alpha/2} \sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$
4. $100(1-\alpha)\%$ CI for $\mu_1 - \mu_2$ (large samples): $\overline{x} - \overline{y} \pm z_{\alpha/2} \sqrt{\dfrac{s_X^2}{n_1} + \dfrac{s_Y^2}{n_2}}$

1. 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 and a 99% confidence upper bound for the difference of mean household incomes between these regions.

Date: October 16, 2019.

Similarly, the equation

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

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

1. $100(1-\alpha)\%$ CI for μ (normal population): $\overline{x} \pm t_{\alpha/2,n-1}\dfrac{s}{\sqrt{n}}$
2. $100(1-\alpha)\%$ CI for $\mu_1 - \mu_2$ (normal populations with the same variance):
   $\overline{x} - \overline{y} \pm t_{\alpha/2,n_1+n_2-2}\sqrt{s_p^2\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}$
   $s_p^2 = \dfrac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$ is the pooled estimator of the common variance for the populations
3. $100(1-\alpha)\%$ CI for $\mu_1 - \mu_2$ (normal populations with difference variances): $\overline{x} - \overline{y} \pm t_{\alpha/2,\nu}\sqrt{\dfrac{s_X^2}{m} + \dfrac{s_Y^2}{n}}$
   $\nu \approx \dfrac{\left(\frac{s_X^2}{m} + \frac{s_Y^2}{n}\right)^2}{\frac{\left(\frac{s_X^2}{m}\right)^2}{m-1} + \frac{\left(\frac{s_Y^2}{n}\right)^2}{n-1}}$ (round down to the nearest integer)

2. 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).

3. 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).

4. 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**:

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

a) Calculate a 95% prediction interval for the next score in 2016.
b) Calculate a 95% prediction interval for the next score in 2019.

5. Suppose that this class represents a random sample from a normally distributed population and use our height data to calculate a 95% confidence interval for the mean height of a Gonzaga student. The **R** commands `mean()` and `sd` might be helpful.

Proposition 1. *If X is a binomial random variable with parameters n and θ and both $n\theta \geq 8$ and $n(1-\theta) \geq 8$, then X/n is approximately normally distributed with mean θ and variance $\frac{\theta(1-\theta)}{n}$.*

6. This proposition allows us to find approximate confidence intervals for the proportion θ of a population with a given property. For example, a recent Ipsos poll (Oct 14-15) of $1,115$ Americans found that 602 disapprove of President Trump. This should allow us to calculate a confidence interval for the true percentage of Americans who disapprove of President Trump. We must first work abstractly.

a) Let X be the total number of people in a random sample of size n that disapprove of President Trump. Let θ be the true proportion of people who disapprove of President Trump. This means that X is a random variable with what distribution?
b) By the proposition, X/n is approximately normally distributed with what mean and variance?
c) This means that X/n is an unbiased estimator of what population parameter?
d) Sub this estimator into your formula for the variance of X/n to get an estimated variance S_x.
e) You can now use $Z \approx \dfrac{(X/n) - \theta}{S_x}$ to find a $100(1-\alpha)\%$ confidence interval for θ.
f) Use the Ipsos poll to calculate a 98% confidence interval for the true proportion of Americans who disapprove of President Trump.
g) A YouGov poll (Oct 13-14) of 980 Americans found 480 who disapprove of President Trump. Use this data to calculate a (different) 98% confidence interval.
h) Combine the two polls and generate a third 98% confidence interval.