

10.2: Two Population Means with Unknown Standard Deviations

1. The two independent samples are simple random samples from two distinct populations.
2. For the two distinct populations:
 - if the sample sizes are small, the distributions are important (should be normal)
 - if the sample sizes are large, the distributions are not important (need not be normal)

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 was developed by Aspin-Welch.

The comparison of two population means is very common. A difference between the two samples depends on both the means and the standard deviations. Very different means can occur by chance if there is great variation among the individual samples. In order to account for the variation, we take the difference of the sample means, $\bar{X}_1 - \bar{X}_2$, and divide by the standard error in order to standardize the difference. The result is a t -score test statistic.

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:

$$\sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}} \quad (10.2.1)$$

The test statistic (t -score) is calculated as follows:

$$\frac{(\bar{x} - \bar{x}) - (\mu_1 - \mu_2)}{\sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}}} \quad (10.2.2)$$

where:

- s_1 and s_2 , the sample standard deviations, are estimates of σ_1 and σ_2 , respectively.
- 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 population means.

The number of *degrees of freedom* (df) requires a somewhat complicated calculation. However, a computer or calculator calculates it easily. The df are not always a whole number. The test statistic calculated previously is approximated by the Student's t -distribution with df as follows:

Degrees of freedom

$$df = \frac{\left(\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2} \right)^2}{\left(\frac{1}{n_1 - 1} \right) \left(\frac{(s_1)^2}{n_1} \right)^2 + \left(\frac{1}{n_2 - 1} \right) \left(\frac{(s_2)^2}{n_2} \right)^2} \quad (10.2.3)$$

When both sample sizes n_1 and n_2 are five or larger, the Student's t approximation is very good. Notice that the sample variances $(s_1)^2$ and $(s_2)^2$ are not pooled. (If the question comes up, do not pool the variances.)

It is not necessary to compute the degrees of freedom by hand. A calculator or computer easily computes it.

✓ Example 10.2.1: Independent groups

The average amount of time boys and girls aged seven to 11 spend playing sports each day is believed to be the same. A study is done and data are collected, resulting in the data in Table 10.2.1. Each populations has a normal distribution.

Table 10.2.1

	Sample Size	Average Number of Hours Playing Sports Per Day	Sample Standard Deviation
Girls	9	2	0.8660.866
Boys	16	3.2	1.00

Is there a difference in the mean amount of time boys and girls aged seven to 11 play sports each day? Test at the 5% level of significance.

Answer

The population standard deviations are not known. Let g be the subscript for girls and b be the subscript for boys. Then, μ_g is the population mean for girls and μ_b is the population mean for boys. 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 girls and boys play sports each day.

- $H_0 : \mu_g = \mu_b$
- $H_0 : \mu_g - \mu_b = 0$
- $H_a : \mu_g \neq \mu_b$
- $H_a : \mu_g - \mu_b \neq 0$

The words "**the same**" tell you H_0 has an "=". Since there are no other words to indicate H_a , assume it says "**is different.**" 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. Using a calculator, df is approximately 18.8462. **Do not pool the variances.**

Calculate the p -value using a Student's t -distribution: $p\text{-value} = 0.0054$

Graph:

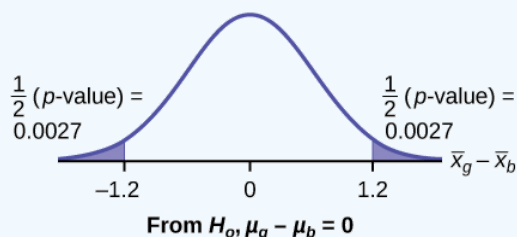


Figure 10.2.1: Normal distribution curve representing the difference in the average amount of time girls and boys play sports all day

$$s_g = 0.866 \quad (10.2.4)$$

$$s_b = 1 \quad (10.2.5)$$

So,

$$\bar{x}_g - \bar{x}_b = 2 - 3.2 = -1.2 \quad (10.2.6)$$

Half the p -value is below -1.2 and half is above 1.2 .

Make a decision: Since $\alpha > p\text{-value}$, reject H_0 . This means you reject $\mu_g = \mu_b$. The means are different.

Press **STAT**. Arrow over to **TESTS** and press **4:2-SampTTest**. Arrow over to **Stats** and press **ENTER**. Arrow down and enter **2** for the first sample mean, **√0.866** for $Sx1$, **9** for $n1$, **3.2** for the second sample mean, **1** for $Sx2$,

and 16 for n_2 . Arrow down to μ_1 : and arrow to **does not equal** μ_2 . Press **ENTER**. Arrow down to Pooled: and **No**. Press **ENTER**. Arrow down to **Calculate** and press **ENTER**. The p -value is $p = 0.0054$, the dfs are approximately 18.8462, and the test statistic is -3.14. Do the procedure again but instead of Calculate do Draw.

Conclusion: At the 5% level of significance, the sample data show there is sufficient evidence to conclude that the mean number of hours that girls and boys aged seven to 11 play sports per day is different (mean number of hours boys aged seven to 11 play sports per day is greater than the mean number of hours played by girls OR the mean number of hours girls aged seven to 11 play sports per day is greater than the mean number of hours played by boys).

? Exercise 10.2.1

Two samples are shown in Table. 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.

Table 10.2.2

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

Answer

The p -value is 0.4125, which is much higher than 0.05, so we decline to reject the null hypothesis. There is not sufficient evidence to conclude that the means of the two populations are not the same.

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 .

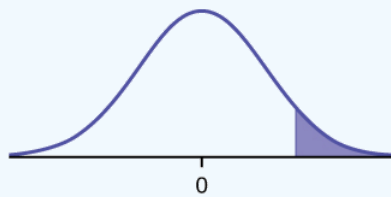
✓ Example 10.2.2

A study is done by a community group in two neighboring colleges to determine which one graduates students with more math classes. College A samples 11 graduates. Their average is four math classes with a standard deviation of 1.5 math classes. College B samples nine graduates. Their average is 3.5 math classes with a standard deviation of one math class. The community group believes that a student who graduates from college A **has taken more math classes**, on the average. Both populations have a normal distribution. Test at a 1% significance level. Answer the following questions.

- Is this a test of two means or two proportions?
- Are the populations standard deviations known or unknown?
- Which distribution do you use to perform the test?
- What is the random variable?
- What are the null and alternate hypotheses? Write the null and alternate hypotheses in words and in symbols.
- Is this test right-, left-, or two-tailed?
- What is the p -value?
- Do you reject or not reject the null hypothesis?

Solutions

- two means
- unknown
- Student's t
- $\bar{X}_A - \bar{X}_B$
- $H_0 : \mu_A \leq \mu_B$ and $H_a : \mu_A > \mu_B$



$$\bar{x}_A - \bar{x}_B = 0.5^*$$

Note: $\bar{x}_A - \bar{x}_B = 4 - 3.5 = 0.5$

f.

right

g. g. 0.1928

h. h. Do not reject.

i. i. At the 1% level of significance, from the sample data, there is not sufficient evidence to conclude that a student who graduates from college A has taken more math classes, on the average, than a student who graduates from college B.

? Exercise 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.

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

Answer

- They are unknown.
- The p -value = 0.0878. At the 5% level of significance, there is insufficient evidence to conclude that the workers of Company A stay longer with the company.

✓ Example 10.2.3

A professor at a large community college wanted to determine whether there is a difference in the means of final exam scores between students who took his statistics course online and the students who took his face-to-face statistics class. He believed that the mean of the final exam scores for the online class would be lower than that of the face-to-face class. Was the professor correct? The randomly selected 30 final exam scores from each group are listed in Table 10.2.3 and Table 10.2.4

Table 10.2.3: Online Class

67.6	41.2	85.3	55.9	82.4	91.2	73.5	94.1	64.7	64.7
70.6	38.2	61.8	88.2	70.6	58.8	91.2	73.5	82.4	35.5
94.1	88.2	64.7	55.9	88.2	97.1	85.3	61.8	79.4	79.4

Table 10.2.4: Face-to-face Class

77.9	95.3	81.2	74.1	98.8	88.2	85.9	92.9	87.1	88.2
69.4	57.6	69.4	67.1	97.6	85.9	88.2	91.8	78.8	71.8
98.8	61.2	92.9	90.6	97.6	100	95.3	83.5	92.9	89.4

Is the mean of the Final Exam scores of the online class lower than the mean of the Final Exam scores of the face-to-face class? Test at a 5% significance level. Answer the following questions:

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

(See the conclusion in Example, and write yours in a similar fashion)

Be careful not to mix up the information for Group 1 and Group 2!

Answer

- a. two means
- b. unknown
- c. Student's t
- d. $\bar{X}_1 - \bar{X}_2$
- e.
 - i. $H_0 : \mu_1 = \mu_2$ Null hypothesis: the means of the final exam scores are equal for the online and face-to-face statistics classes.
 - ii. $H_a : \mu_1 < \mu_2$ Alternative hypothesis: the mean of the final exam scores of the online class is less than the mean of the final exam scores of the face-to-face class.
- f. left-tailed
- g. p -value = 0.0011

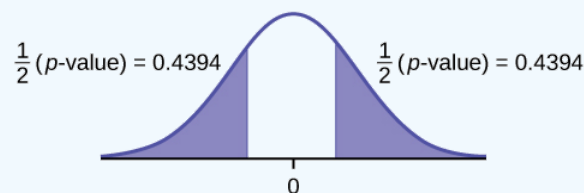


Figure 10.2.3.

- h. Reject the null hypothesis
- i. The professor was correct. The evidence shows that the mean of the final exam scores for the online class is lower than that of the face-to-face class.

At the 5% level of significance, from the sample data, there is (is/is not) sufficient evidence to conclude that the mean of the final exam scores for the online class is less than the mean of final exam scores of the face-to-face class.

First put the data for each group into two lists (such as L1 and L2). Press STAT. Arrow over to TESTS and press 4:2SampTTest. Make sure Data is highlighted and press ENTER. Arrow down and enter L1 for the first list and L2 for the second list. Arrow down to μ_1 : and arrow to $\neq \mu_1$ (does not equal). Press ENTER. Arrow down to Pooled: No. Press ENTER. Arrow down to Calculate and press ENTER.

Cohen's Standards for Small, Medium, and Large Effect Sizes

Cohen's d is a measure of effect size based on the differences between two means. Cohen's d , named for United States statistician Jacob Cohen, measures the relative strength of the differences between the means of two populations based on sample data. The calculated value of effect size is then compared to Cohen's standards of small, medium, and large effect sizes.

Table 10.2.5: Cohen's Standard Effect Sizes

Size of effect	d
Small	0.2
medium	0.5

Size of effect	d
Large	0.8

Cohen's d is the measure of the difference between two means divided by the pooled standard deviation: $d = \frac{\bar{x}_2 - \bar{x}_1}{s_{\text{pooled}}}$ where

$$s_{\text{pooled}} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

✓ Example 10.2.4

Calculate Cohen's d for Example. Is the size of the effect small, medium, or large? Explain what the size of the effect means for this problem.

Answer

$$\mu_1 = 4s_1 = 1.5n_1 = 11$$

$$\mu_2 = 3.5s_2 = 1n_2 = 9$$

$$d = 0.384$$

The effect is small because 0.384 is between Cohen's value of 0.2 for small effect size and 0.5 for medium effect size. The size of the differences of the means for the two colleges is small indicating that there is not a significant difference between them.

✓ Example 10.2.5

Calculate Cohen's d for Example. Is the size of the effect small, medium or large? Explain what the size of the effect means for this problem.

Answer

$d = 0.834$; Large, because 0.834 is greater than Cohen's 0.8 for a large effect size. The size of the differences between the means of the Final Exam scores of online students and students in a face-to-face class is large indicating a significant difference.

✓ Example 10.2.6

Weighted alpha is a measure of risk-adjusted performance of stocks over a period of a year. A high positive weighted alpha signifies a stock whose price has risen while a small positive weighted alpha indicates an unchanged stock price during the time period. Weighted alpha is used to identify companies with strong upward or downward trends. The weighted alpha for the top 30 stocks of banks in the northeast and in the west as identified by Nasdaq on May 24, 2013 are listed in Table and Table, respectively.

Northeast

94.2	75.2	69.6	52.0	48.0	41.9	36.4	33.4	31.5	27.6
77.3	71.9	67.5	50.6	46.2	38.4	35.2	33.0	28.7	26.5
76.3	71.7	56.3	48.7	43.2	37.6	33.7	31.8	28.5	26.0

West

126.0	70.6	65.2	51.4	45.5	37.0	33.0	29.6	23.7	22.6
116.1	70.6	58.2	51.2	43.2	36.0	31.4	28.7	23.5	21.6
78.2	68.2	55.6	50.3	39.0	34.1	31.0	25.3	23.4	21.5

Is there a difference in the weighted alpha of the top 30 stocks of banks in the northeast and in the west? Test at a 5% significance level. Answer the following questions:

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

Answer

- two means
- unknown
- Student's- t
- $\bar{X}_1 - \bar{X}_2$
- $H_0 : \mu_1 = \mu_2$ Null hypothesis: the means of the weighted alphas are equal.
 - $H_a : \mu_1 \neq \mu_2$ Alternative hypothesis : the means of the weighted alphas are not equal.
- two-tailed
- p -value = 0.8787
- Do not reject the null hypothesis
- This indicates that the trends in stocks are about the same in the top 30 banks in each region.


 This is a normal distribution curve with mean equal to zero. Both the right and left tails of the curve are shaded. Each tail represents $1/2(p\text{-value}) = 0.4394$.

Figure 10.2.4.

5% level of significance, from the sample data, there is not sufficient evidence to conclude that the mean weighted alphas for the banks in the northeast and the west are different

- $d = 0.040$, Very small, because 0.040 is less than Cohen's value of 0.2 for small effect size. The size of the difference of the means of the weighted alphas for the two regions of banks is small indicating that there is not a significant difference between their trends in stocks.

References

- Data from Graduating Engineer + Computer Careers. Available online at www.graduatingengineer.com
- Data from *Microsoft Bookshelf*.
- Data from the United States Senate website, available online at www.Senate.gov (accessed June 17, 2013).
- "List of current United States Senators by Age." Wikipedia. Available online at en.Wikipedia.org/wiki/List_of...enators_by_age (accessed June 17, 2013).
- "Sectoring by Industry Groups." Nasdaq. Available online at www.nasdaq.com/markets/barcha...&base=industry (accessed June 17, 2013).
- "Strip Clubs: Where Prostitution and Trafficking Happen." Prostitution Research and Education, 2013. Available online at www.prostitutionresearch.com/ProsViolPosttrauStress.html (accessed June 17, 2013).
- "World Series History." Baseball-Almanac, 2013. Available online at <http://www.baseball-almanac.com/ws/wsmenu.shtml> (accessed June 17, 2013).

Review

Two population means from independent samples where the population standard deviations are not known

- Random Variable: $\bar{X}_1 - \bar{X}_2 =$ the difference of the sampling means
- Distribution: Student's t -distribution with degrees of freedom (variances not pooled)

Formula Review

Standard error:

$$SE = \sqrt{\frac{(s_1^2)}{n_1} + \frac{(s_2^2)}{n_2}} \quad (10.2.7)$$

Test statistic (*t*-score):

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}}} \quad (10.2.8)$$

Degrees of freedom:

$$df = \frac{\left(\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}\right)^2}{\left(\frac{1}{n_1 - 1}\right)\left(\frac{(s_1)^2}{n_1}\right)^2} + \left(\frac{1}{n_2 - 1}\right)\left(\frac{(s_2)^2}{n_2}\right)^2} \quad (10.2.9)$$

where:

- s_1 and s_2 are the sample standard deviations, and n_1 and n_2 are the sample sizes.
- \bar{x}_1 and \bar{x}_2 are the sample means.

Cohen's d is the measure of effect size:

$$d = \frac{\bar{x}_1 - \bar{x}_2}{s_{\text{pooled}}} \quad (10.2.10)$$

where

$$s_{\text{pooled}} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \quad (10.2.11)$$

Glossary

Degrees of Freedom (df)

the number of objects in a sample that are free to vary.

Standard Deviation

A number that is equal to the square root of the variance and measures how far data values are from their mean; notation: s for sample standard deviation and σ for population standard deviation.

Variable (Random Variable)

a characteristic of interest in a population being studied. Common notation for variables are upper-case Latin letters X, Y, Z, \dots . Common notation for a specific value from the domain (set of all possible values of a variable) are lower-case Latin letters x, y, z, \dots . For example, if X is the number of children in a family, then x represents a specific integer 0, 1, 2, 3, Variables in statistics differ from variables in intermediate algebra in the two following ways.

- The domain of the random variable (RV) is not necessarily a numerical set; the domain may be expressed in words; for example, if X = hair color, then the domain is {black, blond, gray, green, orange}.
- We can tell what specific value x of the random variable X takes only after performing the experiment.

This page titled [10.2: Two Population Means with Unknown Standard Deviations](#) 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.

- [10.2: Two Population Means with Unknown Standard Deviations](#) by [OpenStax](#) is licensed [CC BY 4.0](#). Original source: <https://openstax.org/details/books/introductory-statistics>.