Stat 401, section 7.1 Basic Properties of Confidence Intervals

notes by Tim Pilachowski

In the real world, researchers often don't have enough data and resources to determine actual population means and proportions (**population parameters**).

Likewise, it isn't possible to construct an entire sampling distribution to determine values of parameters. Instead, a sample will be taken and the **sample statistics** calculated.

The questions are always, "How well does this sample reflect the population? How reliable are the sample statistics as indicators of the population parameters?"

Chapter 7 uses the work of chapters 5 and 6 to begin answering these questions.

Vocabulary:
A statistic used for estimating a parameter is called a **point estimator** or **estimator**.
The standard deviation of an estimator is called its **standard error**.

The point estimate for a population mean μ will be a sample mean $\bar{x}$.
The point estimate for a population proportion p will be a sample proportion $\hat{p}$.

In this section we'll be looking at a sample as one of all of the possible samples in a **sampling distribution** and be asking, "How likely is it that our sample mean is within a given percent of the actual, but unknown, population parameter?"

Fill in the blanks.

population parameter =

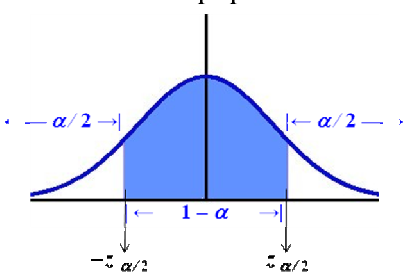
point estimate = sample mean =

standard error =

Before we can begin developing a process for determining how reliable our point estimates are, we need some more vocabulary and notation.

a = probability of error

$1 - a$ = confidence level = probability that a random interval will capture the true value of the population parameter μ.
On this picture of a sampling distribution, the confidence level is $100(1 - \alpha)\%$ surrounding μ in the middle.



Because a normal distribution is symmetric, each tail must contain $\alpha/2$.

The left and right boundaries will be the negative and positive values of z that mark the lower and upper $100(\alpha/2)\%$.

Important note: The vocabulary above and the rest of this Lecture will apply to a **two-sided confidence interval**. We'll take a look at one-sided confidence intervals in Lecture 7.2.

Example A: Determine the value of $z_{a/2}$ for a (two-sided) confidence level of 80%.   *answer*: 1.28

Using a similar process, we could find corresponding values for common two-sided confidence levels.

$1-\alpha$	0.80	0.85	0.90	0.95	0.99
$z_{\alpha/2}$					

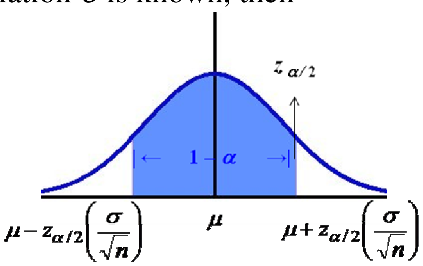
Back in section 5.4 we encountered the Central Limit Theorem, which assures us of three things.

1) As the sample size n increases, or as the number of trials n approaches infinite, the shape of a sampling distribution becomes increasingly like a normal distribution.

2) The mean of a sampling distribution = the mean of population: $E(\overline{X}) = \mu$.

3) The standard deviation of a sampling distribution = $\sigma(\overline{X}) = \dfrac{\sigma_X}{\sqrt{n}}$.

If we assume (for the moment) that a population has a normal distribution and that the population standard deviation σ is known, then



$$P\left[\ \mu - z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) \ < \ \overline{X} \ < \ \mu + z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)\ \right] = 1-\alpha \qquad (1)$$

Now we're going to do some mathematical manipulating of the formula on the right, line (1).

$$\mu - z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) \ < \ \overline{X}$$

$$\mu \ < \ \overline{X} + z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

$$\overline{X} \ < \ \mu + z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

$$\overline{X} - z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) \ < \ \mu$$

$$P\left[\ \overline{X} - z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) \ < \ \mu \ < \ \overline{X} + z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)\ \right] = 1-\alpha \qquad (2)$$

Lines (1) and (2) are mathematically equivalent!
Line (2) gives us the lower and upper limits of the $100(1-\alpha)\%$ *confidence interval.*

Example B – confidence interval: A sample has mean = 150 and the population has known standard deviation = 22. For a random sample of size 47, find the 90% two-sided confidence interval. *answer*: (144.721, 155.279)

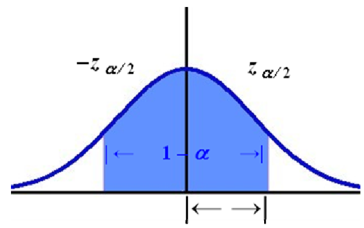
work:

interpretation

You should definitely read the text's explanation of "Interpreting a Confidence Interval".
highlights:

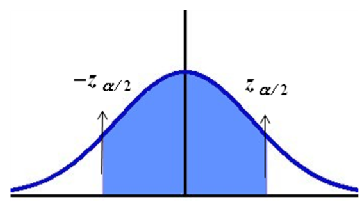
The probability involved in a confidence interval needs to be interpreted as a long-run relative frequency, which is why studies will be repeated by several researchers attempting to duplicate results. Part of hand-in homework #1 involves using Minitab to run a simulation to construct confidence intervals.

With a little further work, we can calculate the amount of a specified *margin of error*. Our question becomes, "The mean plus what amount takes us to the right-side boundary?" We can use the z-score formula for a sampling distribution to derive a formula for **margin of error**.



Important point: We're looking at a sampling distribution (which has a normal probability distribution) and are using a sample mean $\bar{x}$ (an established value) as our point estimate for the population parameter μ.

Example B – margin of error: A population has a known standard deviation = 22. For a random sample of size 47, find the 90% margin of error.
answer: ≈ 5.279



Earlier, it was noted that the Central Limit Theorem tells us, "As the sample size n increases, or as the number of trials n approaches infinite, the shape of a sampling distribution becomes increasingly like a normal distribution."

The question now becomes, "What value of n is 'large enough' for a desired confidence level"?
We'll answer by solving the margin of error formula for n, using w to represent the desired margin of error (width of the confidence interval).

Example B – sample size: A population has standard deviation $\sigma = 22$. If we wanted to be 99% confident in our results, with a margin of error of no more than 2, how large should our sample be?
answer: 806

Always take your result up to the next whole number to make sure the entire error margin is included.