Hypothesis Test: Difference Between Means

This lesson explains how to conduct a hypothesis test for the difference between two means. The test procedure, called the **two-sample t-test**, is appropriate when the following conditions are met:

- The sampling method for each sample is simple random sampling.
- The samples are independent.
- Each population is at least 20 times larger than its respective sample.
- The sampling distribution is approximately normal, which is generally the case if any of the following conditions apply.
  - The population distribution is normal.
  - The population data are symmetric, unimodal, without outliers, and the sample size is 15 or less.
  - The population data are slightly skewed, unimodal, without outliers, and the sample size is 16 to 40.
  - The sample size is greater than 40, without outliers.

This approach consists of four steps: (1) state the hypotheses, (2) formulate an analysis plan, (3) analyze sample data, and (4) interpret results.

State the Hypotheses

Every hypothesis test requires the analyst to state a null hypothesis and an alternative hypothesis. The hypotheses are stated in such a way that they are mutually exclusive. That is, if one is true, the other must be false; and vice versa.

The table below shows three sets of null and alternative hypotheses. Each makes a statement about the difference *d* between the mean of one population μ_1 and the mean of another population μ_2. (In the table, the symbol ≠ means " not equal to ".)

Set	Null hypothesis	Alternative hypothesis	Number of tails
1	$\mu_1 - \mu_2 = d$	$\mu_1 - \mu_2 \neq d$	2
2	$\mu_1 - \mu_2 \geq d$	$\mu_1 - \mu_2 < d$	1
3	$\mu_1 - \mu_2 \leq d$	$\mu_1 - \mu_2 > d$	1

The first set of hypotheses (Set 1) is an example of a two-tailed test, since an extreme value on either side of the sampling distribution would cause a researcher to reject the null hypothesis. The other two sets of hypotheses (Sets 2 and 3) are one-tailed tests, since an extreme value on only one side of the sampling distribution would cause a researcher to reject the null hypothesis.

When the null hypothesis states that there is no difference between the two population means (i.e., d = 0), the null and alternative hypothesis are often stated in the following form.

$$H_0: \mu_1 = \mu_2$$
$$H_a: \mu_1 \neq \mu_2$$

Formulate an Analysis Plan

The analysis plan describes how to use sample data to accept or reject the null hypothesis. It should specify the following elements.

- Significance level. Often, researchers choose significance levels equal to 0.01, 0.05, or 0.10; but any value between 0 and 1 can be used.
- Test method. Use the two-sample t-test to determine whether the difference between means found in the sample is significantly different from the hypothesized difference between means.

Analyze Sample Data

Using sample data, find the standard error, degrees of freedom, test statistic, and the P-value associated with the test statistic.

- Standard error. Compute the standard error (SE) of the sampling distribution.

  $$SE = \text{sqrt}[\ (s_1^2/n_1) + (s_2^2/n_2)\ ]$$

  where s_1 is the standard deviation of sample 1, s_2 is the standard deviation of sample 2, n_1 is the size of sample 1, and n_2 is the size of sample 2.

- Degrees of freedom. The degrees of freedom (DF) is:

  $$DF = (s_1^2/n_1 + s_2^2/n_2)^2 / \{\ [\ (s_1^2 / n_1)^2 / (n_1 - 1)\ ] + [\ (s_2^2 / n_2)^2 / (n_2 - 1)\ ]\ \}$$

  If DF does not compute to an integer, round it off to the nearest whole number. Some texts suggest that the degrees of freedom can be approximated by the smaller of $n_1 - 1$ and $n_2 - 1$; but the above formula gives better results.

- Test statistic. The test statistic is a t statistic (t) defined by the following equation.

  $$t = [\ (\bar{x}_1 - \bar{x}_2) - d\ ] / SE$$

  where $\bar{x}_1$ is the mean of sample 1, $\bar{x}_2$ is the mean of sample 2, d is the hypothesized difference between population means, and SE is the standard error.

- P-value. The P-value is the probability of observing a sample statistic as extreme as the test statistic. Since the test statistic is a t statistic, use the t Distribution Calculator to assess the probability associated with the t statistic, having the degrees of freedom computed above. (See sample problems at the end of this lesson for examples of how this is done.)

Interpret Results

If the sample findings are unlikely, given the null hypothesis, the researcher rejects the null hypothesis. Typically, this involves comparing the P-value to the significance level, and rejecting the null hypothesis when the P-value is less than the significance level.

Test Your Understanding

In this section, two sample problems illustrate how to conduct a hypothesis test of a difference between mean scores. The first problem involves a two-tailed test; the second problem, a one-tailed test.

Problem 1: Two-Tailed Test

Within a school district, students were randomly assigned to one of two Math teachers - Mrs. Smith and Mrs. Jones. After the assignment, Mrs. Smith had 30 students, and Mrs. Jones had 25 students.

At the end of the year, each class took the same standardized test. Mrs. Smith's students had an average test score of 78, with a standard deviation of 10; and Mrs. Jones' students had an average test score of 85, with a standard deviation of 15.

Test the hypothesis that Mrs. Smith and Mrs. Jones are equally effective teachers. Use a 0.10 level of significance. (Assume that student performance is approximately normal.)

Solution: The solution to this problem takes four steps: (1) state the hypotheses, (2) formulate an analysis plan, (3) analyze sample data, and (4) interpret results. We work through those steps below:

- **State the hypotheses.** The first step is to state the null hypothesis and an alternative hypothesis.

  Null hypothesis: $\mu_1 - \mu_2 = 0$
  Alternative hypothesis: $\mu_1 - \mu_2 \neq 0$

  Note that these hypotheses constitute a two-tailed test. The null hypothesis will be rejected if the difference between sample means is too big or if it is too small.

- **Formulate an analysis plan**. For this analysis, the significance level is 0.10. Using sample data, we will conduct a two-sample t-test of the null hypothesis.

- **Analyze sample data**. Using sample data, we compute the standard error (SE), degrees of freedom (DF), and the t statistic test statistic (t).

  $$SE = \text{sqrt}[(s_1^2/n_1) + (s_2^2/n_2)]$$
  $$SE = \text{sqrt}[(10^2/30) + (15^2/25] = \text{sqrt}(3.33 + 9) = \text{sqrt}(12.33) = 3.51$$

$$DF = (s_1^2/n_1 + s_2^2/n_2)^2 / \{ [(s_1^2 / n_1)^2 / (n_1 - 1)] + [(s_2^2 / n_2)^2 / (n_2 - 1)] \}$$
$$DF = (10^2/30 + 15^2/25)^2 / \{ [(10^2 / 30)^2 / (29)] + [(15^2 / 25)^2 / (24)] \}$$
$$DF = (3.33 + 9)^2 / \{ [(3.33)^2 / (29)] + [(9)^2 / (24)] \} = 152.03 / (0.382 + 3.375) = 152.03/3.757 = 40.47$$

$$t = [(\bar{x}_1 - \bar{x}_2) - d] / SE = [(78 - 85) - 0] / 3.51 = -7/3.51 = -1.99$$

where s_1 is the standard deviation of sample 1, s_2 is the standard deviation of sample 2, n_1 is the size of sample 1, n_2 is the size of sample 2, $\bar{x}_1$ is the mean of sample 1, $\bar{x}_2$ is the mean of sample 2, d is the hypothesized difference between the population means, and SE is the standard error.

Since we have a two-tailed test, the P-value is the probability that a t statistic having 40 degrees of freedom is more extreme than -1.99; that is, less than -1.99 or greater than 1.99.

We use the t Distribution Calculator to find $P(t < -1.99) = 0.027$, and $P(t > 1.99) = 0.027$. Thus, the P-value = 0.027 + 0.027 = 0.054.

- **Interpret results**. Since the P-value (0.054) is less than the significance level (0.10), we cannot accept the null hypothesis.

Note: If you use this approach on an exam, you may also want to mention why this approach is appropriate. Specifically, the approach is appropriate because the sampling method was simple random sampling, the samples were independent, the sample size was much smaller than the population size, and the samples were drawn from a normal population.

Problem 2: One-Tailed Test

The Acme Company has developed a new battery. The engineer in charge claims that the new battery will operate continuously for *at least* 7 minutes longer than the old battery.

To test the claim, the company selects a simple random sample of 100 new batteries and 100 old batteries. The old batteries run continuously for 190 minutes with a standard deviation of 20 minutes; the new batteries, 200 minutes with a standard deviation of 40 minutes.

Test the engineer's claim that the new batteries run at least 7 minutes longer than the old. Use a 0.05 level of significance. (Assume that there are no outliers in either sample.)

Solution: The solution to this problem takes four steps: (1) state the hypotheses, (2) formulate an analysis plan, (3) analyze sample data, and (4) interpret results. We work through those steps below:

- **State the hypotheses.** The first step is to state the null hypothesis and an alternative hypothesis.

  Null hypothesis: $\mu_1 - \mu_2 >= 7$
  Alternative hypothesis: $\mu_1 - \mu_2 < 7$

  Note that these hypotheses constitute a one-tailed test. The null hypothesis will be rejected if the mean difference between sample means is too small.

- **Formulate an analysis plan**. For this analysis, the significance level is 0.05. Using sample data, we will conduct a two-sample t-test of the null hypothesis.

- **Analyze sample data**. Using sample data, we compute the standard error (SE), degrees of freedom (DF), and the t statistic test statistic (t).

$$SE = \text{sqrt}[(s_1^2/n_1) + (s_2^2/n_2)]$$
$$SE = \text{sqrt}[(40^2/100) + (20^2/100] = \text{sqrt}(16 + 4) = 4.472$$

$$DF = (s_1^2/n_1 + s_2^2/n_2)^2 / \{ [(s_1^2 / n_1)^2 / (n_1 - 1)] + [(s_2^2 / n_2)^2 / (n_2 - 1)] \}$$
$$DF = (40^2/100 + 20^2/100)^2 / \{ [(40^2 / 100)^2 / (99)] + [(20^2 / 100)^2 / (99)] \}$$
$$DF = (20)^2 / \{ [(16)^2 / (99)] + [(2)^2 / (99)] \} = 400 / (2.586 + 0.162) = 145.56$$

$$t = [(\bar{x}_1 - \bar{x}_2) - d] / SE = [(200 - 190) - 7] / 4.472 = 3/4.472 = 0.67$$

where s_1 is the standard deviation of sample 1, s_2 is the standard deviation of sample 2, n_1 is the size of sample 1, n_2 is the size of sample 2, $\bar{x}_1$ is the mean of sample 1, $\bar{x}_2$ is the mean of sample 2, d is the hypothesized difference between population means, and SE is the standard error.

Here is the logic of the analysis: Given the alternative hypothesis ($\mu_1 - \mu_2 < 7$), we want to know whether the observed difference in sample means is small enough (i.e., sufficiently less than 7) to cause us to reject the null hypothesis.

The observed difference in sample means (10) produced a t statistic of 0.67. We use the t Distribution Calculator to find $P(t \leq 0.67) = 0.75$.

This means we would expect to find an observed difference in sample means of 10 or less in 75% of our samples, if the true difference were actually 7. Therefore, the P-value in this analysis is 0.75.

- **Interpret results**. Since the P-value (0.75) is greater than the significance level (0.05), we cannot reject the null hypothesis.

Note: If you use this approach on an exam, you may also want to mention why this approach is appropriate. Specifically, the approach is appropriate because the sampling method was simple random sampling, the samples were independent, the sample size was much smaller than the population size, and the sample size was large without outliers.

< Previous lesson Next lesson >