ANSWER KEY: PAIRED T-TEST FOR POPULATION MEANS
This answer key provides solutions to the corresponding activity sheet.

Paired *t*-Test

The data for these exercises are in the Minitab file ***PairedTTest_Activity.mtw***.

Exercise 1

For each of the following scenarios, determine if the two samples are **independent or paired.** If the data is:

- **Paired**, determine the paired differences, and proceed to the **1-Sample *t*-Test on One Population Mean** lesson and/or activity sheet.
- **Independent**, proceed to the **2-Sample *t*-Test on Two Population Means** lesson and/or activity sheet.

(a) In the small town where I grew up, companies were dumping their trash and chemical waste in a field that was a couple miles from our house. Later it was determined that trace metals were leaking into drinking water wells of homes and businesses near the dumping site. The waste not only affected the flavor of the water, but it posed a health hazard to community members.

To test the depth of the problem, city officials randomly chose twelve different household water wells within one mile of the dumpsite. The water in the wells varied in the concentration of the trace metals depending on their distance from the site.

To test the metal concentrations (mg/L) of zinc in the wells, a container was lowered into the water at each well, and a first sample was drawn from the bottom of the well. At the same well, another sample of water was drawn from the top of the well. The zinc concentrations are available in the Minitab worksheet for this activity sheet.

Are the two samples of data that they collected paired or independent?

Solution: Paired. Zinc concentrations are recorded twice at each well. One sample of data is from the bottom of the well and the other is from the top of the well. The two groups of data are paired by well.

(b) A business owner wonders whether female customers tend to make larger purchases than males. She randomly selects 40 receipts taken from female customers during the last week and 40 from male customers and compares the average expenditures.

Solution: Independent. There is no relationship or connection between the 40 female customers and 40 male customers. These are two independent groups.

(c) An educator believes that by using Minitab in statistics courses that students will obtain a better understanding of the material. She is assigned to teach two sections of Introductory Statistics at her college the following semester. In one class, there are 21 students, and Minitab is used as part of the curriculum. In the other class of 23 students, she follows the same curriculum without using Minitab. At the end of the semester all students are given the same exam to measure their understanding of the material.

Solution: Independent. One big indicator of independence is that the two classes have different sample sizes. Pairing is impossible since there has to be a one-to-one relationship between members of both classes. These are two independent groups. Although they are learning the same material, there is nothing connecting student i in Class 1 to student i in Class 2.

(d) A shoe company is interested in comparing the number of shoes owned by adult women (over 18) versus adult men in the U.S. In a large research study, company representatives obtain a random sample of 250 adults and ask each individual if they identify as a woman or a man. Next, they ask each person how many pairs of shoes they own.

Solution: Independent. The men and women in this study are in two independent groups. They randomly sample 250 adults in the U.S. and there is likely no connection between any two of them. Also, it would be unusual if their sample contains the same number of women as men.

(e) The shoe company from part (d) decides to invest in a second study with a different research design. This time they take a random sample of 250 heterosexual adult (both over 18) married couples in the U.S. (i.e. 250 husbands and 250 wives). They record the number of shoes owned by each husband and each wife. Do the women or men own more shoes, on average?

Solution: Paired. In this new research design, the woman/man data is paired by married couple.

(f) A study on facial expression and age was conducted to determine if 1-year-olds exhibit more or fewer facial expressions compared to 2-year-olds when given silly emotional cues by adults. The number of facial expressions exhibited in response to 16 silly emotional cues for 25 randomly selected 1-year-old subjects and 25 randomly selected 2-year-old subjects was recorded.

Solution: Independent. No connection is indicated between the 1-year-olds and 2-year-olds selected for this study.

(g) Several years ago, I broke some bones in my left hand. When I went to physical therapy, the therapist measured the amount of water displaced by my left hand when I put it in a volumeter. In order to determine if there was excessive swelling, she had me put my healthy right hand in the volumeter to see how much water that it displaced.

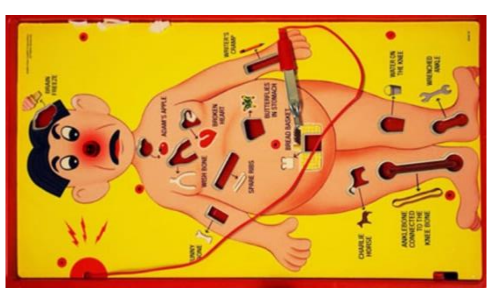
I decided that I wanted to determine if a person's dominant (writing) hand displaces more water in a volumeter than their non-dominant hand. I randomly selected 12 students in my statistics course and had them record the amount of water displaced by both their dominant and non-dominant hands. We determined that there was no significant difference.

Solution: Paired. Both hands are "connected" to each student.

(h) Suppose that I'm interested in performing a hypothesis test to compare the mean final exam score of females and the mean exam score of males in a large statistics class. I randomly select 10 females from the class and then randomly select 10 males. I arrange the females' names alphabetically and use this list to assign each female a number between 1 and 10. I do the same thing for the males.

Solution: Independent. No matter how the instructor arranges the female and male names, the two people have no connection to each other. In other words, female i and male i in position i are not related.

(i) In my Six Sigma course, we use Operation games to collect data on times to remove parts from Cavity Sam. Each student completes surgeries both with and without surgical gloves.



One part that is particularly hard to remove is the "Charlie Horse." In order to determine if the surgical gloves affect their times to remove a part, I randomly select 20 students and ask them to report their times to remove the Charlie Horse with and without the gloves. Do the gloves affect their times?

Solution: Paired. For each student, they have a time to remove the Charlie Horse with gloves and without gloves. The two times are paired by student.

(j) Since I've improved at playing pickleball, I rewarded myself with a pair of court shoes. I bought court shoes because I seemed to be falling down a lot when playing with running shoes. Throughout the summer, I randomly switched between wearing my court shoes and running shoes. Each night I recorded the number of times I fell while playing and what shoe I was wearing. In the fall, I'm going to randomly select 30 days from the summer and compare the number of falls for each of the two shoes. Did I actually fall less with the court shoes?

Solution: Independent. Each day I'm either wearing one pair of shoes or the other: court or running shoes. When I randomly select the 30 days, there is no connection between day i and day $i + 1$. It's very likely that the random days selected won't have half of them being court shoes days and the other half being running shoes days.

(k) There is both a state college and private college in my town. Both colleges teach Calculus I with the same textbook and technology. Suppose the instructors from the two schools write a common final exam. They want to compare how students at the two different schools perform on this exam. To do so, they randomly select 40 students from each school and obtain their final exam scores. Does one of the two schools have a higher final score average?

Solution: Independent. The two groups of students are independent. There is no connection between student i at the state college and student i at the private college.

(l) A medical assistant sampled the blood pressures of 20 randomly selected patients with high blood pressure before and after they receive a dose of a new medication. Which hypothesis test should she run to determine if the medication affected the patients' blood pressures?

Solution: Paired. The two blood pressures, before and after, are paired by patient. This is a "classic" paired t-test example. The researchers do a before/after experiment on their subjects and compare the before and after results.

(m) In my opinion, to be good at the game Scrabble, you have to have a healthy vocabulary of obscure two and three letter words. In general, is a good vocabulary dependent on age?

An investigator thinks that people under the age of forty have vocabularies that are different than those of people over sixty years of age. The investigator administers a vocabulary test to a randomly selected group of 31 younger subjects and to a randomly selected group of 31 older subjects. Higher scores reflect better performance.

Solution: Independent. There are 31 younger subjects and 31 older subjects. There is no indication that the younger and older subjects are paired in any way. If both the younger and older subject came from the same family, then we would have some connection.

(n) A manufacturer of socks wants to determine if a special type of wool socks lasts longer than the cotton socks that they manufacture. He makes pairs of socks in which the left foot is cotton and the right foot is wool. We sample 24 women from those who have volunteered to wear these pairs of socks. We want to determine if there is a difference in "number of wears before tears" for each type of material.

Solution: Paired. For each woman, they have a cotton sock on their left foot and wool sock on their right foot. There are two data points associated with each of the 24 women.

(o) I meet an exercise trainer 3 mornings a week. He has dozens of other clients, and we do similar exercise routines. A graduate student studying exercise science wanted to know how different routines affect heartbeat rates. She randomly selected twelve of his trainees and asked them to perform two different exercise routines while monitoring their heartbeats. The aim of the study was to test for a difference in the effect of the routines on heartbeat rate.

Solution: Paired. For the twelve trainees who are selected, their heartbeat rates are measured twice after two different exercise routines.

(p) A drug company is concerned about the length of time a particular drug retains its potency. A random sample of 10 bottles of the product is drawn from the current production line and analyzed for potency. A second sample is obtained, stored for 1 year, and then analyzed. Does the drug lose potency, on average, over time?

Solution: Independent. The original 10 bottles sampled are not connected to the second sample drawn. If the company sampled 10 bottles and tested one pill from that bottle now and one pill from that bottle in a year, then the before/after data would be paired by bottle. As the problem is written, the two samples of 10 bottles are not paired.

(q) For a random sample of nine autos, the mileage (in 1000's of miles) at which the original front brake pads were worn to 10% of their original thickness was measured, as was the mileage at which the original rear brake pads were worn to 10% of their original thickness. The manufacturer suspects that the average mileage at which the front pads reach 10% of their thickness is less than the average mileage for the rear pads.

Solution: Paired. For the nine autos, the mileages associated with their front brake pad wear and rear brake pad wear are recorded. Both pieces of information are linked to each auto.

(r) Cedar-apple rust is a (non-fatal) disease that affects apple trees. Its most obvious symptom is rust-colored spots on apple leaves. Red cedar trees are the immediate source of the fungus that infects the apple trees. If you could remove all red cedar trees within a few miles of the orchard, you should eliminate the problem. In the first year of this experiment the number of affected leaves on 8 trees was counted. The following winter all red cedar trees within 100 yards of the orchard *were removed* and the following year the same trees were examined for the number of affected leaves.

Solution: Paired. The orchard is counting the number of affected leaves on 8 specific trees. It counts the number in the first year of the experiment. Then the next winter the orchard removes the red cedar trees that could be affecting the apple trees' condition. In the second year, the same 8 trees are examined for the number of affected leaves. The orchard is surely hoping that the number of affected leaves is lower after the removal of the red cedar trees.

Exercise 2: Repair Shop Charges for Males vs Females

A study was conducted to determine whether automobile repair charges are higher for female customers than for male customers. Twenty auto repair shops were randomly selected from the same city. Two cars of the same age, brand, and engine problem were used in the study. For each repair shop, the two cars were randomly assigned to a male and female participant and then taken to the identical shop for an estimate of the repair costs. Some of the repair costs (in dollars) are given here. There were n = 20 auto repair shops.

Repair Shop	**1**	**2**	**3**	**4**	**5**	**6**	**…**	**17**	**18**	**19**	**20**
Female customers	871	684	795	838	1033	917	…	1157	932	1089	770
Male customers	792	765	511	520	618	447	…	884	702	839	878

(a) The data (female repair cost, male repair cost) is paired by what item?

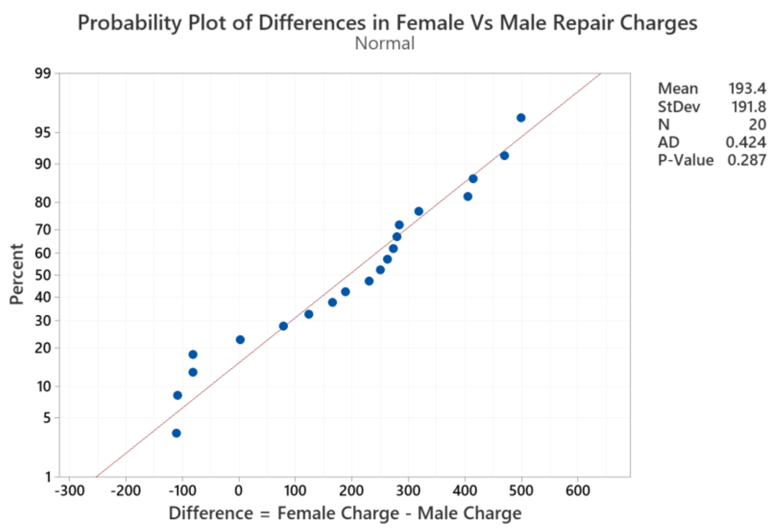
Solution: The data is paired by the repair shop. There are 20 pairs.

(b) Are the repair charges higher for female customers than male customers, on average?

Solution: There are n = 20 pairs of data. Since the data is paired, I prefer to do a **1-sample *t*-test** (rather than a paired *t*-test) on the mean difference in female versus male repair charges. Let D be the difference in repair costs, where D = female repair charge – male repair charge. We want to have a "greater than" alternative hypothesis due to the way in which we defined the variable D. Note: If we defined D as "male repair charge – female repair charge," then the alternative hypothesis would be a "less than" alternative.

$$H_0\text{: } \mu_D = 0 \text{ versus } H_a\text{: } \mu_D > 0$$

Since the sample size is small, we need to determine if the differences are from a normally distributed population. Use Minitab's **Stat > Basic Statistics > Normality Test** and select the column of differences in the Minitab worksheet.



Since the variable D is from a normal population, then we can use a 1-sample *t*-test and base our decision on its *p*-value. Use Minitab's **Stat > Basic Statistics > 1-sample t** and select the column of differences for the hypothesis test. Check the box **Perform hypothesis test** and select **Options**. The **Alternative Hypothesis** needs to be a "greater than" alternative. The resulting Minitab output is:

Test

Null hypothesis	H_0: $\mu = 0$
Alternative hypothesis	H_1: $\mu > 0$

T-Value	P-Value
4.51	0.000

Since the *p*-value is 0, our evidence suggests that the mean difference is greater than 0. This means, on average, that females are paying more for repair charges than males.

Exercise 3: Full Moon Behavior

Many people believe that the moon influences actions of some individuals. A study of dementia patients in nursing homes recorded various types of behavior every day for 12 weeks. Days were classified as moon days if they were in a 3-day period centered at the day of a full moon. For each patient, the average number of disruptive behaviors was computed for moon days and for all other days.*

* From *Introduction to the Practice of Statistics Excel Manual with Macros*
By David S. Moore, Linda Getch Dawson, George P. McCabe

In the Minitab columns labeled "Moon Days" and "All Other Days," the average number of disruptive behaviors were recorded for 15 randomly selected patients. Determine if there is a difference in the average number of disruptive behaviors for "Moon Days" and "All Other Days."

Let μ_{moon} be the true average number of disruptive behaviors for dementia patients in nursing homes during moon days, μ_{other} be the true average number of disruptive behaviors for dementia patients in nursing homes during the "other" days, and μ_{diff} be the true mean difference in the number of disruptive behaviors in "all other days" versus "moon days."

(a) Which choice below best represents the null and alternative hypotheses that we are testing? Use the parameters defined in the above paragraph.

A. H_0: $\mu_{diff} = 0$ versus H_a: $\mu_{diff} \neq 0$
B. H_0: $\mu_{other} - \mu_{moon} = 0$ versus H_a: $\mu_{other} - \mu_{moon} < 0$
C. H_0: $\mu_{other} - \mu_{moon} = 0$ versus H_a: $\mu_{other} - \mu_{moon} > 0$
D. H_0: $\mu_{diff} = 0$ versus H_a: $\mu_{diff} > 0$, where diff is defined as "moon days" – "all other days"
E. H_0: $\mu_{diff} = 0$ versus H_a: $\mu_{diff} > 0$, where diff is defined as "other days" – "moon days"
F. None of these is correct.

Solution: A. The data is paired by patient. We are told in the problem's instructions to *determine if there is a difference in the average number of disruptive behaviors for "Moon Days" and "All Other Days."* Since we are only asked to determine if there is a difference (and not to determine if there are more or less disruptive behaviors on the various nights), the most correct answer is A.

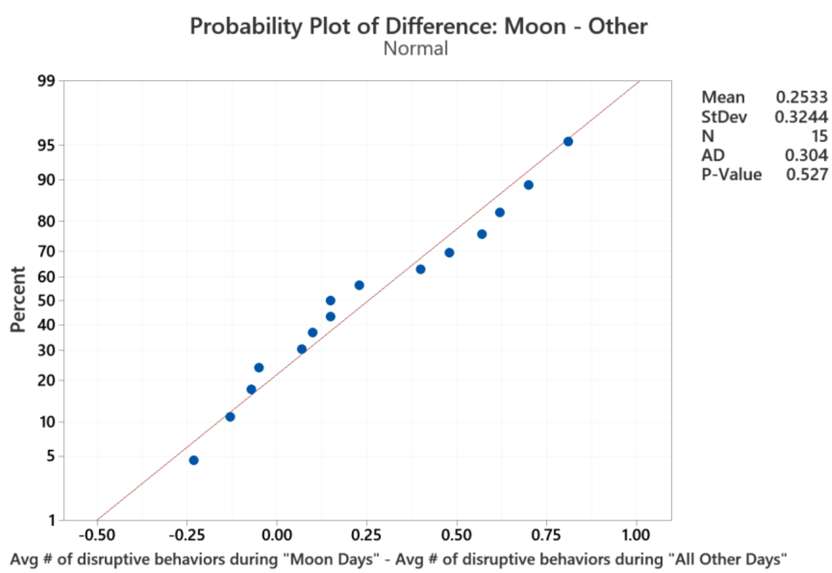
(b) Circle the test that you will be conducting in addressing this data analysis situation.

1-sample *z*-test	1-sample *t*-test	2-sample *z*-test
2-sample *t* test	paired *t* test or 1-sample *t*-test on the differences	

(c) Circle ALL the choices below that justify your reasons for deciding on this test from part (b). There IS more than one correct choice.

A. The patients were randomly selected.
B. The data for the average number of disruptive behaviors on "Moon Days" and the average number of disruptive behaviors on "All Other Days" is paired by patient.
C. The average number of disruptive behaviors on "Moon Days" and the average number of disruptive behaviors on "All Other Days" are independent.
D. The sample size *n* is small (*n* < 30).
E. The sample size n is large ($n \geq 30$).
F. The differences for the average number of disruptive behaviors on "Moon Days" and the average number of disruptive behaviors on "All Other Days" come from a normal distribution according to a normality test.
G. The population variance is unknown.
H. The population variance can be determined.
I. We can use the Central Limit Theorem to say that the differences come from a normal distribution.

Solution: The following letters must be selected.
A. We are told that the patients were randomly selected in the problem statement.
B. This is paired data – the avg # of disruptive behaviors during each type of day is paired by patient.
D. The sample size is $n = 15$; there are 15 patients, and we record two data points for each patient.
F. The data is from a normal distribution. Students should check this using a normality test. An AD normality test is shown at the bottom of this solution.
G. The population variance is unknown. We can compute a sample variance for the differences.



(d) In the space below, provide the statistical documentation associated with your analysis. For full credit, this should include the test statistic(s), including the formula you have used to compute the test statistic, and Minitab output.

Solution: $\bar{x}_{diff} \cong 0.2533, s_{diff} \cong 0.3244$

$$t_0 = \frac{0.2533 - 0}{\frac{0.3244}{\sqrt{15}}} \cong \mathbf{3.024}$$

Descriptive Statistics

N	Mean	StDev	SE Mean	95% CI for μ
15	**0.2533**	**0.3244**	0.0838	(0.0737, 0.4330)

μ: population mean of Difference: Moon - Other

Test

Null hypothesis	H_0: μ = 0
Alternative hypothesis	H_1: μ ≠ 0

T-Value	P-Value
3.02	0.009

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

Solution: 0.009

(f) Based on the significance level α = 0.05, what decision do you reach? Circle one:

Reject H_0 Fail to Reject H_0

(g) In the space below, provide the appropriate confidence interval corresponding to α = 0.05, including the formula you have used to compute it.

Solution: From the Minitab output for part (d), we know the values in the confidence interval. In computing the confidence interval by-hand, we have all the necessary pieces from part (d), except the *t* critical value (bolded below). The *t* critical value is $t_{0.025,14}$ because we are modeling the differences by a *t* distribution with *df* = 14.

$$0.2533 \pm \mathbf{2.145} \cdot \frac{0.3244}{\sqrt{15}} \cong [\mathbf{0.0736}, \mathbf{0.4330}]$$

Exercise 4: Who's on First?

Charlie Brown manages potentially the worst baseball team to have ever played. In fact, they have managed to win only a handful of games. Charlie Brown (who also pitches for the team) believes Lucy van Pelt (who plays outfield on the team) is the worst player in the history of the game, and Lucy often berates Charlie Brown from the outfield. Charlie Brown decides to collect data to show that he is the better player.

Specifically, Charlie Brown is interested in showing that, on average, Lucy makes more errors than he does. He collects a random sample of 45 games and records the number of errors Lucy makes (LvP Errors) and the number he makes during the game (CB Errors). The data is provided in the Minitab worksheet for this activity topic.

(a) State the null and alternative hypothesis appropriate for addressing the question of interest. Be sure to define the parameter(s) of interest.

Solution: Let μ_d be the average difference in errors (Lucy minus Charlie Brown); then, we are interested in testing:

$$H_0: \mu_d = 0 \text{ vs. } H_a: \mu_d > 0$$

Note: The alternative hypothesis will be "less than" if the student defines the average difference in errors as "Charlie Brown – Lucy."

(b) Charlie Brown believes he took a good random sample but is concerned that the data may not be consistent with the assumption of normality. Why is it okay for him to move forward with his analysis even if the data is not from a normal distribution?

Solution: Since the sample size is $n = 45$, the Central Limit Theorem states that the distribution of the sample means will be normally distributed.

(c) Is the data paired or independent?

Solution: Paired. The number of errors that Charlie Brown makes and the number of errors that Lucy makes is paired by game.

(d) Compute the corresponding confidence interval for assessing the hypothesis stated in part (a) at the **α = 0.05 significance level**.

Solution: Since this is a ***one-sided "greater than" hypothesis test***, we want to construct a $100(1 - 2\alpha)$% CI. We can determine the 90% two-sided confidence interval in Minitab using the column of differences. Also, since the data is paired by game, I suggest determining the column of differences first (Lucy – Charlie Brown) and constructing the confidence interval with the differences.

Descriptive Statistics

N	Mean	StDev	SE Mean	90% CI for μ
45	7.69	18.87	2.81	(2.96, 12.41)

μ: population mean of Difference (LvP - CB)

(e) Using the 90% two-sided confidence interval that you obtained in part (d), what conclusions can Charlie Brown draw regarding his question of interest?

Solution: As the interval does not include 0, we would reject the null hypothesis and conclude that there is evidence that Charlie Brown makes fewer errors, on average, in a game compared with Lucy.

(f) Based on the 90% two-sided confidence interval that you obtained in part (d), what can we say about the p-value for addressing the hypothesis stated in part (a)?

A. The *p*-value is less than 0.05.
B. The *p*-value is equal to 0.05.
C. The *p*-value is greater than 0.05.
D. We cannot make a conclusion regarding the *p*-value relative to 0.05.

Solution: A. Since the 90% CI (constructed for a significance level of 0.05) did not include 0, we rejected the null, which means that the *p*-value must be less than the significance level.

Exercise 5: Alternative Diet from "Finding Nemo"

Despite his minor slip, Bruce, along with pals Anchor and Chum (three sharks) have decided to re-institute their fish-free diet. Dory, who has agreed to help, is investigating two types of seaweed to include in their diet. Specifically, she is curious if see if Winged Kelp is preferred to Wakame, on average. She gets a random sample of 25 sharks to agree to participate in the study. Each shark is served a clump of Wakame and a clump of Winged Kelp; before eating, the shark flips a sand-dollar. If the sand-dollar lands face-up, they eat the Wakame first, if it lands face-down, they eat the Winged Kelp first. Their level of satisfaction (on a scale of 1-10, 10 being best) is recorded. The data is summarized below.

Seaweed	N	Mean	Std. Dev.
Wakame	25	7.92	1.37
Winged Kelp	25	8.63	2.10
Diff. (Wakame – Kelp)	25	-0.71	1.125

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

Solution: Let μ_1 be the mean score for Wakame and μ_2 be the mean score for Winged Kelp; we are interested in $\mu_d = \mu_1 - \mu_2$, which is the mean difference in the scores (Wakame – Kelp). Specifically, to establish a preference for Kelp, we want to know:

$$H_0: \mu_d = 0 \text{ vs. } H_a: \mu_d < 0$$

It's okay to have the following hypothesis test as well, but it's not clear that you're testing the mean difference.

$$H_0: \mu_1 - \mu_2 = 0 \text{ vs. } H_a: \mu_1 - \mu_2 < 0$$

(b) Assuming that the differences are normally distributed, compute the test statistic appropriate for addressing the hypothesis stated in part (a). Write down *both* the formula and the value of the test statistic.

Solution: Since this is a paired analysis, we should use the following test statistic:

$$t^* = \frac{(\bar{x}_d - 0)}{\frac{s_d}{\sqrt{n}}} = \frac{-0.71 - 0}{\frac{1.125}{\sqrt{25}}} = -3.156$$

The incorrect 2-sample t test statistic is:

$$t^* = \frac{(\bar{x}_1 - \bar{x}_2) - (0)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(7.92 - 8.63) - (0)}{\sqrt{\frac{1.37^2}{25} + \frac{2.10^2}{25}}} = -1.416$$

Note: It is possible to have the positive versions if a student switched the order of the parameters in the hypothesis test above, using Kelp – Wakame.

(c) Mr. Ray, the schoolteacher, states that we didn't need to assume normality of the differences in part (b). He claims that "the test statistic in part (b) can be modeled using a normal distribution since we have a total of 50 observations randomly collected." Why is this statement invalid?

Solution: In order to use a normal distribution, we would need the ***number of differences*** to exceed 30 (or each group individually to exceed 30 if we were doing a 2-sample test).

(d) The *p*-value associated with performing the correct analysis for the hypothesis in part (a) is 0.0017. What conclusions can Dory draw at $\alpha = 0.05$ significance level? State your conclusions in context of the problem.

Solution: With a low *p*-value, we conclude that the sharks will be more content, on average, if we place them on a Winged Kelp diet.

(e) Which of the following **is** an appropriate interpretation of the *p*-value for this test?

A. The probability that the alternative hypothesis is true is 0.0017.
B. In repeated sampling, the probability of observing a test statistic less than or equal to t^* if the null hypothesis is true is 0.0017.
C. The probability that the null hypothesis is true is 0.0017.
D. In repeated sampling, the probability of observing a test statistic less than or equal to t^* if the alternative hypothesis is true is 0.0017.

Solution: B. This is the definition of a *p*-value.

Exercise 5: Lung Capacities

Long-distance runners have contended that moderate exposure to ozone increases lung capacity. To investigate this possibility, a researcher exposed 12 rats to ozone at the rate of 2 parts per million for a period of 30 days. The lung capacity of the rats was determined at the beginning of the study and again after the 30 days of ozone exposure. The lung capacities (in mL) are given here.
* From *An Introduction to Statistical Methods and Data Analysis*
By R. Lyman Ott, Micheal T. Longnecker

Rat	1	2	3	4	5	6	7	8	9	10	11	12
Before exposure	8.7	7.9	8.3	8.4	9.2	9.1	8.2	8.1	8.9	8.2	8.9	7.5
After exposure	9.4	9.8	9.9	10.3	8.9	8.8	9.8	8.2	9.4	9.9	12.2	9.3

(a) Is there sufficient evidence to support the conjecture that ozone exposure increases lung capacity in rats? Set up the null and alternative hypothesis. Make it clear what you are testing; for example, $\mu_{after} - \mu_{before,}$, or $\mu_{before} - \mu_{after}$, or μ_{diff}.

Solution: I'll let diff = "Lung Capacity After – Lung Capacity Before." Then the correct hypothesis test for this definition of diff is:

$$H_0: \mu_{diff} = 0 \text{ versus } H_a: \mu_{diff} > 0$$

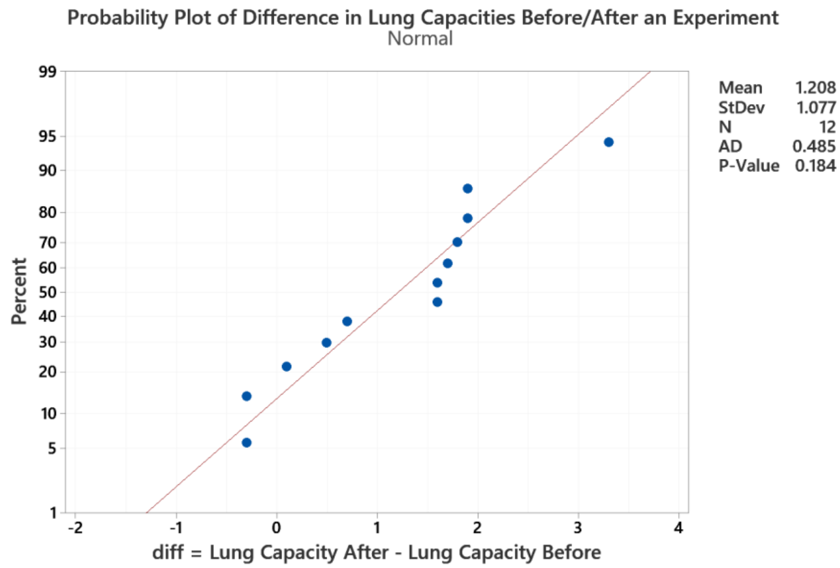
You may have written this as H_0: $\mu_{after} - \mu_{before} = 0$ versus H_a: $\mu_{after} - \mu_{before} > 0$, but it's not clear that you're testing the mean difference.

Another alternative test is H_0: $\mu_{before} - \mu_{after} = 0$ versus H_a: $\mu_{before} - \mu_{after} < 0$. The standardized test statistic for this set up will be the negative value of the one I obtain in part (b).

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

Solution: Since I'm performing a 1-sample *t*-test on diff = "Lung Capacity After – Lung Capacity

Before," I first create a column in Minitab for the differences. Since the sample size is small, we have to determine if the differences are from a normal distribution. Perform a normality test.



Since the p-value of the normality test is greater than 0.05, then we can move onto conduct the 1-sample t-test on the differences.

Test

Null hypothesis	H_0: $\mu = 0$
Alternative hypothesis	H_1: $\mu > 0$

T-Value	**P-Value**
3.89	0.001

According to Minitab, the standardized test statistic is **3.89** and its associated p-value is **0.001**.

(c) After completing the study, the researcher claims that ozone causes increased lung capacity. Is this statement supported by this experiment?

Solution: Yes … well, sort of. Since the p-value is 0.001, we can reject the null (that there is no difference) and conclude that there is evidence that exposure to ozone increases lung capacity in RATS. But what about humans? The evidence collected from the rats hints that the same may be true for humans, but from this experiment alone we can't automatically conclude that exposure to ozone increases lung capacity in humans. If you answered no, it should because of this rat/human stretch – not because of the p-value computed for the test. Also, be careful about using the word "causes" in a study of two variables. There appears to be a correlation or relationship between ozone levels and lung capacity, but we wouldn't say ozone levels *cause* increased lung capacity.

Exercise 6: Throwing vs Spitting Distance of Sunflower Seeds

Can students in my statistics classes spit or throw sunflower seeds farther? During class, take a tape measure and measure (in cm) the distance you can spit and throw a sunflower seed. Record your data (in cm) on the board. Perform a hypothesis test to determine if the average distance students can throw sunflower seeds is different than the average distance that students can spit sunflowers seeds.

Student	1	2	3	4	5	6	7
Throw Distance	398	847	968	527	421	693	704
Spit Distance	501	378	496	803	540	589	277

(a) What is the null and alternative hypothesis? Make it clear what you are testing; for example, $\mu_{after} - \mu_{before}$,, or $\mu_{before} - \mu_{after}$, or μ_{diff}.

Solution: I'll let diff = Throw – Spit distances. Then my hypotheses are:

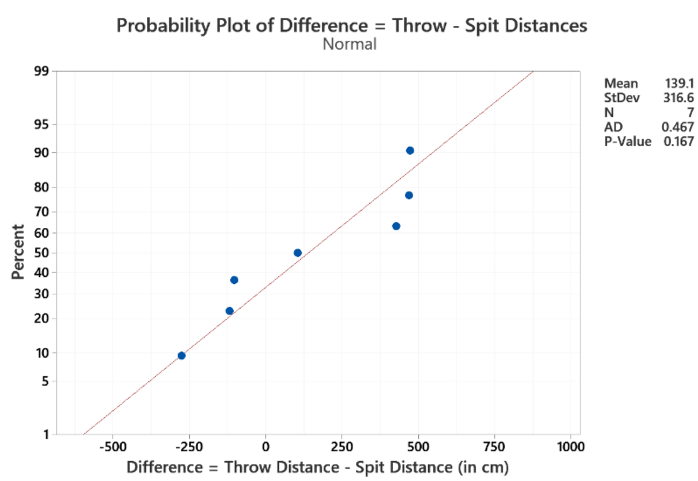
$$H_0: \mu_{diff} = 0 \text{ versus } H_a: \mu_{diff} \neq 0$$

You may have written this as H_0: $\mu_{Throw} - \mu_{Spit} = 0$ versus H_a: $\mu_{Throw} - \mu_{Spit} \neq 0$. Writing it this way makes it appear as if you are going to perform a 2-sample *t*-test instead of a 1-sample *t* test on the mean difference.

An alternative test is H_0: $\mu_{Spit} - \mu_{Throw} = 0$ versus H_a: $\mu_{Spit} - \mu_{Throw} \neq 0$. The only difference is that your standardized test statistic in part (b) will be the negative value of the one I obtain.

(b) Compute a *p*-value for this test. Based on the *p*-value, would you reject the null hypothesis?

Solution: I'm performing a 1-sample *t*-test on the difference = Throw – Spit that I created in another column. Since the sample size is small, I performed a normality test on the differences to confirm that the data was consistent with coming from a normal population.



In Minitab, we can determine the standardized test statistic and the *p*-value that accompanies it.

Test

Null hypothesis	H_0: $\mu = 0$
Alternative hypothesis	H_1: $\mu \neq 0$

T-Value	P-Value
1.16	0.289

At $\alpha = 0.05$, we do not reject the null hypothesis with a *p*-value of 0.289. Our data does not provide evidence to suggest that the mean difference is 0. In other words, we cannot conclude that the mean throwing distance and the mean spitting distance are different.

Exercise 7: Cost for Name Brand vs Store Brand Items

To investigate the amount of savings due to purchasing store brands versus name brand products, *Consumer Reports* shopped for a list of items at an A&P grocery store. One cart was filled with national brand products and the other cart was filled with the store brands of the same products (*Consumer Reports*, September 1993). The sample data obtained is:

Product, Size	***Name Brand Price***	***Store Brand Price***
Ketchup, 2 lb.	1.69	0.79
Coffee, 12 oz.	2.79	1.59
Soda, 6-pack	2.79	1.64
Paper Towels, 90 sheets	1.39	0.50
Ice Cream, 1/2 gal.	3.99	2.39
American cheese, 1 lb.	3.99	2.99
Thin spaghetti, 1 lb.	0.89	0.53
Butter, 1 lb.	2.39	1.69
White rice, 5 lb.	3.99	1.59
Vegetable oil, qt.	2.19	1.69

Develop a 95% confidence interval for the mean difference in price between name brand products and store brand products.

Solution: The name brand product price and the store brand product price are paired by food item. I defined the difference as: Diff = Name Brand $ - Store Brand $. I constructed a column of Differences in Minitab and performed a normality test on it. The normality test agrees with the Differences data being from a normal population.

We can now use a 1-sample *t* test on the Differences to construct the 95% confidence interval. I'm going to just use Minitab:

Descriptive Statistics

N	Mean	StDev	SE Mean	95% CI for μ
10	1.070	0.587	0.186	**(0.650, 1.490)**

μ: population mean of Diff = Name - Store

If we defined the Difference in the reverse order, Store Brand Price – Name Brand Price, the 95% confidence interval for the mean difference would be (-1.490, -0.650).

We are 95% certain that the interval (0.650, 1.490) contains the true mean difference in Name Brand Price – Store Brand Price. Thus, our data suggests that true mean difference is non-zero. The mean Name Brand Price does appear to be larger than the mean Store Brand Price at α = 0.05.

Exercise 8: Parallel Parking Cars with Different Turning Radii

The journal *Human Factors* (1962, pp. 375-380) reports a study in which n = 14 subjects were asked to parallel park two cars having very different wheelbases and turning radii. The time in seconds for each subject to park the two different autos was recorded and is given below.

Person	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Auto 1	37.0	35.8	16.2	24.2	22.0	33.4	25.8	58.2	33.6	24.4	23.4	21.2	36.2	29.8
Auto 2	17.8	20.2	16.8	41.4	21.4	38.4	16.8	32.2	27.8	23.2	29.6	20.6	32.2	53.8

(a) Let **d_i = time for person i to park auto 1 – time for person i to park auto 2** for each person i. Set up the null and alternative hypothesis for testing whether the true mean difference μ_d in parking times for the two autos is zero.

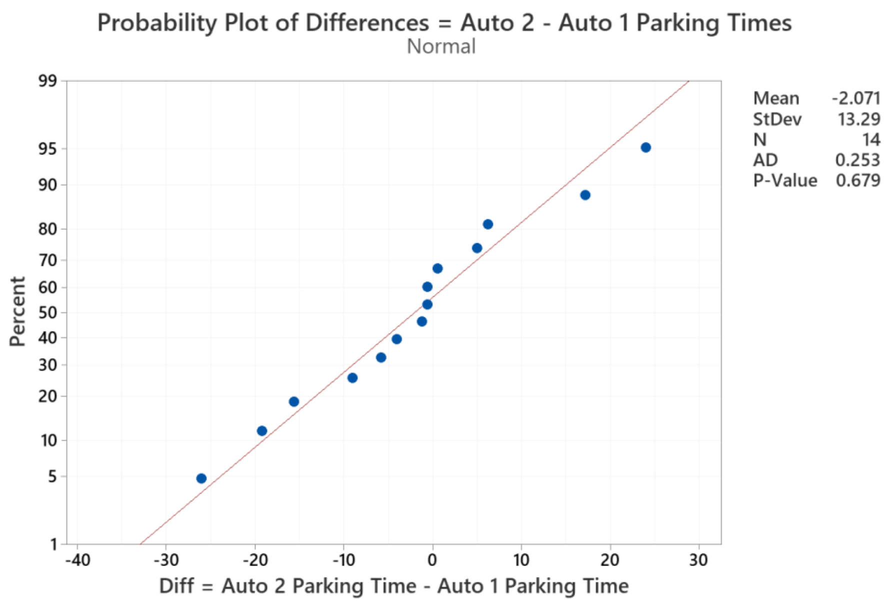
Solution: H_0: $\mu_d = 0$ versus H_a: $\mu_d \neq 0$

(b) Determine the appropriate standardized test statistic for your hypothesis test.

Solution: The data is paired by person. In order to determine the test statistic, I'll first compute the differences in parking times as "Auto 2 – Auto 1" for each person i, where i = 1, 2, ..., 12. In order to use a 1-sample t-test on the differences, we need to first check if they are from a normal distribution. At α = 0.05, it appears that the differences are from a normal population.

The standardized test statistic is:

$$t_0 = \frac{-2.071 - 0}{\frac{13.29}{\sqrt{14}}} \cong -0.58.$$



(c) In Minitab, determine the approximate *p*-value for your test statistic.

Solution: In Minitab, the *p*-value is given as 0.570.

Test

Null hypothesis	H_0: $\mu = 0$
Alternative hypothesis	H_1: $\mu \neq 0$

T-Value	**P-Value**
-0.58	0.570

(d) At level of significance α = 0.05, would you REJECT or NOT REJECT that the true mean difference in parking times is zero?

Solution: At α = 0.05, we would NOT reject that the true mean difference in parking times is 0 since the *p*-value for this hypothesis test is much greater than 0.05.