ONE SAMPLE: MEANS (Quantitative variables)

Formula for confidence interval is

\[
\left( \bar{x} - z \frac{\sigma}{\sqrt{n}}, \bar{x} + z \frac{\sigma}{\sqrt{n}} \right)
\]

The margin of error is \( z \frac{\sigma}{\sqrt{n}} \).

For 95% confidence, \( z = 1.96 \), for 99% confidence \( z = 2.58 \).

If the standard deviation of the population, \( \sigma \), is not known, we use the sample standard deviation, \( s \), instead:

\[
\left( \bar{x} - t \frac{s}{\sqrt{n}}, \bar{x} + t \frac{s}{\sqrt{n}} \right)
\]

The quantity \( s/\sqrt{n} \) is the standard error.

Steps for Hypothesis Test

Step 1: Choose null and alternate hypotheses and significance level

Step 2: Construct test statistic (\( \bar{x} \) or \( \bar{p} \)) from the sample, and calculate the \( z \)-value or \( t \)-value assuming the null hypothesis.

Step 3: Calculate \( p \)-value

Step 4: Make a Decision: Reject or not to Reject the Null Hypothesis.

Notes:

Step 1:

- When choosing the null hypothesis, it must contain an equal sign (because the sampling distribution is based on it).
- The alternate hypothesis is generally what someone doing the experiment wants to show is true.
- One sided tests have \( H_a \) with < or >; Two sided tests have \( \neq \).

Step 2:

- Later we will have \( \chi^2 \) and \( F \) values, in addition to \( z \) and \( t \)-values.

Step 3:

- Make sure you know how to find the \( p \)-value for a one-sided test and a two-sided test. (For two-sided, \( p \)-values have a factor of 2.)

Step 4:

- We never accept \( H_0 \), just fail to reject it. We can’t show \( H_0 \) is true, only find, or fail to find, evidence against it.
TWO SAMPLES: MEANS (Quantitative variables)

**Matched Pairs:** Use differences for each individual, then a one-sample test.

**Two Independent Samples:** If the standard deviations of the two populations are $\sigma_1$ and $\sigma_2$, then the standard deviation of the sampling distribution of the difference in means, $D = \bar{x}_1 - \bar{x}_2$, is

$$\sigma_D = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

**Formula for confidence interval** for the difference in means is

$$\left( \bar{x}_1 - \bar{x}_2 - z \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, \bar{x}_1 - \bar{x}_2 + z \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right)$$

The margin of error is $z \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$.

For a Hypothesis Test, use

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

If the standard deviations of the two populations are unknown, we use the standard deviations of the two samples, $s_1$ and $s_2$, giving the standard error of the sampling distribution of the differences:

$$SE_D = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

so the confidence interval, with degree of freedom $= \min(n_1 - 1, n_2 - 1)$, is

$$\left( \bar{x}_1 - \bar{x}_2 - t \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}, \bar{x}_1 - \bar{x}_2 + t \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right)$$

For a hypothesis test, the degree of freedom is $\min(n_1 - 1, n_2 - 1)$ and

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$
**ONE SAMPLE: PROPORTIONS (Categorical Variable)**

For a categorical variable, a sample gives us a count, \( X \), and a proportion \( \hat{p} = \frac{X}{n} \). If \( n \) is reasonably large, (30 or more), then \( \hat{p} \) is normally distributed, with mean \( p \), the proportion in the population, and standard deviation

\[
\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}.
\]

If you often don’t know \( p \), you can use \( \hat{p} \) or 0.5. This gives the **standard error**, which is an approximation to the standard deviation

\[
SE_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \quad \text{or} \quad SE_{\hat{p}} = \sqrt{\frac{0.5(1-0.5)}{n}}.
\]

(Using 0.5 gives the largest standard error for that sample size.) The **confidence interval** is:

\[
\left( \hat{p} - z \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right) \quad \text{or} \quad \left( \hat{p} - z \sqrt{\frac{0.5(1-0.5)}{n}}, \hat{p} + z \sqrt{\frac{0.5(1-0.5)}{n}} \right)
\]

The **margin of error** is

\[
z \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.
\]

We do **Hypothesis Tests** using:

\[
z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}
\]

Note: In the confidence interval, you do not know \( p \) and so you use \( \hat{p} \) and the standard error. For a hypothesis test, you use the value of \( p \) in the null hypothesis for the standard deviation.

**TWO SAMPLES: PROPORTIONS (Categorical Variables)**

The standard deviation of the difference in proportions, \( D = \hat{p}_1 - \hat{p}_2 \), is

\[
\sigma_D = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}
\]

If we don’t know \( p_1 \) and \( p_2 \), we use the standard error:

\[
SE_D = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}
\]

Then the **confidence interval** for the difference in proportions in the population is

\[
\left( \hat{p}_1 - \hat{p}_2 - z \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}, \hat{p}_1 - \hat{p}_2 + z \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \right)
\]

**Hypothesis tests** use:

\[
z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p}) \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}}, \text{ where the pooled proportion } \hat{p} = \frac{X_1 + X_2}{n_1 + n_2}
\]