Minitab

ANSWER KEY: TWO-SAMPLE T-TESTS FOR POPULATION MEANS
This answer key provides solutions to the corresponding activity sheet.

Two-Sample *t*-Test for Population Means

The data for these exercises is in the Minitab file ***TwoSampleTTest_PopMeans_Activity.mtw***.

Exercise 1: Noise Level of Planes

In an environmental impact study for a new airport, the noise level of various jets was measured just seconds after their wheels left the ground. The jets were either wide-bodied or narrow-bodied. The noise levels in decibels (dB) are recorded here for 15 wide-bodied and 12 narrow-bodied jets.*

Wide	109.5	107.3	105.0	117.3	105.4	113.7	121.7	109.2	108.1	106.4	104.6	110.5	110.9	111.0	112.4
Narrow	131.4	126.8	114.1	126.9	108.2	122.0	106.9	116.3	115.5	111.6	124.5	116.2			

* An Introduction to Statistical Methods and Data Analysis
By R. Lyman Ott, Micheal T. Longnecker

(a) Is the data for the wide-bodied and narrow-bodied jets normally distributed so that we can perform a 2-sample *t*-test comparing their noise levels? How are you checking this condition?

Solution: You can perform a normality test on the data sets separately or together. The null and alternative hypotheses for a normality test are:

H_0: The data is from a normally distributed population
H_a: The data is not from a normally distributed population

The *p*-values for the Anderson-Darling normality tests for the wide and narrow-bodied planes are 0.283 and 0.593, respectively. These values suggest that there is NOT enough evidence to reject the null hypothesis, which is that both data sets are normally distributed. So, yes, there is support for both data sets being from normally distributed populations. We'd like larger *p*-values, but the 0.283 isn't too bad!

(b) Do the two types of jets have different mean noise levels? Set up the null and alternative hypotheses to exam this question. Clearly define the population parameters.

Solution: We're testing H_0: $\mu_{wide} - \mu_{narrow} = 0$ versus H_a: $\mu_{wide} - \mu_{narrow} \neq 0$, which may also be written as H_0: $\mu_{wide} = \mu_{narrow}$ versus H_a: $\mu_{wide} \neq \mu_{narrow}$.

(c) Determine the standardized test statistic and the *p*-value associated with this test statistic.

Solution: The appropriate test is a 2-sample *t*-test. We don't know the population variances, and we'll assume that they are NOT equal. We can determine the test statistic using Minitab's **Stat > Basic Statistics > 2-sample t.**

Method

μ_1: population mean of Wide-Bodied Plane Noise
μ_2: population mean of Narrow-Bodied Plane Noise
Difference: μ_1 - μ_2

*Equal variances are **not** assumed for this analysis.*

Descriptive Statistics

Sample	N	Mean	StDev	SE Mean
Wide-Bodied Plane Noise	15	110.20	4.71	1.2
Narrow-Bodied Plane Noise	12	118.37	7.87	2.3

Estimation for Difference

Difference	95% CI for Difference
-8.17	(-13.60, -2.73)

Test

Null hypothesis	H_0: $\mu_1 - \mu_2 = 0$
Alternative hypothesis	H_1: $\mu_1 - \mu_2 \neq 0$

T-Value	DF	P-Value
-3.17	17	0.006

The standardized test statistic is t_0 = **-3.17** and the associated *p*-value is **0.006**.

If the null and alternative are written as: **H_0: $\mu_{narrow} - \mu_{wide} = 0$ versus H_a: $\mu_{narrow} - \mu_{wide} \neq 0$**, then the test statistic is t_0 = +3.17.

(d) State the conclusion with respect to the research question.

Using the provided sample data, the *p*-value for the hypothesis test is ~0.006. At $\alpha = 0.05$, we reject H_0, suggesting that the mean noise levels from the two different types of planes are not the same.

Exercise 2: Amount Spent on Gifts by Gender

Independent random samples of 23 females and 23 males (no connections to each other) were asked how much money they spent on Valentine's Day gifts this year. Assume the amount spent for each gender is normally distributed and the population variances for the amounts spent are unknown and not equal. A hypothesis test was performed to determine whether the true mean amount that females spent on gifts is equal to the true mean amount that males spent on gifts. The following Minitab output was obtained, but the *p*-value and degrees of freedom did not print.

Sample	N	Mean	StDev	SE Mean
Females	23	25.00	2.00	0.42
Males	23	22.00	3.00	0.63

Test

Null hypothesis	H_0: $\mu_F - \mu_M = 0$
Alternative hypothesis	H_1: $\mu_F - \mu_M \neq 0$

T-Value	DF	P-Value
3.99	xx	x.xxx

What are the degrees of freedom for this hypothesis test, and what is your decision regarding the H_0?

A. We don't need to compute the degrees of freedom because we would use a 2-sample *z*-test to perform this hypothesis test. In performing a 2-sample *z*-test, our decision would be to reject H_0.
B. Without the actual data, we can't determine the degrees of freedom or make a decision about H_0.
C. The degrees of freedom is 22. We would reject H_0.
D. The degrees of freedom is 23. We would reject H_0.
E. The degrees of freedom is 22. We would not reject H_0.
F. The degrees of freedom is 38. We would reject H_0.
G. The degrees of freedom is 44. We would reject H_0.
H. The degrees of freedom is 46. We would not reject H_0.

Solution: F. I'd suggest performing this 2-sample *t*-test in Minitab with the sample statistics provided, and the degrees of freedom and *p*-value will be computed.

Exercise 3: Linus's Twitter Followers

Linus is seldom seen without his trusty blue blanket. But he is concerned that it is affecting his social life. Linus decides to conduct a study to determine if children who carry blankets daily have less Twitter followers, on average, than children who do not carry a blanket daily. He obtains a random sample of 20 children who carry a blanket daily and records the number of Twitter followers each has; then, he obtains a random sample of 20 children who do not carry a blanket daily and records the number of Twitter followers each has. The data is provided in Minitab worksheet for this activity document. Note: The order in which the data was collected is unknown.

(a) State the null and alternative hypothesis that addresses the question of interest. Be sure to define the parameters of interest.

Solution: Let μ_B be the average number of Twitter followers that a child who carries a blanket daily has and let μ_{NB} be the average number of Twitter followers that a child who does not carry a blanket daily has. Then, we are interested in:

H_0: $\mu_B - \mu_{NB} = 0$ vs. H_a: $\mu_B - \mu_{NB} < 0$

(b) Indicate (by circling the choice below) whether you feel the data is paired or independent. Then, compute the corresponding test statistic for assessing the hypothesis stated in part (a).

Paired Data Two Independent Samples

Solution: The data comes from **two independent samples**. Therefore, the appropriate hypothesis test to perform is a 2-sample *t*-test. The standardized test statistic is t_0 = -3.53.

Method

μ_B: population mean of Blanket: Num Followers

μ_{NB}: population mean of No Blanket: Num Followers

Difference: $\mu_B - \mu_{NB}$

Equal variances are not assumed for this analysis.

Descriptive Statistics

Sample	N	Mean	StDev	SE Mean
Blanket: Num Followers	20	40.1	19.5	4.4
No Blanket: Num Followers	20	64.6	24.2	5.4

Test

Null hypothesis	H_0: $\mu_B - \mu_{NB} = 0$
Alternative hypothesis	H_1: $\mu_B - \mu_{NB} \neq 0$

T-Value	DF	P-Value
-3.53	36	0.001

We can also compute the test statistics by-hand as follows:

$$t_0 = \frac{(40.1 - 64.6) - 0}{\sqrt{\frac{19.5^2}{20} + \frac{24.2^2}{20}}} \cong \mathbf{-3.525}$$

Note: If a student incorrectly chose paired data, the test statistic is -3.11.
Note: Also, the sign of the test statistic may be flipped depending on the direction of the hypothesis defined in (a).

(c) Woodstock (a big fan of Twitter's mascot) looks over the data Linus has collected and states that "the test statistic in part (b) can be modeled using a standard normal distribution since we have a total of 40 observations randomly collected." Why is this statement invalid?

Solution: In order to apply the Central Limit Theorem, we need both samples to be large; it is not enough to have the total sample size be large.

(d) Assuming all assumptions for modeling the test statistic in part (b) using a *t* distribution are reasonable, Lucy computes a *p*-value that is less than her significance level of $\alpha = 0.05$; she concludes that "his data suggests that if children stopped carrying blankets, then their number of Twitter followers would increase, on average." Is her conclusion justified? Explain.

Solution: No. In order to conclude cause-and-effect, we would need a controlled experiment. Since whether a subject carried a blanket daily was not randomly assigned, we cannot conclude a cause-and-effect relationship.

Exercise 4: Height Data

In class, we examined data from a study conducted in class to estimate the average increase in a student's height when wearing shoes (compared to being barefoot). The data is available in the Minitab worksheet for this activity document.

When designing the study, there was some concern that since males and females wear different types of shoes, the increased height when wearing shoes may differ between genders. As a result, in addition to each student's height with and without shoes, we also recorded each student's gender. We are interested in determining whether the increased height when wearing shoes (compared to being barefoot) differs, on average, between males and females.
Let μ_F be the average increase in height when wearing shoes for females. And, let μ_M be the average increase in height when wearing shoes for males. Then, we are interested in testing the following hypothesis:

$$H_0: \mu_F = \mu_M \text{ versus } H_1: \mu_F \neq \mu_M, \text{ or}$$

$$H_0: \mu_F - \mu_M = 0 \text{ versus } H_1: \mu_F - \mu_M \neq 0$$

At first, addressing this question can feel a bit daunting. Since each subject in the study had their height measured with and without shoes, the data is paired (in a sense). This is how we examined the data in class. However, regarding the question of interest in this problem (does the increased height differ between males and females), the data (that is, the difference in height between wearing shoes and not wearing shoes) is not paired since each male is not paired with a specific female. Again, the idea is that once differences have been taken, the pairing disappears.

Compute the increase in height when wearing shoes for each subject. Tip: Use **Calc > Calculator** to compute the differences. This will be the response moving forward.

Explain why a 2-sample *t*-test is not appropriate for addressing the question of interest. Be sure to include any supporting material (graphics) needed.

Solution: Conducting a 2-sample *t*-test carries with it 3 assumptions: the two groups are independent, the sample is random, and each sample comes from a normal population.

It should be easy to argue that the two samples are independent of one another from the context. The males and females in the class are not connected in any way.

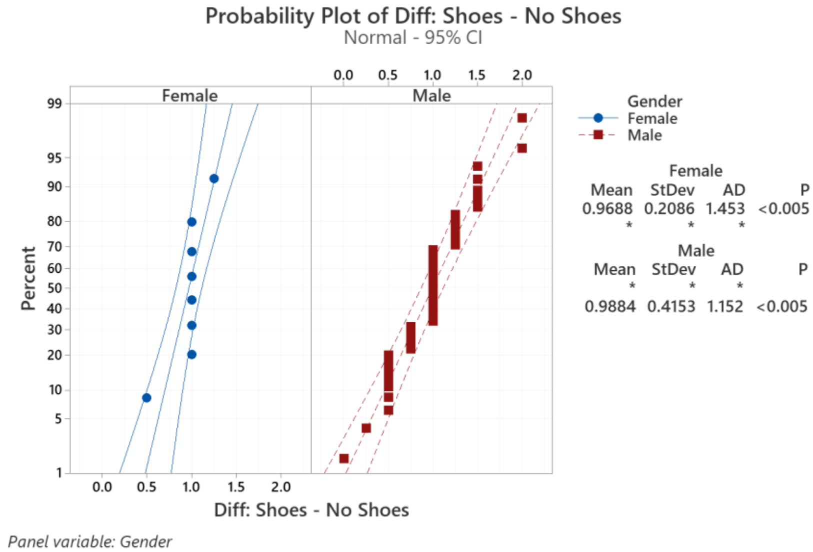
The students in these two statistics class sections are not necessarily a random selection of males and females in all the statistics sections. To move forward, we'll assume that although it is not random, it is fairly representative of students taking this statistics class this quarter.

Keeping in mind that our groups for this question are formed by the gender of the subject (column C11 contains each subject's gender), we see that while we have a large number of males (n = 43), the number of females (n = 8) is quite small. Therefore, we cannot rely on the Central Limit Theorem to guarantee normality for the distribution of sample means for the females.

Since we are now comparing two groups of differences, one of which has a small sample size, we need the differences from each group to have been drawn from a population which follows a normal distribution. A probability plot of each sample is shown below.

From the probability plot of the differences, neither of these samples follows a normal distribution. Although we can still invoke the Central Limit Theorem for the male heights, the sample size for the females is too small. Therefore, modeling the test statistic as a *t* distribution is not appropriate.

Note: You can use the data from this example to perform a hypothesis test in the paired *t* lesson to address the question of there being a mean difference in heights for shoes versus without shoes.



Exercise 5: Baby Walkers

Baby walkers are seats hanging from frames that allow babies to sit upright with their legs dangling and feet touching the floor. Walkers have wheels on their legs that allow the infant to propel the walker around the house long before he or she can walk or even crawl. Typically, babies use walkers between the ages of 4 months and 11 months.

Because most walkers have tray tables in front that block babies' views of their feet, child psychologists have begun to question whether walkers affect infants' cognitive development. One study compared mental skills of a random sample of those who never used walkers with a random sample of those who used walkers. Mental skill scores averaged 123 for 55 babies who did not use walkers (standard deviation of 15) and 113 for 54 babies who used walkers (standard deviation of 12).

If we want to construct a 99% two-sided confidence interval for the difference between the true mean mental score for babies who did not use walkers and the true mean mental score for babies who did use walkers, which test would we use?

A. 1-sample *t*-test on differences (or paired *t*)
B. 2-sample *z*-test
C. 2-sample *t*-test

Solution: C. The group of babies who used walkers is independent of the group of babies who didn't use walkers. So, the data is not paired, and a paired *t* is not the appropriate text. Since the sample sizes are large for each group of babies, we can assume that the distribution of the sample means of mental skill scores for each group is normal. Since we don't know the population standard deviation for either group, the more appropriate test to use is a 2-sample *t*-

test instead of a 2-sample *z* test. Because the degrees of freedom for the *t* distribution will be so large, both the 2-sample *t* test and 2-sample *z*-test will yield similar *p*-values.

Exercise 6: Ice Cream Comparisons

Suppose you want to determine if, on average, there is more or less fat in one brand of ice cream versus another. So, you take a random sample of 8 tubs of Brand A ice cream and determine the following percentages of fat in these 8 tubs.

A: 5.7, 4.5, 6.2, 6.3, 7.3, 6.1, 5.6, 4.7.

Next, you take a random sample of 8 tubs of Brand B ice cream and determine the following percentages of fat in these 8 tubs.

B: 6.3, 5.7, 5.9, 6.4, 5.1, 5.8, 5.6, 5.7.

Which of the following procedures is most appropriate to test equal versus unequal average fat content in the two types of ice cream?

A. 2-sample *t*-test with 8 degrees of freedom
B. Paired *t* test or 1-sample *t* test on differences with 7 degrees of freedom
C. 2-sample *t*-test with 9 degrees of freedom
D. Two sample *z*-test
E. 2-sample *t*-test with 16 degrees of freedom
F. None of these methods is correct

Solution: C. The 8 tubs of ice cream A and 8 tubs of ice cream B are not connected – so they are independent. As long as normality tests show that both data sets are from normal populations, then a 2-sample *t*-test is appropriate. We can verify normality in Minitab, and then we are left with determining the degrees of freedom. By using the data in a 2-sample *t*-test, we can see that the correct degrees of freedom is 9. We can compute this value by-hand (since the formula is in the 2-sample *t*-test on population means lesson notes), but it's much more efficient to simply use Minitab's **Stat > Basic Statistics > 2-sample t**.

Exercise 7: Weight-Loss Results

To test the effectiveness of the new weight-loss drug "Reducts," 40 women were randomly split into two groups (e.g. names drawn out of a hat): Group A, the control group, took a placebo drug, and Group B, the experimental group, took Reducts. The amount of weight lost by each member of the groups over a 6-month period is given below.

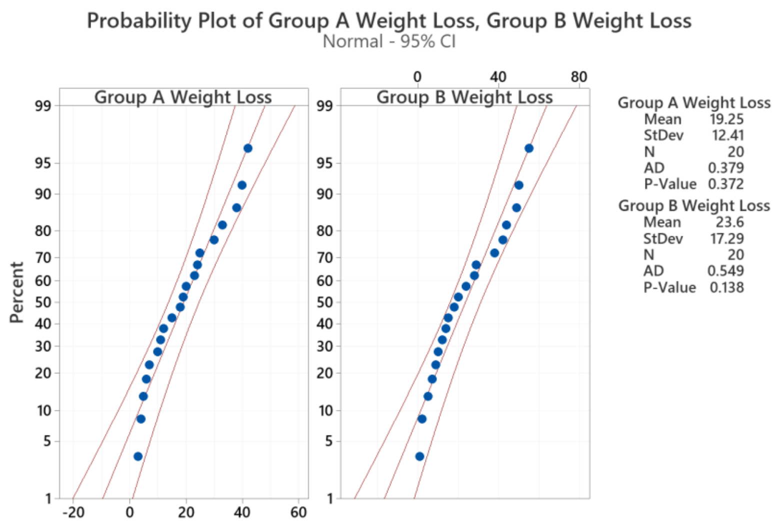
Group A: 3, 4, 5, 6, 7, 10, 11, 12, 15, 18, 19, 20, 23, 24, 25, 30, 33, 38, 40, 42

Group B: 1, 2, 5, 7, 9, 10, 12, 14, 15, 18, 20, 24, 28, 29, 38, 42, 44, 49, 50, 55

Based on these results, is Reducts ***effective*** (i.e. increased weight loss)? We'll run a hypothesis test to determine this.

Using the appropriate hypothesis test, provide the standardized test statistic and corresponding *p*-value for that test statistic. Make sure that all necessary assumptions have been met to run the given hypothesis test that you have chosen.

Solution: The 40 women were randomly assigned to the two groups. There is no connection between the women in Group A and the women in Group B, so the data is independent. The appropriate test to perform is a 2-sample *t*-test as long as the data from both groups is from a normal population. The AD test statistic *p*-values for the normality tests for Group A and Group B are 0.372 and 0.138, respectively. Thus, we can assume their underlying distributions are normal.



Using the data already in the Minitab worksheet for this activity, we can select the 2-sample *t*-test with a "greater than" hypothesis if the null and alternative are written as:

$$H_0: \mu_B = \mu_A \text{ versus } H_1: \mu_B > \mu_A, \text{ or } H_0: \mu_B - \mu_A = 0 \text{ versus } H_1: \mu_B - \mu_A > 0$$

The standardized test statistic is $t_0 = 0.91$ with *p*-value = 0.184.

Test

Null hypothesis	$H_0: \mu_B - \mu_A = 0$
Alternative hypothesis	$H_1: \mu_B - \mu_A > 0$

T-Value	DF	P-Value
0.91	34	0.184

Exercise 8: Lab vs Golden Retriever Weights

After seeing a picture of my student's Labrador Retriever on his computer desktop, I was curious as to whether Labrador Retrievers (L) or Golden Retrievers (G) weigh more as adult dogs. So, I decided to collect data to determine if their true mean weights are the same or not. The data I collected for each breed is in the Minitab worksheet for this activity. Perform the following hypothesis test.

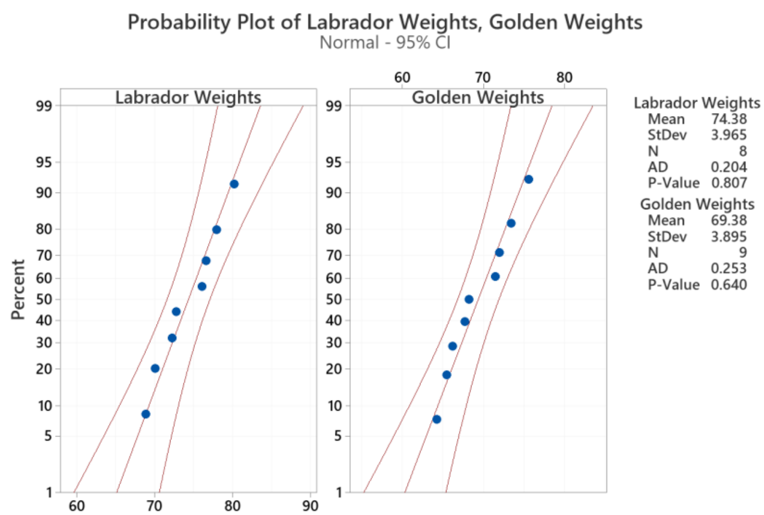
$$H_0: \mu_L = \mu_G \text{ versus } H_a: \mu_L \neq \mu_G$$

(a) Is the data paired or independent?

Solution: The randomly selected dogs are not paired. In fact, the sample size for each group is different, so the two groups can't be paired because there's not a "one-to-one" linkage.

(b) Determine the standardized test statistic.

Solution: In Minitab, perform a 2-sample *t*-test after verifying that the weights for each group appear to be from normal distributions. Since the *p*-values for the AD normality tests are much larger than α = 0.05, we don't have evidence suggesting that normality assumption is not valid.



Using Minitab's **Stat > Basic Statistics > 2-sample t**, we obtain the standardized test statistic t_0 = 2.63.

Test

Null hypothesis	$H_0: \mu_L - \mu_G = 0$
Alternative hypothesis	$H_1: \mu_L - \mu_G \neq 0$

T-Value	DF	P-Value
2.62	14	0.020

(c) Determine the *p*-value associated with this test statistic.

Solution: Minitab's 2-sample *t*-test reports the "not equal to" alternative hypothesis's *p*-value as 0.020.

(d) At $\alpha = 0.05$, can we reject H_0?

Solution: Yes. Since the *p*-value is less than 0.05, then our data suggests the rejection of H_0. Our evidence suggests that Labrador Retriever weights are not equal to Golden Retriever weights, on average.

Exercise 9: Mice Maze Times

The times (in minutes) it took six white mice to complete a maze and the times it took six brown mice to complete the same maze are given below. At $\alpha = 0.05$, does the color of the mice make a difference in their learning rate?

White Mice	18	24	20	13	15	12
Brown Mice	25	16	19	14	16	10

Use a 95% confidence interval for the difference of their mean times to complete the maze to answer this question. Assume the times for both groups are from normally distributed populations.

Solution: Let μ_W be the mean time through the maze for white mice, and μ_B the mean time for brown mice. I'm going to build the two-sided 95% confidence interval for $\mu_W - \mu_B$.

Although the sample sizes of the two groups are the same, the groups are not paired. There is no connection between the time to complete the maze for white mouse *i* and brown mouse *i*, where *i* = 1, 2, …, 6. Unlike most of the other examples in this activity document, we do not need to perform normality tests on the two groups. We are told in the problem statement that the times are from normally distributed populations. Since we have independent groups from normally distributed populations (and σ is unknown), the appropriate test to use is a 2-sample *t*-test. The 95% two-sided confidence interval for the difference in means, $\mu_W - \mu_B$, is:

Estimation for Difference

Difference	**95% CI for Difference**
0.33	(-5.95, 6.62)

Since the 95% two-sided confidence interval for the difference contains the value 0, then we don't have enough evidence at $\alpha = 0.05$ to reject H_0: $\mu_W - \mu_B = 0$ (or H_0: $\mu_W = \mu_B$). Our data doesn't allow us to state that white mice and brown mice complete the maze at different times, on average. Note: The 95% confidence interval for $\mu_B - \mu_W$ is (-6.52, 5.86) minutes. We still conclude that we don't have evidence to suggest that their maze completion times are different, on average.

Exercise 10: Employee Sick Days

A large corporation is interested in determining whether the average days of sick leave taken annually is more for the nightshift employees than for the dayshift employees. It is assumed that the distribution of the days of sick leave is normal for both shifts and that the variances are unknown and assumed unequal.

A random sample of 12 employees from the nightshift yields an average sick leave of 16.4 days with a standard deviation of 2.2 days. A random sample of 15 employees from the dayshift yields an average sick leave of 12.3 days with a standard deviation of 3.5 days.

The hypotheses are: H_0: $\mu_{NS} = \mu_{DS}$ vs H_1: $\mu_{NS} > \mu_{DS}$, or H_0: $\mu_{NS} - \mu_{DS} = 0$ vs H_1: $\mu_{NS} - \mu_{DS} > 0$.

(a) Determine the standardized test statistic for the hypothesis test.

Solution: Since the sample sizes are small but the distribution of sick leave is normal for both shifts, then a 2-sample *t*-test with non-equal variances is required here. I'll compute the test statistic by-hand for this example.

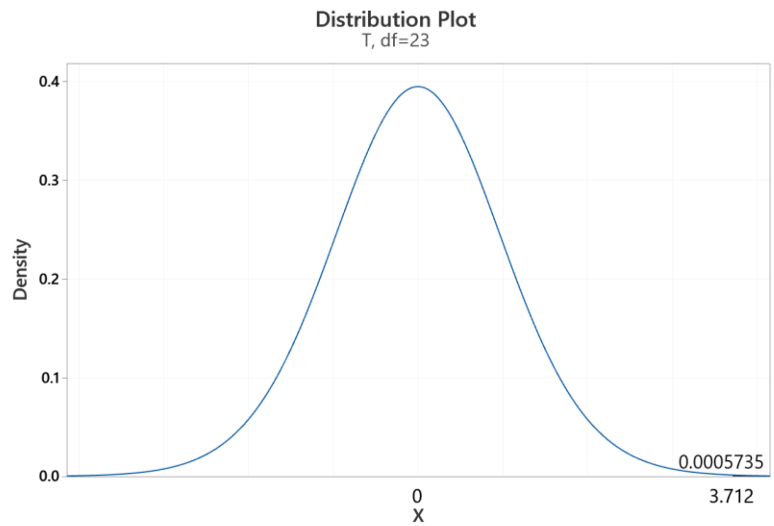
$$t_0 = \frac{(16.4 - 12.3) - (0)}{\sqrt{\frac{2.2^2}{12} + \frac{3.5^2}{15}}} \cong 3.712.$$

The degrees of freedom for the *t* distribution is:

$$df = \text{floor}\left(\frac{\left(\frac{2.2^2}{12} + \frac{3.5^2}{15}\right)^2}{\frac{\left(\frac{2.2^2}{12}\right)^2}{11} + \frac{\left(\frac{3.5^2}{15}\right)^2}{14}}\right) \cong \text{floor}(23.842) = 23.$$

(b) Should we reject H_0 at level of significance $\alpha = 0.01$?

Solution: Since we know the test statistic and degrees of freedom, we can use Minitab's **Graph > Probability Distribution Plot > View Probability > t** Distribution with **23** Degrees of freedom to determine the *p*-value associated with $t_0 = 3.712$. It is ~0.00057.



Since the *p*-value is less than α = 0.01, then **reject H_0** at level of significance α = 0.01. We have data that supports nightshift employees taking more sick days than dayshift employees, on average.

Exercise 11: DDT Found in Pelicans

A study is being conducted by the Florida Game and Fish Commission to assess the amounts of chemical residues found in the brain tissue of brown pelicans. Specifically, the commission is interested in determining if the mean amount of DDT found in juvenile pelicans is larger than the mean amount of DDT found in nestling pelicans. [A nestling pelican is one that is between the hatching and leaving the nest period.] This test has important implications regarding the accumulation of DDT over time.

A random sample of 10 juvenile pelicans is selected and the DDT amounts in parts per million (ppm) are recorded in the column "Juvenile" in the Minitab worksheet for this activity. Likewise, a random sample of 10 nestling pelicans is selected and the DDT amounts in parts per million (ppm) are in the column "Nestling."

Let $\mu_{juvenile}$ be the true average amount of DDT (in ppm) found in juvenile pelicans in this area, $\mu_{nestling}$ be the true average amount of DDT (in ppm) found in nestling pelicans in this area, and μ_{diff} be the true mean difference in the amount of DDT in juveniles versus nestlings. Specifically, let "diff = juvenile – nestling."

(a) Select the appropriate null and alternative hypotheses for this experiment using the parameters defined in the above paragraph. Since the alternative can be written in various ways depending on the order of subtraction for the variable "diff," there may be more than one correct answer. Select all the correct choices.

A. H_0: $\mu_{juvenile} - \mu_{nestling} = 0$ versus H_a: $\mu_{juvenile} - \mu_{nestling} < 0$
B. H_0: $\mu_{nestling} - \mu_{juvenile} = 0$ versus H_a: $\mu_{nestling} - \mu_{juvenile} > 0$
C. H_0: $\mu_{nestling} - \mu_{juvenile} = 0$ versus H_a: $\mu_{nestling} - \mu_{juvenile} \neq 0$
D. H_0: $\mu_{juvenile} - \mu_{nestling} = 0$ versus H_a: $\mu_{juvenile} - \mu_{nestling} > 0$
E. H_0: $\mu_{diff} = 0$ versus H_a: $\mu_{diff} < 0$
F. H_0: $\mu_{juvenile} - \mu_{nestling} > 0$ versus H_a: $\mu_{juvenile} - \mu_{nestling} = 0$

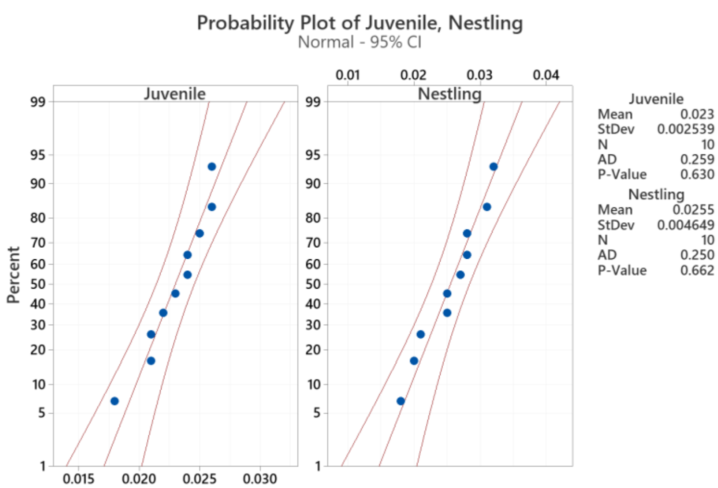
Solution: D. It is the only correct answer. As stated in the last paragraph on the problem statement, I'm letting "diff = juvenile – nestling."

(b) Is the data paired or independent?

Solution: Independent. There is no indication that the nestling and juvenile pelicans are related to each other in any way. The commission didn't try to take a nestling and juvenile pelican from the same "family" of pelicans. Thus, DDT amounts in juveniles and DDT amounts in nestlings are independent.

(c) Determine the standardized test statistic for the appropriate hypothesis test.

Solution: Since the two groups of data are independent, then the appropriate statistical test is the 2-sample *t*-test as long as we can show that both groups come from normal distributions. We'll confirm this first with normality tests.



The standardized test statistic is:

$$t_0 = \frac{0.0255 - 0.023}{\sqrt{\frac{0.00465^2}{10} + \frac{0.00254^2}{10}}} \cong \mathbf{1.492}$$

Using Minitab, we can also obtain the standardized test statistic and the p-value associated with it.

Test

Null hypothesis	H_0: $\mu_J - \mu_N = 0$
Alternative hypothesis	H_1: $\mu_J - \mu_N > 0$

T-Value	DF	P-Value
1.49	13	0.080

(d) What is the p-value for the test you have conducted?

Solution: 0.080

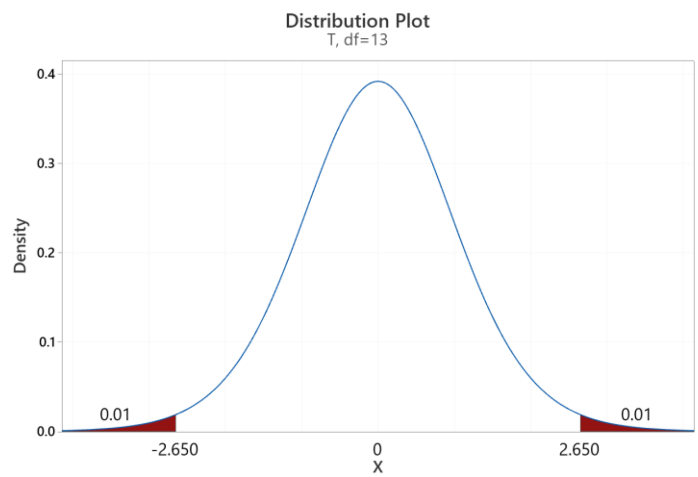
(e) Based on the significance level $\alpha = 0.05$, what decision should the commission reach?

Solution: Do not reject H_0. We do not have enough evidence at $\alpha = 0.05$ to conclude that the average DDT amount in juveniles is greater than the average DDT amount in nestling pelicans.

(f) Determine the correct t critical value (cv) in the expression below to construct a 98% two-tailed confidence interval for the difference in the mean amounts of DDT.

$$0.0025 \pm \boldsymbol{cv} \cdot \sqrt{\frac{0.00465^2}{10} + \frac{0.00254^2}{10}}$$

Solution: We have a t distribution with $df = 13$. In constructing a 98% two-sided confidence interval, we'll want to leave 0.01 area in each tail of the t distribution. We can use Minitab's **Graph > Probability Distribution Plot > View Probability > t** Distribution with **13** Degrees of freedom to determine the correct critical value, which is 2.650.



Exercise 12: Drinking Water and Weight Loss

A new study published in the journal *Obesity* found that pre-loading water *before* meals helps you lose weight. The study looked at 84 obese adults and had 41 members of the group drink around 16 ounces of water before meals, while the other 43 adults were asked simply to imagine being full before digging into their food.

Over the course of the 12 weeks, those who filled up on water prior to eating the three main meals a day lost an average of 9.48 pounds with a standard deviation of 3.12 pounds, whereas those imagining they were full before meals resulted in an average loss of 1.76 pounds with a standard deviation of 0.21 pounds.

Let μ_1 be the average weight loss for the participants drinking 16 ounces of water before meals and let μ_2 be the average weight loss for participants who imagined they were full before meals. The authors were interested in results of the following hypothesis test. Their intention was to show that the average weight loss for participants who drank water prior to meals was greater than the average weight loss for participants imagining they were full before meals.

$$H_0: \mu_1 - \mu_2 \leq 0 \qquad \text{vs.} \qquad H_1: \mu_1 - \mu_2 > 0$$

(a) Compute an appropriate confidence interval for addressing this hypothesis at $\alpha = 0.025$ significance level. Write down *both* the formula and the final answer. Use the *most* appropriate method to model the underlying distribution, e.g. normal, Student's *t*, *F*, …

Solution: Since the alternative hypothesis is a "greater than," then the two-sided confidence interval should be a 95% CI.

Since both sample sizes are large (*n*'s > 30), then by the Central Limit Theorem, each group's distribution of sample means is normal. We can compute the interval "by-hand" using the confidence interval formula for a 2-sample *t*-test.

$$(\bar{x}_1 - \bar{x}_2) \pm t_{0.025,40} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \rightarrow (9.48 - 1.76) \pm 2.021 \cdot \sqrt{\frac{3.12^2}{41} + \frac{0.21^2}{43}} \rightarrow 7.72 \pm 2.021 \cdot \sqrt{0.488}$$

(6.73, 8.71) lbs

(b) Given the confidence interval that you computed in part (a), is it possible that the difference in mean weight loss amounts could be as much as 8.5 pounds at $\alpha = 0.025$?

Solution: Yes, it's possible since 8.5 pounds is in the 95% two-sided confidence interval above.

Exercise 13: Fish Tank Filters from "Finding Nemo"

After the "Tank Gang's" escape, the dentist P. Sherman decides to compare two filters for use in his fish tank. Specifically, he is interested in determining if there is evidence that a power filter will remove a higher percentage of contaminants compared to an internal filter. He takes a sample of 40 fish tanks and observes the type of filter currently in use and the percentage of contaminants removed. Below is a summary of the data.

Filter	N	Mean	Std. Dev.
Power	20	98.5	6.91
Internal	20	92.3	10.63
Diff. (Power – Internal)	20	6.2	8.27

(a) State the null and alternative hypothesis that best addresses the question of interest. Be sure to define any mathematical notation used.

Solution: Let μ_1 be the average percentage of contaminants removed using the Power filter and μ_2 be the average percentage of contaminants removed using the Internal filter. Then, we are interested in the following hypothesis test:

$$H_0: \mu_1 - \mu_2 = 0 \qquad \text{vs.} \qquad H_1: \mu_1 - \mu_2 > 0$$

(b) Compute an appropriate confidence interval for addressing the hypothesis stated in part (a) at the 0.01 significance level. Be sure to state *both* the formula and the final answer. You may assume that the percentage of contaminants removed by each filter is from a normal distribution.

Solution: Since we have two independent groups and the sample size is small, then Student's t is appropriate given percentage of contaminants removed by each filter is from a normal distribution. The formula for constructing a 2-sample *t*-test confidence interval is:

$$(\bar{x}_1 - \bar{x}_2) \pm t_{v,0.99} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}},$$

where υ represents the degrees of freedom of the *t* distribution. The problem just states to provide the formula for the confidence interval and the final answer. Since determining the degrees of freedom υ can be tedious, I would just find the corresponding interval in Minitab with **Stat > Basic Statistics > 2 Sample t**. Select **"Summarized data"** from the drop-down menu and enter the values given in the table above. Since the alternative hypothesis is "greater than" and we want a one-sided confidence interval to address the hypothesis at α = 0.01. We obtain:

Estimation for Difference

Difference	99% Lower Bound
6.20	-0.74

Notice that Minitab also print the degrees of freedom υ = 32 if you'd like to compute the *t* critical value using the Minitab's **Probability Distribution Plot**.

Test

Null hypothesis	H_0: $\mu_1 - \mu_2 = 0$
Alternative hypothesis	H_1: $\mu_1 - \mu_2 > 0$

T-Value	DF	P-Value
2.19	32	0.018

The by-hand construction of the confidence interval is:

$$(\bar{x}_1 - \bar{x}_2) - t_{\nu,0.99} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \rightarrow (98.5 - 92.3) - t_{32,0.99} \cdot \sqrt{\frac{6.91^2}{20} + \frac{10.63^2}{20}} \rightarrow 6.2 - 2.449 \cdot 2.835 \rightarrow$$

-0.74.

(c) Using the interval constructed in part (b), what conclusions can be drawn? Be sure to state your conclusions in the context of the problem.

Solution: Since 0 is within the interval (-0.74 →), we cannot reject H_0 at $\alpha = 0.01$ to conclude that the power filter removes a greater percentage of contaminants than the internal filter, on average. There is no evidence to suggest that the power filter remove a higher percentage of contaminants, on average, at $\alpha = 0.01$.