For one of the simulated sets of data (from last time): $n = 1000,\ \hat{p} = 0.502$ ➜

$$SE_{\hat{p}} = \sqrt{\frac{0.502 \times .498}{1000}} = 0.0158$$

Fixed sample size, increasing CI

$100(1-\alpha)$	α	$\alpha/2$	$z_{\alpha/2}$	Interval
80%	.20	.10	1.285	.482,.522
90%	.10	.05	1.645	.476,.528
95%	**.05**	**.025**	**1.960**	**.471,.533**
99%	.01	.005	2.575	.461,.543

Fixed CL at 95%, increasing sample size
(assuming $\hat{p} = 0.502$ every time, which is unlikely)

Sample size	$SE_{\hat{p}}$	Interval
50	0.0707	.363, .641
75	0.0577	.389, .615
150	0.0408	.422,.582
250	0.0316	.440, .564

Estimating a population mean, σ known
Requires

- SRS

- The value of the population SD is known

- The population is normally distributed or $n > 30$; method robust wrt this requirement meaning that as long as the population is unimodal and approximately symmetric with no extreme outliers, then the methods should work just fine

  Remember that n > 30 is just a guideline. If the underlying population is heavily skewed then a larger sample may be required. For many populations samples between 15 and 30 suffice

Define

$\mu =$ population mean

$\bar{x} =$ sample mean usually provides
   the best point estimate of the population mean

- $\bar{x}$ is an unbiased estimator of μ
- For many populations $\bar{x}$ tends to be more consistent then distributions of other sample statistics

Margin of Error

Again, denoted E—with probability $1-\alpha$, the maximum likely difference between the observed sample mean $\bar{x}$ and the true value of the population mean μ

$$E = z_{\alpha/2}\left(\text{SD}_{\bar{x}}\right) = z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

Use this to construct a CI:

$$\bar{x} \pm E$$

$$\bar{x} - E < \mu < \bar{x} + E$$

$$\left(\bar{x} - E, \bar{x} + E\right)$$

Interpreting a CI—basically the same as for proportions

Must be careful. We are 95% confident that the population mean falls in …

Not: there is a 95% probability that the population mean falls between…
Not: there is a 95% chance that the population mean falls between…

After you have used sample data to construct the interval, the interval either contains the truth or it does not. Confidence comes from the method/procedure.

Determining sample size

We can use the expression for margin of error to help determine how large of a sample we need to obtain an estimate of μ with a particular level of confidence.

$$E = z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) \Rightarrow E^2 = \left(z_{\alpha/2}\right)^2\left(\frac{\sigma}{\sqrt{n}}\right)^2$$

$$\Rightarrow n = \left(\frac{z_{\alpha/2}\sigma}{E}\right)^2$$

If we are in the process of designing a study, where do we get an estimate of $\overline{x}$

- From a pilot study
- From the literature
- From an expert
- Range rule of thumb: $\sigma \approx \dfrac{\text{Range}}{4}$

Practice 1

The Tyco Video Game Corporation finds that it is losing income because of slugs used in its video games. The machines must be adjusted to accept coins only if they fall within set limits. In order to set those limits, the mean weight of quarters in circulation must be estimated. A sample of quarters will be weighed in order to determine the mean. How many quarters must we randomly select and weigh if we want to be 99% confident that the sample mean is within 0.025 g of the true population mean for all quarters? Based on results from a sample of quarters, we can estimate the population SD as 0.068 g.

Section 7-4

Estimating a population mean, σ unknown

Requires that sample is a SRS and that sample is normally distributed or $n > 30$

We will use s, the sample SD in place of σ, the population SD

$$SD_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Define $SE_{\bar{x}} = \dfrac{s}{\sqrt{n}}$

SE has more uncertainty so we compensate by making the CI a bit wider, especially for small samples

We will use the Student t distribution instead of the standard normal distribution—unimodal, symmetric around zero, but has "fatter" tails

t distribution varies for each value of n

Must use degrees of freedom (df) to in addition to CL to find the critical value

©2006 Thomson Higher Education
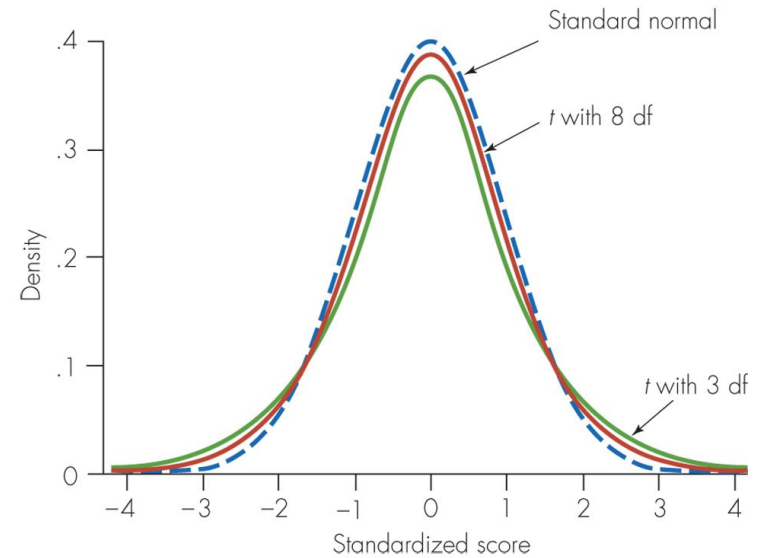
For a confidence interval for a single mean based on a collection of sample data, the df = n - 1

df have to do with the number of sample values that can vary after certain restrictions have been imposed on all data values

When estimating the mean from a sample we have the following situation:

Say we have 5 sample values and we know that the mean is 250.

This ➜ that the sum of the 5 sample values must be 1250 (5 x 250).

We can freely assign 4 of the sample values after which the 5th value will be determined as 1250 – sum of first 4 values.

Since 4 scores can be freely selected, we say there are 4 df = n - 1

The critical value is: $t_{\alpha/2,\text{df}}$ and

$$E = t_{\alpha/2,\text{df}}\text{SE}_{\bar{x}} = t_{\alpha/2,\text{df}}\left(\frac{s}{\sqrt{n}}\right)$$

Example 2

Because cardiac deaths appear to increase after heavy snowfalls, an experiment was designed to compare cardiac demands of snow shoveling to those of using an electric snow thrower. Ten subjects cleared tracts of snow using both methods, and their maximum heart rates (beats per minute) were recorded during both activities. The following results were obtained:

Manual: $n = 10$; $\bar{x} = 175$; $s = 15$

Electric: $n = 10$; $\bar{x} = 124$; $s = 18$

a) Find the 95% CI estimate of the population mean for those people who shovel snow manually.

b) Find the 95% CI estimate of the population mean for those people who use the electric snow thrower.

c) If you are a physician with concerns about cardiac deaths fostered by manual snow shoveling, what single value in the confidence interval from part (a) would be of greatest concern?

d) Compare the confidence intervals from parts (a) and (b) and interpret your findings.

Using a CI to find the point estimate and the margin of error

$$\bar{x} = \frac{(\text{Upper confidence limit}) + (\text{Lower confidence limit})}{2}$$

$$E = \frac{(\text{Upper confidence limit}) - (\text{Lower confidence limit})}{2}$$