How much confidence do we have in an interval?

- We now have a two-standard-error bound on the parameter $\mu_1 - \mu_2$
- For what percentage of samples will the two-standard-error bound actually enclose the parameter?
- We're asking: what is π in the expression $P(\varepsilon < b) = \pi$ (where $b = 2SE$)?
- Since the sampling distribution of the estimator $\bar{x}_1 - \bar{x}_2$ is approximately normal, this is about 95%
- This percentage is called the *confidence coefficient*
- The confidence coefficient is generally given by $1 - \alpha$, where α is known as the *level of significance*

Finding confidence intervals

How can we find a confidence interval for a parameter θ given a certain confidence coefficient/significance level?

Example

We want to compare the proportion of families earning above \$40,000 in two states. For state 1, we had a random sample of $n_1 = 100$ families with a sample proportion of $\hat{p}_1 = 0.30$ making above \$40,000. For state 2, we had a random sample of $n_2 = 144$ families with a sample proportion of $\hat{p}_2 = 0.32$.

Find the 99% confidence interval for $p_1 - p_2$.

One-sided confidence intervals

- One-sided confidence intervals do exist, but aren't used very often
- Upper bound only: $P(\frac{\theta-\hat{\theta}}{SE_{\hat{\theta}}} \leq z_\alpha) = 1 - \alpha$ yields an interval of the form $(-\infty,$ $\hat{\theta} + z_\alpha SE_{\hat{\theta}}]$
- Lower bound only: $P(\frac{\theta-\hat{\theta}}{SE_{\hat{\theta}}} \geq -z_\alpha) = 1 - \alpha$ yields an interval of the form $[\hat{\theta} - z_\alpha SE_{\hat{\theta}}, \infty)$
- Can be useful when you're interested in the percentage of samples for which the parameter is less/greater than the one boundary point

Example

We want to compare the proportion of families earning above \$40,000 in two states. For state 1, we had a random sample of $n_1 = 100$ families with a sample proportion of $\hat{p}_1 = 0.30$ making above \$40,000. For state 2, we had a random sample of $n_2 = 144$ families with a sample proportion of $\hat{p}_2 = 0.32$.

Find a lower bound for $p_1 - p_2$ in which you have 99% confidence.

Summary of confidence interval strategy

Two-sided confidence interval for parameter θ is given by $\hat{\theta} \pm z_{\frac{\alpha}{2}} SE_{\hat{\theta}}$

One-sided confidence intervals for parameter θ are given by $(-\infty, \hat{\theta} + z_\alpha SE_{\hat{\theta}}]$ or $[\hat{\theta} - z_\alpha SE_{\hat{\theta}}, \infty)$

where

- $\hat{\theta}$ is the point estimator
- $SE_{\hat{\theta}}$ is the standard error for the estimator
- Confidence coefficient: $1 - \alpha$

What if the sample is small?

- When sample sizes are large, we can assume normality of the sampling distribution for sample means/proportions (via CLT)
- Cases sometimes arise in which we can only obtain a small random sample
- If we know that the population is normal or near normal, we can form confidence intervals based on the t distribution

Small sample confidence interval for μ

- Assume we have a small random sample ($n < 30$) from a population for which σ is unknown and the distribution is assumed to be mound-shaped and near normal
- In this case, we say that is $\frac{\bar{X}-\mu}{\frac{s}{\sqrt{n}}}$ has approximately a t distribution with $n-1$ degrees of freedom
- Want a way to construct bounds such that, if we use the procedure with many different random samples, $1-\alpha$ proportion of samples will yield intervals that contain the mean
- These bounds are given by $(\bar{x} - t^{n-1}_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}, \bar{x} + t^{n-1}_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}})$

Example: Small sample conf. interval

A bottling factory fills thousands of 20oz bottles daily with soda, but not all the bottles are filled to the same level. A random sample of bottles was taken from the factory line, containing the following amounts of soda (in oz):

19.8 20.1 19.7 19.2 19.9 20.0 19.8 19.9 19.7

Assuming that the distribution of amounts of soda is approximately normal, find a 95% confidence interval for the mean amount of soda contained in the bottles. What does this say about the dependability of the factory's process?

Small sample confidence interval for $\mu_1 - \mu_2$

- Assume we have small random samples from 2 populations for which the distributions are assumed to be mound-shaped and near normal
- As long as the variances of the populations can be assumed to be equal, the same general confidence interval strategy is effective
- We can use the same general format for the confidence interval, which is $(\bar{x}_1 - \bar{x}_2) \pm t_{\frac{\alpha}{2}}^{n_1+n_2-1} SE_{\bar{x}_1 - \bar{x}_2}$, but we need the estimator for the standard error of the difference in means

Standard error estimator for $\mu_1 - \mu_2$

- In the past, our estimator for standard error for the difference in means was $\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}})$
- Since we're assuming equal variances, this simplifies to $\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$
- In most cases, we won't know σ so we need an estimator that takes the information from both samples into account
- This estimator is $s_p = \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}$
- So when we have small samples from near-normal populations assumed to have equal variances, the standard error estimate is $s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$

Example: Matched design

A task in a factory can be performed in 2 ways; management is interested in knowing which method takes less time. A random sample of employees is timed performing the task using method A and method B.

Method A	6.0	5.0	7.0	6.2	6.0	6.4
Method B	5.4	5.2	6.5	5.9	6.0	5.8
Difference	0.6	-0.2	0.5	0.3	0.0	0.6

Find a 90% confidence interval for the difference in the completion time. Assume that the time differences are distributed approximately normally.