10.2: Comparing Two Independent Population Means

The comparison of two independent population means is very common and provides a way to test the hypothesis that the two groups differ from each other. Is the night shift less productive than the day shift, are the rates of return from fixed asset investments different from those from common stock investments, and so on? An observed difference between two sample means depends on both the means and the sample standard deviations. Very different means can occur by chance if there is great variation among the individual samples. The test statistic will have to account for this fact. The test comparing two independent population means with unknown and possibly unequal population standard deviations is called the Aspin-Welch t-test. The degrees of freedom formula we will see later was developed by Aspin-Welch.

When we developed the hypothesis test for the mean and proportions we began with the Central Limit Theorem. We recognized that a sample mean came from a distribution of sample means, and sample proportions came from the sampling distribution of sample proportions. This made our sample parameters, the sample means and sample proportions, into random variables. It was important for us to know the distribution that these random variables came from. The Central Limit Theorem gave us the answer: the normal distribution. Our Z and t statistics came from this theorem. This provided us with the solution to our question of how to measure the probability that a sample mean came from a distribution with a particular hypothesized value of the mean or proportion. In both cases that was the question: what is the probability that the mean (or proportion) from our sample data came from a population distribution with the hypothesized value we are interested in?

Now we are interested in whether or not two samples have the same mean. Our question has not changed: Do these two samples come from the same population distribution? To approach this problem we create a new random variable. We recognize that we have two sample means, one from each set of data, and thus we have two random variables coming from two unknown distributions. To solve the problem we create a new random variable, the difference between the sample means. This new random variable also has a distribution and, again, the Central Limit Theorem tells us that this new distribution is normally distributed, regardless of the underlying distributions of the original data. A graph may help to understand this concept.

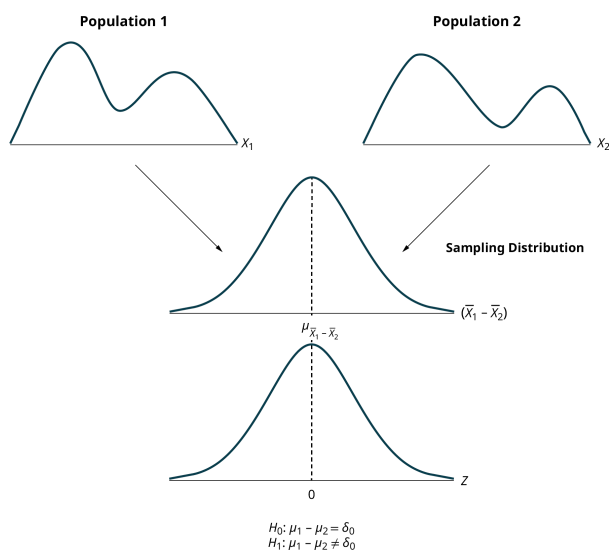


Figure 10.2.1

Pictured are two distributions of data, X_1 and X_2, with unknown means and standard deviations. The second panel shows the sampling distribution of the newly created random variable $\left(\bar{X}_1 - \bar{X}_2\right)$. This distribution is the theoretical distribution of many sample means from population 1 minus sample means from population 2 . The Central Limit Theorem tells us that this theoretical sampling distribution of differences in sample means is normally distributed, regardless of the distribution of the actual population data shown in the top panel. Because the sampling distribution is normally distributed, we can develop a standardizing formula and calculate probabilities from the standard normal distribution in the bottom panel, the Z distribution. We have seen this same analysis before in The Central Limit Theorem Figure 7.2.

The Central Limit Theorem, as before, provides us with the standard deviation of the sampling distribution, and further, that the expected value of the mean of the distribution of differences in sample means is equal to the differences in the population means. Mathematically this can be stated:

$$E\left(\mu_{\bar{x}_1}-\mu_{\bar{x}_2}\right)=\mu_1-\mu_2 \tag{10.2.1}$$

Because we do not know the population standard deviations, we estimate them using the two sample standard deviations from our independent samples. For the hypothesis test, we calculate the estimated standard deviation, or standard error, of the difference in sample means, $\bar{X}_1-\bar{X}_2$.
The standard error is:

$$\sqrt{\frac{\left(s_1\right)^2}{n_1}+\frac{\left(s_2\right)^2}{n_2}} \tag{10.2.2}$$

We remember that substituting the sample variance for the population variance when we did not have the population variance was the technique we used when building the confidence interval and the test statistic for the test of hypothesis for a single mean back in Confidence Intervals and Hypothesis Testing_with One Sample. The test statistic (t-score) is calculated as follows:

$$t_c=\frac{\left(\bar{x}_1-\bar{x}_2\right)-\delta_0}{\sqrt{\frac{\left(s_1\right)^2}{n_1}+\frac{\left(s_2\right)^2}{n_2}}} \tag{10.2.3}$$

where:

- s_1 and s_2, the sample standard deviations, are estimates of σ_1 and σ_2, respectively and
- σ_1 and σ_2 are the unknown population standard deviations.
- $\bar{x}_1$ and $\bar{x}_2$ are the sample means. μ_1 and μ_2 are the unknown population means.

$\delta=$ the hypothesized differences between the two populations means.

Substituting the sample variances of the two sample groups when the populations variances are unknown and the sample size is large, the t-distribution will provide good estimates using degrees of freedom:

$$v=n_1+n_2-2. \tag{10.2.4}$$

Large in this case would be greater than 100 observations from the combined two groups (n_1+n_2) although other statisticians suggest that the definition of large is greater than 30 observations for each of the two groups.

The number of degrees of freedom (df) requires a somewhat complicated calculation. The df are not always a whole number. The test statistic above is approximated by the Student's t-distribution with df as follows:
Degrees of Freedom

$$\text{The standard error is: } df=\frac{\left(\frac{\left(s_1\right)^2}{n_1}+\frac{\left(s_2\right)^2}{n_2}\right)^2}{\left(\frac{1}{n_1 1}\right)\left(\frac{\left(s_1\right)^2}{n_1}\right)^2+\left(\frac{1}{n_2-1}\right)\left(\frac{\left(s_2\right)^2}{n_2}\right)^2} \tag{10.2.5}$$

When both sample sizes n_1 and n_2 are 30 or larger, the Student's t approximation is very good. If each sample has more than 30 observations then the degrees of freedom can be calculated as n_1+n_2-2 .

The format of the sampling distribution, differences in sample means, specifies that the format of the null and alternative hypothesis is:
The test hypothesis for differences between two means

$$\begin{aligned} H_0: \mu_1-\mu_2=\delta_0 \\ H_{\mathrm{a}}: \mu_1-\mu_2 \neq \delta_0 \end{aligned} \tag{10.2.6}$$

where δ_0 is the hypothesized difference between the two means. If the question is simply "is there any difference between the means?" then $\delta_0=0$ and the null and alternative hypotheses becomes:

$$\begin{aligned} H_0: \mu_1=\mu_2 \\ H_{\mathrm{a}}: \mu_1 \neq \mu_2 \end{aligned} \tag{10.2.7}$$

An example of when δ_0 might not be zero is when the comparison of the two groups requires a specific difference for the decision to be meaningful. Imagine that you are making a capital investment. You are considering changing from your current model machine to another. You measure the productivity of your machines by the speed they produce the product. It may be that a contender to replace the old model is faster in terms of product throughput, but is also more expensive. The second machine may also have more maintenance costs, setup costs, etc. The null hypothesis would be set up so that the new machine would have to be better than the old one by enough to cover these extra costs in terms of speed and cost of production. This form of the null and alternative hypothesis shows how valuable this particular hypothesis test can be. For most of our work we will be testing simple hypotheses asking if there is any difference between the two distribution means.

Exercise 10.2.1

Independent groups

The Kona Iki Corporation produces coconut milk. They take coconuts and extract the milk inside by drilling a hole and pouring the milk into a vat for processing. They have both a day shift (called the B shift) and a night shift (called the G shift) to do this part of the process. They would like to know if the day shift and the night shift are equally efficient in processing the coconuts. A study is done sampling 9 shifts of the G shift and 16 shifts of the B shift. The results of the number of hours required to process 100 pounds of coconuts is presented in Table 10.1. A study is done and data are collected, resulting in the data in Table 10.1.

	Sample Size	Average Number of Hours to Process 100 Pounds of Coconuts	Sample Standard Deviation
G Shift	9	2	0.866
B Shift	16	3.2	1.00

Problem

Is there a difference in the mean amount of time for each shift to process 100 pounds of coconuts? Test at the 5% level of significance.

Answer

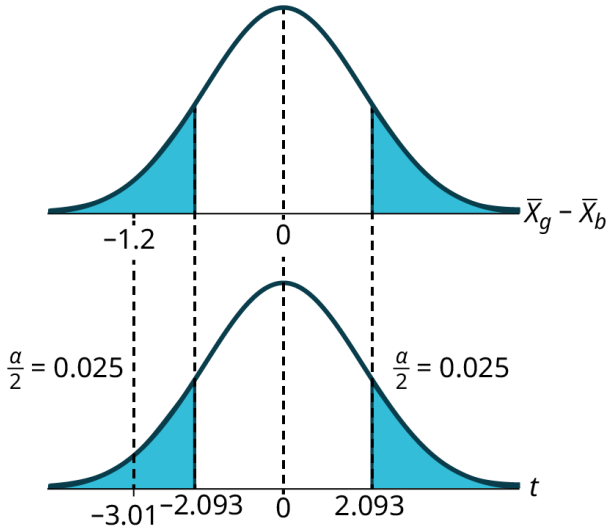
The population standard deviations are not known and cannot be assumed to equal each other.
Let g be the subscript for the G Shift and b be the subscript for the B Shift. Then, μ_g is the population mean for G Shift and μ_b is the population mean for B Shift. This is a test of two independent groups, two population means.

Random variable: $\bar{X}_g - \bar{X}_b =$ difference in the sample mean amount of time between the G Shift and the B Shift takes to process the coconuts.

$$\begin{aligned} H_0 : \mu_g = \mu_b \quad & H_0 : \mu_g - \mu_b = 0 \\ H_a : \mu_g \neq \mu_b \quad & H_a : \mu_g - \mu_b \neq 0 \end{aligned} \tag{10.2.8}$$

The words "**the same**" tell you H_0 has an " $=$ ". Since there are no other words to indicate H_a, is either faster or slower. This is a two tailed test.

Distribution for the test: Use t_{df} where df is calculated using the df formula for independent groups, two population means above. Using a calculator, df is approximately 18.8462.



$$H_0: \mu_g - \mu_b = 0 \quad \text{or} \quad H_0: \mu_g = \mu_b$$
$$H_a: \mu_g - \mu_b \neq 0 \qquad H_a: \mu_g \neq \mu_b$$

Figure 10.2.2

We next find the critical value on the t-table using the degrees of freedom from above. The critical value, 2.093, is found in the .025 column, this is α/2, at 19 degrees of freedom. (The convention is to round up the degrees of freedom to make the conclusion more conservative.) Next we calculate the test statistic and mark this on the t-distribution graph.

Make a decision: Since the calculated t-value is in the tail we cannot accept the null hypothesis that there is no difference between the two groups. The means are different.

The graph has included the sampling distribution of the differences in the sample means to show how the t-distribution aligns with the sampling distribution data. We see in the top panel that the calculated difference in the two means is -1.2 and the bottom panel shows that this is 3.01 standard deviations from the mean. Typically we do not need to show the sampling distribution graph and can rely on the graph of the test statistic, the t-distribution in this case, to reach our conclusion.

Conclusion: At the 5% level of significance, the sample data show there is sufficient evidence to conclude that the mean number of hours that the G Shift takes to process 100 pounds of coconuts is different from the B Shift (mean number of hours for the B Shift is greater than the mean number of hours for the G Shift).

Try It 10.2.1

Two samples are shown in the table below. Both have normal distributions. The means for the two populations are thought to be the same. Is there a difference in the means? Test at the 5% level of significance.

	Sample Size	Sample Mean	Sample Standard Deviation
Population A	25	5	1
Population B	16	4.7	1.2

Note

When the sum of the sample sizes is larger than $30\,(n_1 + n_2 > 30)$ you can use the normal distribution to approximate the Student's t. Use the Student's t distribution, however, whenever possible.

Exercise 10.2.2

A study is done to determine if Company A retains its workers longer than Company B. It is believed that Company A has a higher retention than Company B. The study finds that in a sample of 11 workers at Company A their average time with the company is four years with a standard deviation of 1.5 years. A sample of 9 workers at Company B finds that the average time with the company was 3.5 years with a standard deviation of 1 year. Test this proposition at the 1% level of significance.

Problem

a. Is this a test of two means or two proportions?

▼

b. Are the populations standard deviations known or unknown?

▼

c. Which distribution do you use to perform the test?

▼

d. What is the random variable?

e. What are the null and alternate hypotheses?

f. Is this test right-, left-, or two-tailed?

g. What is the value of the test statistic?

h. Can you accept/reject the null hypothesis?

i. **Conclusion:**

Answer

a. two means because time is a continuous random variable.

b. unknown

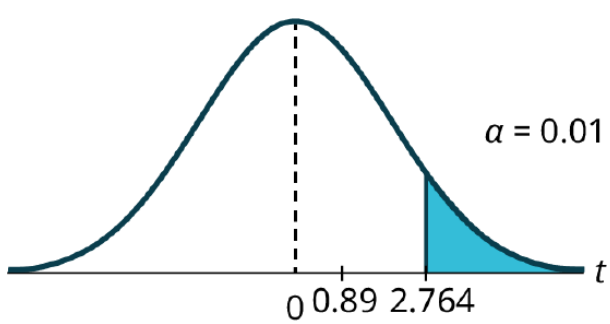
c. Student's *t*

d. $\bar{X}_A - \bar{X}_B$

e.

- $H_\varnothing : \mu_A \leq \mu_B$
- $H_a : \mu_A > \mu_B$

f. right one-tailed test



$$H_0: \mu_A \leq \mu_B$$

$$H_a: \mu_A > \mu_B$$

Figure 10.2.3: Copy and Paste Caption here. (Copyright; author via source)

g. $t_c = \dfrac{(\bar{X}_1 - \bar{X}_2) - \delta_0}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}} = 0.89$

h. Cannot reject the null hypothesis that there is no difference between the two groups. Test statistic is not in the tail. The critical value of the *t*-distribution is 2.764 with 10 degrees of freedom. This example shows how difficult it is to reject a null hypothesis with a very small sample. The critical values require very large test statistics to reach the tail.

i. At the 1% level of significance, from the sample data, there is not sufficient evidence to conclude that the retention of workers at Company A is longer than Company B, on average.

Try It 10.2.2

A study is done to determine if Company A retains its workers longer than Company B. Company A samples 15 workers, and their average time with the company is five years with a standard deviation of 1.2. Company B samples 20 workers, and their average time with the company is 4.5 years with a standard deviation of 0.8. The populations are normally distributed.

a. Are the population standard deviations known?
b. Conduct an appropriate hypothesis test. At the 5% significance level, what is your conclusion?

Exercise 10.2.3

An interesting research question is the effect, if any, that different types of teaching formats have on the grade outcomes of students. To investigate this issue one sample of students' grades was taken from a hybrid class and another sample taken from a standard lecture format class. Both classes were for the same subject. The mean course grade in percent for the 35 hybrid students is 74 with a standard deviation of 16. The mean grades of the 40 students form the standard lecture class was 76 percent with a standard deviation of 9. Test at 5% to see if there is any significant difference in the population mean grades between standard lecture course and hybrid class.

Answer

We begin by noting that we have two groups, students from a hybrid class and students from a standard lecture format class. We also note that the random variable, what we are interested in, is students' grades, a continuous random variable. We could have asked the research question in a different way and had a binary random variable. For example, we could have studied the percentage of students with a failing grade, or with an A grade. Both of these would be binary and thus a test of proportions and not a test of means as is the case here. Finally, there is no presumption as to which format might lead to higher grades so the hypothesis is stated as a two-tailed test.

$$\begin{aligned} \mathrm{H}_0 &: \mu_1 = \mu_2 \\ \mathrm{H}_a &: \mu_1 \neq \mu_2 \end{aligned} \tag{10.2.9}$$

As would virtually always be the case, we do not know the population variances of the two distributions and thus our test statistic is:

$$t_c = \frac{(\bar{x}_1 - \bar{x}_2) - \delta_0}{\sqrt{\dfrac{s^2}{n_1} + \dfrac{s^2}{n_2}}} = \frac{(74 - 76) - 0}{\sqrt{\dfrac{16^2}{35} + \dfrac{9^2}{40}}} = -0.65 \tag{10.2.10}$$

To determine the critical value of the Student's t we need the degrees of freedom. For this case we use: $\mathrm{df} = \mathrm{n1} + \mathrm{n2} - 2 = 35 + 40 - 2 = 73$. This is large enough to consider it the normal distribution thus $t_{a/2} = 1.96$. Again as always we determine if the calculated value is in the tail determined by the critical value. In this case we do not even need to look up the critical value: the calculated value of the difference in these two average grades is not even one standard deviation apart. Certainly not in the tail.

Conclusion: Cannot reject the null at $\alpha = 5\%$. Therefore, evidence does not exist to prove that the grades in hybrid and standard classes differ.

Try It 10.2.2

Two professors, A and B, teach classes on the same subjects. Scores of 10 students from each class are selected randomly. The final exam scores of the students are as follows:

Professor A:

97	62	73	58	84
74	66	93	73	85

Professor B:

85	64	74	55	76
67	72	84	71	98

Professor A says that the mean score of their class is more than the mean of professor B's class. Is professor A correct? Test at a 5% significance level. Answer the following questions:

a. Is this a test of two means or two proportions?
b. Are the population standard deviations known or unknown?
c. Which distribution do you use to perform the test?
d. What is the random variable?
e. What are the null and alternative hypotheses? Write the null and alternative hypotheses in words and in symbols.
f. Is this test right-, left-, or two-tailed?
g. What is the *p*-value?
h. Do you reject or not reject the null hypothesis?
i. At the ___ level of significance, from the sample data, there ___ (is/is not) sufficient evidence to conclude that ___.

This page titled 10.2: Comparing Two Independent Population Means is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.