

10.22: Putting It Together- Inference for Means

Let's Summarize

The focus of this module, *Inference for Means*, is inference for a population mean or a difference between two populations means. We began this module with a discussion of the sampling distribution of sample means. We then developed a probability model based on this sampling distribution. We used the probability model with an actual sample mean to test a claim about population mean in a hypothesis test or to estimate a population mean with a confidence interval. We then moved to inference for a difference in two population means (or a treatment effect.)

Sampling Distribution of Means

If we have a quantitative data set from a population with mean μ and standard deviation σ , the model for the theoretical sampling distribution of means of all random samples of size n has the following properties:

- The mean of the sampling distribution of means is μ .
- The standard deviation of the sampling distribution of means is $\frac{\sigma}{\sqrt{n}}$.
 - Notice that as n grows, the standard error of the sampling distribution of means shrinks. That means that larger samples give more accurate estimates of a population mean.
- For large enough sample size, the sampling distribution of means is approximately normal (even if population is not normal). This is called the *central limit theorem*.
 - If a variable has a skewed distribution for individuals in the population, a larger sample size is needed to ensure that the sampling distribution has a normal shape.
 - The general rule is that if n is at least 30, then the sampling distribution of means will be approximately normal. However, if the population is already normal, then any sample size will produce a normal sampling distribution.
- We practiced finding a probability associated with a range of sample means, which is similar to finding a P-value in hypothesis testing. The process is as follows.
 - Convert a sample mean \bar{X} into a z-score: $Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$
 - Use technology to find a probability associated with a given range of z-scores.

Confidence Intervals

The Form

A confidence interval approximates a population mean by giving us a range of values that likely contains the population mean μ . The general form of the confidence interval is

$$\bar{x} \pm (\text{margin of error}) = \bar{x} \pm (\text{critical value}) \cdot (\text{standard error})$$

We covered three different types of confidence intervals:

One-sample Z-interval: $\bar{x} \pm Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$, where σ is the population standard deviation (when it is known).

One-sample T-interval: $\bar{x} \pm T_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$, where s is the sample standard deviation.

Two-sample T-interval: $(\bar{x}_1 - \bar{x}_2) \pm T_{\alpha/2} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$, where we use the sample statistics from two independent samples.

The T-Model

When the standard deviation of the population is unknown, which is often the case, we use the T-model to find the critical values. When using the T-model to find critical values, we need to select an appropriate number of degrees of freedom.

- In the one-sample case, the number of degrees of freedom is 1 less than the sample size ($df = n - 1$).
- In the two-independent-sample case, the degrees of freedom come from a complicated formula, and we often use technology to find df .

Conclusions

To say we are 95% confident that the population mean falls within our confidence interval really means that about 95% of all confidence intervals computed in this way will capture the true population mean.

Conditions

The population must be normally distributed, or the sample size must be large enough (larger than 30). In the case of the two-sample T-interval, both populations/samples must meet these conditions. In practice, we use T-procedures with smaller samples if the distribution of the variable in the sample(s) is not heavily skewed and is without outliers. We take this as an indication that the variable has a fairly normal distribution in the population(s).

Observations about Confidence Interval Structure

- As we saw with other confidence intervals, the width of a confidence interval is twice the margin of error. The smaller the margin of error, the narrower the confidence interval and the more precise the estimate of the population parameter.
- Increasing the confidence level decreases the precision (larger margin of error, so wider interval). Decreasing the confidence level increases the precision (smaller margin of error, so narrower interval).
- Confidence intervals are useful estimates only when they provide a good balance of confidence level and precision. In order to increase precision without losing confidence, we must increase the sample size. In other words, larger samples provide more precise estimates without sacrificing confidence.

Hypothesis Testing (Tests for Statistical Significance)

The process of any hypothesis test consists of four basic steps:

- Define the hypotheses
- Collect the data: We need random samples that are representative of the population. For the two-sample T-test, the samples must be independent.
- Assess the evidence: Assessment includes checking appropriate conditions, computing test statistics, and finding corresponding P-values.
- State the conclusion: We compare the P-value to α , decide whether or not to reject H_0 , then state conclusion in context.

Hypotheses

- **The null hypothesis (H_0):** The null hypothesis gives the value of the parameter we use to create the sampling distribution. In this way, the null hypothesis states what we assume to be true about the population.
- **The alternative hypothesis (H_a):** The alternative hypothesis usually reflects the claim in the research question about the value of the parameter. The alternative hypothesis says the parameter is greater than or less than or not equal to the value we assume to be true in the null hypothesis.
 - When H_a is $\mu < \mu_0$ or $\mu > \mu_0$, the test is called a one-tailed test.
 - For the paired T-test, H_0 would look like $\mu < 0$ or $\mu > 0$ in the case of a one-tailed test.
 - For the two-sample T-test, H_0 would look like $\mu_1 - \mu_2 < 0$ or $\mu_1 - \mu_2 > 0$ in the case of a one-tailed test.
 - When H_a is $\mu \neq \mu_0$, the test is called a two-tailed test.
 - For the paired T-test, H_a would look like $\mu \neq 0$ in the case of a two-tailed test.
 - For the two-sample T-test, H_a would look like $\mu_1 - \mu_2 \neq 0$ in the case of a two-tailed test.

Conditions

Conditions that must be satisfied in order to carryout T-procedures are as follows:

- The population is normally distributed, or the sample is large (at least 30). This applies to both populations for the two-sample T-test.
- The samples must be random in order to avoid bias.
- The samples must be independent in the case of the two-sample T-test.

Test Statistic

The test T-statistic is given by

$$T = \frac{\text{sample statistic} - \text{hypothesized parameter}}{\text{standard error}}$$

We've learned about three different types of T-tests:

One-sample T-test:

$$T = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$

Paired T-test: We calculate the differences, then find the mean and standard deviation.

$$T = \frac{\bar{d} - 0}{s_d / \sqrt{n}}$$

Two-sample T-test:

$$T = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

P-values

The P-value is the probability of finding a random sample with a test statistic at least as extreme as ours, assuming that the null hypothesis is true. We find P-values by using the T-distribution.

To come to a conclusion about H_0 , we compare the P-value to the significance level, α .

- If $P \leq \alpha$, we reject H_0 and conclude there is significant evidence in favor of H_a .
- If $P > \alpha$, we fail to reject H_0 and conclude the sample does not provide significant evidence in favor of H_a .

Error Types

Hypothesis tests are based on random samples, so the conclusions are really statements about probabilities, and it is possible for the conclusions to be wrong.

- If our test results in rejecting a null hypothesis that is actually true, it is called a type I error.
- If our test results in failing to reject a null hypothesis that is actually false, it is called a type II error.

You are now ready to practice what you learned in this module by doing a StatTutor exercise. We design StatTutor exercises to help you apply what you have learned to a real-life data analysis question.

Instructions: One of the first few screens in StatTutor contains a link to download the data set for this StatTutor exercise. When you click that link, a pop-up window will appear asking if you want to open or save the file. Make sure you click “Save,” which allows you to save the file to your hard drive. Then find the downloaded file and double-click it to open it if you’re using R, Minitab, Excel, or StatCrunch, or transfer it to your calculator if you’re using the TI Calculator.

If you are using StatCrunch, please see [Additional Instructions for StatCrunch](#).

Are You Ready for the Checkpoint?

If you completed all of the exercises in this module, you should be ready for the Checkpoint. To make sure that you are ready for the Checkpoint, use the My Response link below to evaluate your understanding of the learning objectives for this module and to submit questions that you may have.

CC licensed content, Shared previously

- Concepts in Statistics. **Provided by:** Open Learning Initiative. **Located at:** <http://oli.cmu.edu>. **License:** [CC BY: Attribution](#)

This page titled [10.22: Putting It Together- Inference for Means](#) is shared under a [CC BY 4.0](#) license and was authored, remixed, and/or curated by [Bill Pelz](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.