

Two-Sample Problems

Diana Mindrila, Ph.D.

Phoebe Balentyne, M.Ed.

Based on Chapter 19 of The Basic Practice of Statistics (6th ed.)

Concepts:

- Two-Sample Problems
- Comparing Two Population Means
- Two-Sample t Procedures
- Using Technology
- Robustness

Objectives:

- Describe the conditions necessary for inference.
- Check the conditions necessary for inference.
- Perform two-sample t procedures.
- Describe the robustness of the t procedures.

References:

Moore, D. S., Notz, W. I, & Flinger, M. A. (2013). *The basic practice of statistics* (6th ed.). New York, NY: W. H. Freeman and Company.

Two-Sample Problems

Researchers may want to compare two independent groups. With matched samples, the same individuals are tested twice, or pairs of individuals who are very similar in some respect are tested.

- **Independent samples** consist of two groups of individuals who are randomly selected from two different populations.
- The term “independent” is used because the individuals in one sample must be completely unrelated to the individuals in the other sample.
- Example: to find out if test scores are significantly different between males and females, researchers would need to randomly select a group of females and randomly select a group of males. There are two groups and they come from two separate populations (one population is males and the other separate population is females). Each of these populations has a mean for the variable.

Population Mean	Sample Mean	Sample Size
μ_1	\bar{x}_1	n_1
μ_2	\bar{x}_2	n_2

- ❖ The parameters of interest are the population means, which are denoted μ_1 and μ_2 .
- ❖ The sample means are denoted as \bar{x}_1 and \bar{x}_2 .
- ❖ The sample sizes are denoted as n_1 and n_2 .

Conditions for Inference Comparing Two Means

Before conducting any statistical analyses, two assumptions must be met:

- 1) The two samples are random and they come from two distinct populations. The samples are independent. That is, one sample has no influence on the other. Matching violates independence, for example. Additionally, the same response variable must be measured for both samples.
- 2) Both populations are Normally distributed. The means and standard deviations of the populations are unknown. In practice, it is enough that the distributions have similar shapes and that the data have no strong outliers.

The Two-Sample t Statistic

When data come from two random samples or two groups in a randomized experiment, the difference between the sample means ($\bar{x}_1 - \bar{x}_2$) is the best estimate of the difference between the population means ($\mu_1 - \mu_2$).

- In other words, since the population means (μ_1 and μ_2) are unknown, the sample means (\bar{x}_1 and \bar{x}_2) must be used to make inferences.
- The inferences that are being made are based on the differences between the sample means: $\bar{x}_1 - \bar{x}_2$

When the Independent condition is met, the standard deviation of the difference $\bar{x}_1 - \bar{x}_2$ is:

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

- However, this formula requires the population standard deviations to be known. If these are unknown, t procedures must be used to make inferences.

If the values of the parameters σ_1 and σ_2 (the population standard deviations) are unknown, they can be replaced with the sample standard deviations. The result is the **standard error** of the difference $\bar{x}_1 - \bar{x}_2$:

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

Degrees of Freedom

- The shape of the t distribution is different for different sample sizes.
- Therefore, when making inferences about the difference between two population means, the size of the two samples must be taken into account.
- This is because the t distribution is used to make these inferences.

There are two ways of computing degrees of freedom:

- 1) Use technology.
- 2) Choose the smaller of: $df_1 = n - 1$ and $df_2 = n - 1$
 - Subtract 1 from each sample size.
 - Use the degrees of freedom that is the smallest (from the smaller sample size).

Confidence Interval for $\mu_1 - \mu_2$

- In some situations, the confidence interval for the difference between the two populations must be estimated.
- To obtain this confidence interval, compute the difference between the two sample means and then add and subtract the margin of error to obtain the upper and lower limit of this interval.

Two-Sample t Interval for a Difference Between Means

When the Random, Normal, and Independent conditions are met, a level C confidence interval for $(\mu_1 - \mu_2)$ is

$$CI = (\bar{x}_1 - \bar{x}_2) \pm t^* \underbrace{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}_{\text{margin of error}}$$

standard error

where t^* is the critical value for confidence level C for the t distribution.

- The margin of error is obtained by multiplying the standard error by t^* .
- T^* can be obtained from a table in a statistical textbook and is found using the confidence level desired and the degrees of freedom for the particular sample.

Confidence Intervals: Example

The following example illustrates how to estimate a confidence interval for the difference between two groups.

- The goal is to make an inference about the difference between two populations (males and females) based on their performance on a certain test.
- To make inferences about each population as a whole, a random sample from the population of females and a random sample from the population of males must be selected.
- Although separate selection procedures are needed, the data sets do not need to be separated. Individuals from both samples can be included in the same data file, as shown below.

Question: *In the population, what is the difference between the females' average score and the males' average score on the test?*

Sample 1: Females:

$$n_1=8 \quad \bar{x}_1=97.25 \quad s_1=3.65$$

Sample 2: Males:

$$n_2=12 \quad \bar{x}_2=87.25 \quad s_2=9.60$$

Gender	N	Mean	Std. Deviation	Std. Error Mean
Test scores Female	8	97.2500	3.65474	1.29215
Test scores Male	12	87.2500	9.60232	2.77195

Name	Test Score	Gender
Dan	95.00	2.00
Mimi	100.00	1.00
Sam	78.00	2.00
Gene	68.00	2.00
Lena	100.00	1.00
Richard	95.00	2.00
Dorian	98.00	2.00
Ernest	79.00	2.00
John	98.00	2.00
Linda	95.00	1.00
Martha	90.00	1.00
Geta	95.00	1.00
Delia	98.00	1.00
Damian	86.00	2.00
Sylvia	100.00	1.00
Wynona	100.00	1.00
Steve	78.00	2.00
Gregory	89.00	2.00
Julian	89.00	2.00
Steve	94.00	2.00

- As illustrated in the table, for each individual an identifier and the value of the variable of interest must be entered.
 - In this case, the variable of interest is the test score.
- Additionally, a categorical variable is needed to indicate which group each individual belongs to.
 - In this example, the categorical variable is gender. The value 1 is used for females and the value 2 is used for males.
 - This is called the **grouping variable** because it shows which group, or random sample, each individual belongs to.

- In order to make inferences about the differences between the two groups in the population, the sample size, mean, and standard deviation for each group must be known.
- These statistics are automatically computed by statistical software when estimating a confidence interval or conducting a test of significance.
- The table labeled “Group Statistics” on the previous page is part of the output provided by SPSS.
- In this example, the mean for females is 10 points higher than the mean for males.
- However, if another two random samples of females and males were selected, the numbers would be slightly different.
- Therefore, instead of only providing the difference in the means, a confidence interval should be computed for this difference.
- For the example, the difference in test scores in the population between males and females will be estimated with 90% confidence.
- If the confidence interval is estimated using statistical software, the results will be slightly different than if it is done by hand. Statistical software uses a more complicated formula to compute the degrees of freedom, and tests for the assumption that the variance of the variable is equal in the two populations.
- The assumption that the variance is equal is almost never met, so the second row in the SPSS output table should be used.

Using Statistical Software:
Confidence level: 90%

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	90% Confidence Interval of the Difference	
									Lower	Upper
Test scores	Equal variances assumed	7.151	.015	2.793	18	.012	10.00000	3.58067	3.79089	16.20911
	Equal variances not assumed			3.270	15.174	.005	10.00000	3.05833	4.64264	15.35736

Interpretation: *Researchers can say with 90% confidence that difference between the females' average test score and the males' average test score ranges between 4.64 and 15.36.*

- The interval in this case is quite large due to the small sample size. The small samples were used to make computations easier for this example, but in reality confidence intervals should not be estimated for such small samples.

- To compute a confidence interval by hand, the size of the two samples and the mean and standard deviation for each group are needed.

<ul style="list-style-type: none"> • Sample 1: Females: • $n_1=8$ $\bar{x}_1=97.25$ $s_1=3.65$ 	$CI = (\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$
<ul style="list-style-type: none"> • Sample 2: Males: • $n_2=12$ $\bar{x}_2=87.25$ $s_2=9.60$ 	
<ul style="list-style-type: none"> • $df=8-1=7$ • Cl: 90% 	$t^*=1.833 \text{ (Table C)}$
	$CI = (97.25 - 87.25) \pm 1.833 * \sqrt{\frac{3.65^2}{8} + \frac{9.60^2}{12}}$ $= 10.00 \pm 5.34 = (4.65; 15.34)$

Interpretation: *It can be said with 90% confidence that difference between the females' average test score and the males' average test score ranges between 4.65 and 15.34.*

- These values are very close to the ones provided by SPSS. The small differences between these values are due to rounding and using a different procedure for computing the degrees of freedom.

Two-Sample t Test

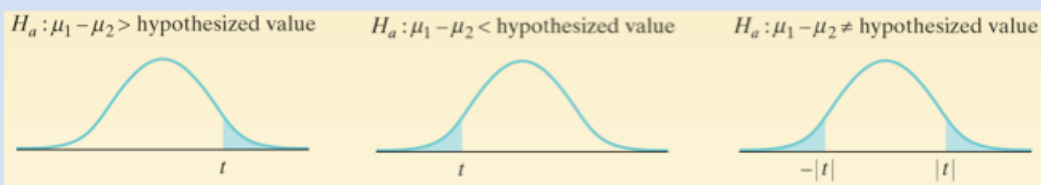
- In many research situations, it is necessary to test whether the difference between two independent groups of individuals is statistically significant.
- The null hypothesis for this test is that the groups have equal means or that there is no significant difference between the average scores of the two groups in the population.
- The alternative hypothesis can be one-sided, stating that the mean of one of the groups is higher or lower than the mean of the other group.
- If there is no information to justify a one-sided alternative hypothesis, a two-sided alternative hypothesis, which states that the two means are significantly different, could be formulated. (A one-sided alternative hypothesis looks for a significant difference only, there is no assumption that one sample has a higher or lower mean than the other sample).

Two-Sample t Test for the Difference Between Two Means

Suppose the Random, Normal, and Independent conditions are met. To test the hypothesis $H_0 : \mu_1 - \mu_2 = \text{hypothesized value}$, compute the t statistic

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

Find the P -value by calculating the probability of getting a t statistic this large or larger in the direction specified by the alternative hypothesis H_a . Use the t distribution with degrees of freedom approximated by technology or the smaller of $n_1 - 1$ and $n_2 - 1$.



- The test statistic t is a standardized difference between the means of the two samples.
- As with any other test of significance, after the test statistic has been computed, it must be determined whether this test statistic is far enough from zero to reject the null hypothesis.
- This is done either by using statistical software or by hand by using a table in a statistics textbook.

Two-Sample t Test

This example will use the same data as the previous example to test whether the difference between females' and males' average test scores is statistically significant.

- The null hypothesis is that there is no significant difference in average test scores between females and males in the population.
- There is no way of anticipating which group will have higher scores, so the alternative hypothesis is two-sided, and states that, in the population, there is a statistically significant difference in average test scores between females and males.

Using Technology:

- $H_0: \mu_1 - \mu_2 = 0$
"In the population, average test scores do not differ significantly across genders."
- $H_a: \mu_1 - \mu_2 \neq 0$ (two-sided)
"In the population, there is a significant difference between female and male average test scores."

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	90% Confidence Interval of the Difference	
									Lower	Upper
Test scores	Equal variances assumed	7.151	.015	2.793	18	.012	10.00000	3.58067	3.79089	16.20911
	Equal variances not assumed			3.270	15.174	.005	10.00000	3.05833	4.64264	15.35736

Test statistic: $t=3.270$ $p_value=.005$

$\alpha = 0.05$ (two-sided) $p_value < \alpha \Rightarrow$ reject H_0

Interpretation: Data from the sample shows that, if the null hypothesis is true, there is only a 0.5% probability of obtaining an average difference of 10 points between the two groups. This probability is smaller than the significance level of 5%, therefore the null hypothesis is rejected, and the alternative hypothesis is accepted.

- If SPSS is used to conduct the test of significance, results are provided in the "Independent Samples Test" table in the section labeled "t-test for equality of means."
- SPSS also reports the confidence interval for the difference between the two means.
- ❖ Since the p-value is .005, which is less than the alpha level of .05, the null hypothesis should be rejected.
- ❖ If the null hypothesis is true (that is, if the two population means are truly equal), there is only a 0.5% probability of obtaining the average difference that has been obtained between these two samples.
- ❖ According to these results, the alternative hypothesis should be accepted. The difference between the two population means is statistically significant.

Two-Sample t Test

- To conduct a test of significance by hand, the sample size, mean, and standard deviation of each sample are required.
- Additionally, researchers must find the critical value of t that corresponds to the degrees of freedom and the chosen level of significance.

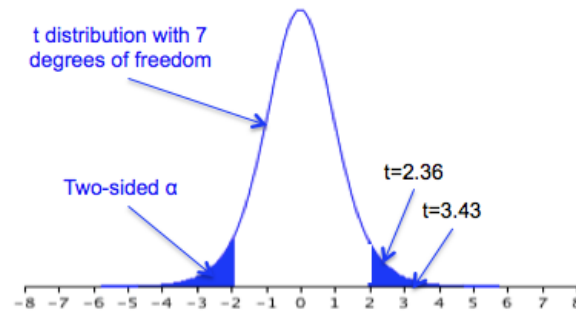
- **Sample 1: Females:**
- $n_1=8$ $\bar{x}_1=97.25$ $s_1=3.65$
- **Sample 2: Males:**
- $n_2=12$ $\bar{x}_2=87.25$ $s_2=9.60$
- $df=8-1=7$
- $\alpha=0.05$ (two-sided) } $t_{cv}=2.365$ (Table C)

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

$$t = \frac{(97.25 - 87.25)}{\sqrt{\frac{3.65^2}{8} + \frac{9.6^2}{12}}} = 3.43$$

$t > t_{cv} \Rightarrow$ reject H_0

Interpretation: Data from the sample shows that, if the null hypothesis is true, the test statistic $t_{(7)}=3.43$ is greater than the critical value $t_{cv(7)}=2.365$, when the two-tailed $\alpha=0.05$. Therefore, the null hypothesis was rejected.



- In the above example, the test statistic t is greater than the critical value. Therefore, the null hypothesis should be rejected.
- Statistical software would result in the same interpretation of the results. However, the evidence provided to support the decision to reject the null hypothesis is different.

Robustness

Two-sample t procedures are more robust than one-sample t procedures, particularly when the distributions are not symmetric.

Using the t Procedures

- Except in the case of small samples, the condition that the data are random samples from the populations of interest is more important than the condition that the population distributions are Normal.
- *Sum of the sample sizes less than 15*: Use t procedures if the data appear close to Normal. If the data are clearly skewed or if outliers are present, do not use t .
- *Sum of the sample size at least 15*: The t procedures can be used except in the presence of outliers or strong skewness.
- *Large samples*: The t procedures can be used even for clearly skewed distributions when the sum of the sample sizes is large.