ILRST/STSCI 2100 Discussion 8: Ch 9 Single Sample Estimation

Central limit theorem: for any distribution, with a sample size n sufficiently large, the sampling distribution of sample means is normally distributed.

Sufficiently large for sampling distribution of means: $n > 30$

For proportions, we want to ensure that $n \times p > 10$ AND $n \times (1 - p) > 10$, both must be shown to be true. In the case that we don't know the true population proportion, we can use $\hat{p}$ and we must show that $n \times \hat{p} > 10$ AND $n \times (1 - \hat{p}) > 10$.

The z-score for proportions is given by $Z = \dfrac{\hat{p} - p}{\sqrt{\dfrac{p(1 - p)}{n}}}$

This chapter introduces the concept of estimation.
The purpose of statistics is to make estimations about population parameters with a sample.
- The population mean μ is estimated by the sample mean $\bar{x}$
- The population variance σ^2 is estimated by the sample variance s^2
- The population median is estimated by the sample median

Those are called point estimates, because your answer is just one value.

Confidence Interval: Provides a range of estimates for the parameter.

The general formula for confidence interval is:

$$\text{point estimate} \pm z_{\alpha/2} * \text{SD}$$

So if estimating the population mean, then the confidence interval is given by:

$$\bar{x} \pm z_{\alpha/2} * \frac{\sigma}{\sqrt{n}}$$

And if estimating the population proportion, then the confidence interval is given by:

$$\hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}$$

For a 95% confidence interval, $\alpha = 0.05$, $\alpha/2 = 0.025$, so $z_{\alpha/2} = -1.96$

Interpretation: We are 95% confident that the true population mean is between $\bar{x} - z_{\alpha/2} * \frac{\sigma}{\sqrt{n}}$ and $\bar{x} + z_{\alpha/2} * \frac{\sigma}{\sqrt{n}}$.

Incorrect, but common misinterpretation: The true population mean falls in the range $\bar{x} - z_{\alpha/2} * \frac{\sigma}{\sqrt{n}}$ and $\bar{x} + z_{\alpha/2} * \frac{\sigma}{\sqrt{n}}$ with 95% probability. This is incorrect because the true population mean is fixed and doesn't change (it isn't random). Instead, $\bar{x}$ is the random variable.

Now suppose you don't know σ, the confidence interval is now:

$$\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$$

This assumes that you have $n > 30$, or if you have an approximately normal population.
Finding $t_{\alpha/2}$ depends on the degrees of freedom df, which is equal to $n - 1$. For a 95% confidence interval,

in table 3, you look under the 95% column because the area in the middle would be 95%.

t-distribution:
Similar to normal distribution: symmetric distribution, centered at 0, bell-curved. But, has heavier tails, and a lower peak compared to the normal curve. The shape changes based on the degrees of freedom, with greater degrees of freedom, the more normal it looks.

The expression to the right of $\pm$ is called the margin of error, denoted by B.

We can find the sample size by $n = \dfrac{z_{\alpha/2}^2 \sigma^2}{B^2}$, in which n is rounded up to the nearest integer.

If you're not given any information about the variance or standard deviation, you might have to use range/4 as an approximation.

For proportions:
The confidence interval can be generalized as (estimator) $\pm z_{\alpha/2} *$ SD of Sampling Distribution

So for proportions, to estimate the true population proportion, the point estimate is $\hat{p}$.

The standard deviation of sample proportions is $\sqrt{\dfrac{p(1-p)}{n}}$, but since p is unknown, we can use $\hat{p}$ as an estimator.

Therefore, the confidence interval for proportions is:

$$\hat{p} \pm z_{\alpha/2} * \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$$

To determine the sample size, you can use the formula:

$$n = p(1-p)\dfrac{z_{\alpha/2}^2}{B^2}$$

But again, you often don't know p, so you can use the sample proportion. If you don't know either, use $p = 0.5$ for the most conservative estimation.

If using $p = 0.5$, the formula then simplifies to:

$$n = \dfrac{1}{4}\dfrac{z_{\alpha/2}^2}{B^2}$$

Practice Problems:

1. Given $\bar{x} = 4$, $\sigma^2 = 16$, $n = 16$ find the following confidence intervals:
   95%
   Ans: $4 \pm 1.96 * \dfrac{\sqrt{16}}{\sqrt{16}} = (2.04, 5.96)$

   99%
   Ans: $4 \pm 2.58 * \dfrac{\sqrt{16}}{\sqrt{16}} = (1.42, 6.58)$

91%

Ans: $4 \pm 1.70 * \dfrac{\sqrt{16}}{\sqrt{16}} = (2.10, 5.70)$

What do you notice about the width of the confidence interval as the percent confidence increases? How can you explain this?

Ans: The width increases with increasing percent confidence. Intuitively, if you're trying to throw a ball into a box, you would have more confidence in the ball landing in the box if the box is bigger. Formally, the width is determined by the margin of error, which increases with increasing $z_{\alpha/2}$. $z_{\alpha/2}$ increases with increasing percent confidence.

2. For a certain population variance and percent confidence, the necessary sample size is 20 for a margin of error of 0.02. Suppose you want the margin of error to be 0.01, how many additional samples would you need (assuming same population variance and confidence level)?

Ans: Note that $n \propto \dfrac{1}{B^2}$, meaning if B doubles, n is quartered.

In this problem, B is halved, so n is quadrupled.

So the new sample size would be 4*20 = 80. And the new samples we need are 80 - 20 = 60.