

## 10.19: Hypothesis Test for a Difference in Two Population Means (1 of 2)

### Learning Objectives

- Under appropriate conditions, conduct a hypothesis test about a difference between two population means. State a conclusion in context.

### The Hypothesis Test for a Difference in Two Population Means

The general steps of this hypothesis test are the same as always. As expected, the details of the conditions for use of the test and the test statistic are unique to this test (but similar in many ways to what we have seen before.)

#### Step 1: Determine the hypotheses.

The hypotheses for a difference in two population means are similar to those for a difference in two population proportions. The null hypothesis,  $H_0$ , is again a statement of “no effect” or “no difference.”

$H_0: \mu_1 - \mu_2 = 0$ , which is the same as  $H_0: \mu_1 = \mu_2$

The alternative hypothesis,  $H_a$ , can be any one of the following.

$H_a: \mu_1 - \mu_2 < 0$ , which is the same as  $H_a: \mu_1 < \mu_2$

$H_a: \mu_1 - \mu_2 > 0$ , which is the same as  $H_a: \mu_1 > \mu_2$

$H_a: \mu_1 - \mu_2 \neq 0$ , which is the same as  $H_a: \mu_1 \neq \mu_2$

#### Step 2: Collect the data.

As usual, how we collect the data determines whether we can use it in the inference procedure. We have our usual two requirements for data collection.

- Samples must be random to remove or minimize bias.
- Samples must be representative of the populations in question.

We use this hypothesis test when the data meets the following conditions.

- The two *random* samples are *independent*.
- The variable is *normally distributed in both populations*. If this variable is not known, *samples of more than 30* will have a difference in sample means that can be modeled adequately by the t-distribution. As we discussed in “Hypothesis Test for a Population Mean,” t-procedures are robust even when the variable is not normally distributed in the population. If checking normality in the populations is impossible, then we look at the distribution in the samples. If a histogram or dotplot of the data does not show extreme skew or outliers, we take it as a sign that the variable is not heavily skewed in the populations, and we use the inference procedure. (Note: This is the same condition we used for the one-sample t-test in “Hypothesis Test for a Population Mean.”)

#### Step 3: Assess the evidence.

If the conditions are met, then we calculate the t-test statistic. The t-test statistic has a familiar form.



$$t = \frac{\text{Observed difference in sample means} - \text{Hypothesized difference in population means}}{\text{Standard error}}$$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Since the null hypothesis assumes there is no difference in the population means, the expression  $(\mu_1 - \mu_2)$  is always zero.

As we learned in “Estimating a Population Mean,” the t-distribution depends on the **degrees of freedom (df)**. In the one-sample and matched-pair cases  $df = n - 1$ . For the two-sample t-test, determining the correct  $df$  is based on a complicated formula that we do not cover in this course. We will either give the  $df$  or use technology to find the  $df$ . With the t-test statistic and the degrees of freedom, we can use the appropriate t-model to find the P-value, just as we did in “Hypothesis Test for a Population Mean.” We can even use the same simulation.

#### Step 4: State a conclusion.

To state a conclusion, we follow what we have done with other hypothesis tests. We compare our P-value to a stated level of significance.

- If the P-value  $\leq \alpha$ , we reject the null hypothesis in favor of the alternative hypothesis.
- If the P-value  $> \alpha$ , we fail to reject the null hypothesis. We do not have enough evidence to support the alternative hypothesis.

As always, we state our conclusion in context, usually by referring to the alternative hypothesis.

### Example

#### “Context and Calories”

Does the company you keep impact what you eat? This example comes from an article titled “Impact of Group Settings and Gender on Meals Purchased by College Students” (Allen-O'Donnell, M., T. C. Nowak, K. A. Snyder, and M. D. Cottingham, *Journal of Applied Social Psychology* 49(9), 2011, [onlinelibrary.wiley.com/doi/10.1111/j.1559-1816.2011.00804.x/full](http://onlinelibrary.wiley.com/doi/10.1111/j.1559-1816.2011.00804.x/full)). In this study, researchers examined this issue in the context of gender-related theories in their field. For our purposes, we look at this research more narrowly.

#### Step 1: Stating the hypotheses.

In the article, the authors make the following hypothesis. “The attempt to appear feminine will be empirically demonstrated by the purchase of fewer calories by women in mixed-gender groups than by women in same-gender groups.” We translate this into a simpler and narrower research question: *Do women purchase fewer calories when they eat with men compared to when they eat with women?*

Here the two populations are “women eating with women” (population 1) and “women eating with men” (population 2). The variable is the calories in the meal. We test the following hypotheses at the 5% level of significance.

The null hypothesis is always  $H_0: \mu_1 - \mu_2 = 0$ , which is the same as  $H_0: \mu_1 = \mu_2$ .

The alternative hypothesis  $H_a: \mu_1 - \mu_2 > 0$ , which is the same as  $H_a: \mu_1 > \mu_2$ .

Here  $\mu_1$  represents the mean number of calories ordered by women when they were eating with other women, and  $\mu_2$  represents the mean number of calories ordered by women when they were eating with men.

Note: It does not matter which population we label as 1 or 2, but once we decide, we have to stay consistent throughout the hypothesis test. Since we expect the number of calories to be greater for the women eating with other women, the difference is positive if “women eating with women” is population 1. If you prefer to work with positive numbers, choose the group with the larger expected mean as population 1. This is a good general tip.

#### Step 2: Collect Data.

As usual, there are two major things to keep in mind when considering the collection of data.

- Samples need to be representative of the population in question.
- Samples need to be random in order to remove or minimize bias.

##### Representative Samples?

The researchers state their hypothesis in terms of “women.” We did the same. But the researchers gathered data by watching people eat at the HUB Rock Café II on the campus of Indiana University of Pennsylvania during the Spring semester of 2006. Almost all of the women in the data set were white undergraduates between the ages of 18 and 24, so there are some definite limitations on the scope of this study. These limitations will affect our conclusion (and the specific definition of the population means in our hypotheses.)

##### Random Samples?

The observations were collected on February 13, 2006, through February 22, 2006, between 11 a.m. and 7 p.m. We can see that the researchers included both lunch and dinner. They also made observations on all days of the week to ensure that weekly customer patterns did not confound their findings. The authors state that “since the time period for observations and the place where [they] observed students were limited, the sample was a convenience sample.” Despite these limitations, the researchers conducted inference procedures with the data, and the results were published in a reputable journal. We will also conduct inference with this data, but we also include a discussion of the limitations of the study with our conclusion. The authors did this, also.

##### Do the data met the conditions for use of a t-test?

The researchers reported the following sample statistics.

- In a sample of 45 women dining with other women, the average number of calories ordered was 850, and the standard deviation was 252.
- In a sample of 27 women dining with men, the average number of calories ordered was 719, and the standard deviation was 322.

One of the samples has fewer than 30 women. We need to make sure the distribution of calories in this sample is not heavily skewed and has no outliers, but we do not have access to a spreadsheet of the actual data. Since the researchers conducted a t-test with this data, we will assume that the conditions are met. This includes the assumption that the samples are independent.

### Step 3: Assess the evidence.

As noted previously, the researchers reported the following sample statistics.

- In a sample of 45 women dining with other women, the average number of calories ordered was 850, and the standard deviation was 252.
- In a sample of 27 women dining with men, the average number of calories ordered was 719, and the standard deviation was 322.

To compute the t-test statistic, make sure sample 1 corresponds to population 1. Here our population 1 is “women eating with other women.” So  $x_1 = 850$ ,  $s_1 = 252$ ,  $n_1 = 45$ , and so on.



$$T = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{850 - 719}{\sqrt{\frac{252^2}{45} + \frac{322^2}{27}}} \approx \frac{131}{72.47} \approx 1.81$$

Using technology, we determined that the degrees of freedom are about 45 for this data. To find the P-value, we use our familiar simulation of the t-distribution. Since the alternative hypothesis is a “greater than” statement, we look for the area to the right of  $T = 1.81$ . The P-value is 0.0385.

### Step 4: State a conclusion.

#### Generic Conclusion

The hypotheses for this test are  $H_0: \mu_1 - \mu_2 = 0$  and  $H_a: \mu_1 - \mu_2 > 0$ . Since the P-value is less than the significance level ( $0.0385 < 0.05$ ), we reject  $H_0$  and accept  $H_a$ .

#### Conclusion in context

At Indiana University of Pennsylvania, the mean number of calories ordered by undergraduate women eating with other women is greater than the mean number of calories ordered by undergraduate women eating with men (P-value = 0.0385).

#### A Comment about Conclusions

In the conclusion above, we did not generalize the findings to all women. Since the samples included only undergraduate women at one university, we included this information in our conclusion. But our conclusion is a cautious statement of the findings. The authors see the results more broadly in the context of theories in the field of social psychology. In the context of these theories, they write, “Our findings support the assertion that meal size is a tool for influencing the impressions of others. For traditional-age, predominantly White college women, diminished meal size appears to be an attempt to assert femininity in groups that include men.” This viewpoint is echoed in the following summary of the study for the general public on National Public Radio (npr.org).

*Both men and women appear to choose larger portions when they eat with women, and both men and women choose smaller portions when they eat in the company of men, according to new research published in the Journal of Applied Social Psychology. The study, conducted among a sample of 127 college students, suggests that both men and women are influenced by unconscious scripts about how to behave in each other’s company. And these scripts change the way men and women eat when they eat together and when they eat apart.*

Should we be concerned that the findings of this study are generalized in this way? Perhaps. But the authors of the article address this concern by including the following disclaimer with their findings: “While the results of our research are suggestive, they should be replicated with larger, representative samples. Studies should be done not only with primarily White, middle-class college students, but also with students who differ in terms of race/ethnicity, social class, age, sexual orientation, and so forth.” This is an example of good statistical practice. It is often very difficult to select truly random samples from the populations of interest. Researchers therefore discuss the limitations of their sampling design when they discuss their conclusions.

In the following activities, you will have the opportunity to practice parts of the hypothesis test for a difference in two population means. On the next page, the activities focus on the entire process and also incorporate technology.

Learn By Doing

## National Health and Nutrition Survey

<https://assessments.lumenlearning.co...sessments/3705>

<https://assessments.lumenlearning.co...sessments/3782>

<https://assessments.lumenlearning.co...sessments/3706>

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 [10.19: Hypothesis Test for a Difference in Two Population Means \(1 of 2\)](#) 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.