Lesson 4: Inference for a Difference in Two Means

Corresponding Workbook Module: 5

Example 4.1 *Hypothesis testing for a paired (dependent) sample*
Mintaek wants to know if there is a difference between population mean Exam 1 scores and population mean Exam 2 scores of his students. He randomly selected 7 students across his classes and collected their Exam 1 and Exam 2 scores. Use the appropriate statistical methods and set up the hypotheses that best match Mintaek's interests.

Student	1	2	3	4	5	6	7
Exam 1	77	70	86	72	92	62	69
Exam 2	81	79	83	82	98	60	76

You first need to recognize that this is a paired sample. It is because each observation in Exam 1 can be paired with one in Exam 2 since both scores came from the same student. In general, you can say that you have a paired data when you can match each individual observation from one group to another from the other group (like in matched pairs experiment or matched pairs observational study). You can match them when they have something in common. For example, when you have two observations from a same person or two people with similar characteristics.

We use the paired t-test for this problem since we have a paired data. When we are working with the paired data, we only concern ourselves with the differences between two exam scores within each pair. In fact, we will not even be considering the original data anymore.

Student	1	2	3	4	5	6	7
Difference	-4	-9	3	-10	-6	2	-7

Using software, we find $\bar{x}_D = -4.43857$ (sample mean of differences), $s_D = 5.12696$ (sample standard deviation of differences), and $n = 7$ (sample size). Note that the sample size is obtained by counting the number of pairs for the paired t-test, NOT the total number of test scores.

Step 1: Since the problem did not explicitly specify how much of a difference we want to see, any difference larger than 0 would suffice for our testing. In this case, any significant difference will do.

- H_0: $\mu_D = 0$, population mean difference in exam scores is **equal to** 0.

- H_A: $\mu_D \neq 0$, population mean difference in exam scores is **different from** 0.

Step 2: We are using the matched paires t-test here (matched pairs t-test and paired t-test mean the same thing). We check all the underlying conditions (See workbook page 149 then 91 for the description of conditions). For the significance level, we will use $\alpha = 0.05$.

We check the following conditions for the matched paires t-test for means:

- Since $n = 7 < 16$, we need to make sure that the distribution of differences appear to have come from normal populations. We can draw a histogram of differences to check this condition.

Histogram of Differences
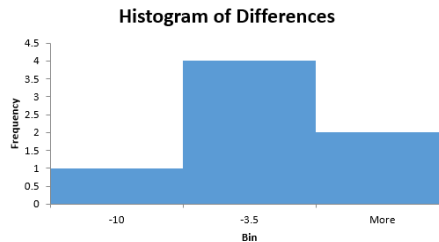
We see nothing to be concerned here. So we conclude that the distribution of differences appear to have come from normal populations.

- Each pair was selected randomly since all 7 students were selected randomly.

- Since there are more than 70 ($= 10 \times n$) students taking Mintaek's class (in fact, there are 116 this semester), the population size (N) is larger than 10 times the sample size.

Since all conditions are met, we can use the matched paires t-test to make an inference on μ_D (population difference between two exam scores).

Step 3: We can now find the t-test statistic value and corresponding p-values.

For $df = 7 - 1 = 6$, we find

$$t = \frac{\bar{x}_D - \mu_D}{s_D/\sqrt{n}} = \frac{-4.42857 - 0}{5.12696/\sqrt{7}} = \frac{-4.42857}{1.93781} = -2.29$$

where μ_D is the hypothesized population mean difference from the hypotheses (or "the null hypothesis value").

At the t-distribution table, look at the row corresponding to $df = 6$. We can see that the positive value of our t-test statistic $t = 2.29$ is between 1.943 and 2.447. Note that I am using the positive value of t. It is because the t-distribution is symmetric. It means the tail area to the right of $t = 2.29$ (upper tail area) is going to be the same as the tail area to the left of $t = -2.29$ (lower tail area).

Since the upper tail probability p corresponding to $t = 1.943$ and $t = 2.447$ are 0.05 and 0.025, respectively, we can guess that the **upper** tail probability p corresponding to our t-test statistic $t = 2.29$ is between 0.05 and 0.025. Therefore, the **lower** tail probability p corresponding to our t-test statistic $t = -2.29$ is also between 0.05 and 0.025.

Since this is a two-sided test, our P-value would be between 2×0.05 and 2×0.025, that is 0.1 and 0.05. We can express our P-value as: $0.05 <$ P-value < 0.1.

For homework, you can use `surfstat.anu.edu.au/surfstat-home/tables/t.php` to find the exact P-value. Make sure to choose the correct graph corresponding to the tail you are interested in. Based on the output shown on the next page, we find the exact P-value of 0.0619. See the last page of Handout 2 for more information. Note that you will not be able to use this website for the exam, so it is important that you understand the process explained in the previous page.

SurfStat *t-distribution* calculator
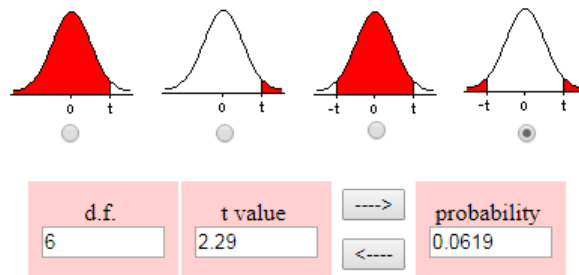
d.f.	t value	---->	probability
6	2.29	<----	0.0619

Step 4: We fail to reject the null hypothesis under the significance level $\alpha = 0.05$. It is because our P-value was larger than our pre-determined significance level (0.05). We do not have a significant evidence to suggest that the difference in exam scores is not 0.

Interpretation of the p-value: The probability of obtaining a sample mean difference this far (or farther) away from 0 is 0.0619 if the population mean difference was 0. That is, there is 6.19% chance of obtaining a sample mean difference of larger than 4.42857 OR smaller than -4.42857 if the true mean difference was 0 ($= H_0$ was true). Since the chance is greater than $\alpha = 0.05$, this would not be a surprising enough result to reject H_0.

IF THIS WAS A ONE-SIDED TEST WHERE $H_A : \mu_D < 0$

Step 1: H_0 is the same, just H_A needs to be changed.

- H_0: $\mu_D = 0$, population mean difference in exam scores is 0.

- H_A: $\mu_D < 0$, population mean difference in exam scores is less than 0. In other words, since $\mu_D = \mu_1 - \mu_2 < 0 \implies \mu_1 < \mu_2$, students did better in Exam 2 than Exam 1.

Step 2: Same

Step 3: Since $t = -2.29$ for $df = 6$, we find the **lower** tail area to be between 0.05 and 0.025. We are only interested in the lower tail since the sign in the H_A is $<$. We do not consider the upper tail area. Therefore, our P-value would just be $0.025 <$ P-value < 0.05.

Similarly, for the website, choose the second graph from the left instead to find the area of only one tail. You should find the exact P-value of 0.031.

Step 4: We reject the null hypothesis under the significance level $\alpha = 0.05$. It is because our P-value was smaller than our pre-determined significance level. We have significant evidence to suggest that the difference in exam scores is less than 0. That is, the average exam 1 score is lower than the average exam 2 score, because $\mu_D = \mu_{exam1} - \mu_{exam2} < 0$ (the alternative hypothesis) would imply $\mu_{exam1} < \mu_{exam2}$.

Interpretation of the p-value: The probability of obtaining a sample mean difference of -4.42857 or lower is 0.031 if the population mean difference was 0. That is, there is 3.1% chance of obtaining a sample mean difference of smaller than -4.42857 if the true mean difference was 0 ($= H_0$ was true). Since the chance is less than $\alpha = 0.05$, this would be a surprising enough result to reject H_0.

Example 4.2 *Confidence intervals for paired (dependent) samples*
Mintaek now wants to estimate the true (or population) difference between the average exam 1 scores and the average exam 2 scores with the 95% certainty.

Technically, you still would need to check all conditions for the paired t-confidence intervals. However, you may notice that they are identical to ones for the matched pairs t-test. It will be omitted here for the sake of space.

We first find the standard error: $SE = \dfrac{s_D}{\sqrt{n}} = \dfrac{5.12696}{\sqrt{7}} = 1.93781$.

We need to find the t critical value, t^*, to find the margin of error. Note that $df = 6$ and confidence level is 95%. So we look at the row corresponding to $df = 6$ on the t-distribution table. We then find a column corresponding to the confidence level 95% from the last row. You should find $t^* = 2.447$ as appropriate critical value.

When you want to find t^* for df's not listed on the t-distribution table, such as $df = 65$, you can use `surfstat.anu.edu.au/surfstat-home/tables/t.php` to find the t^*. Make sure to choose the third graph from the left to find the t critical values. Based on the output shown, we find the 2.447 as our t^* also. To get this value, enter 6 under d.f. and 0.95 under probability, then hit the <- - - - button

SurfStat *t-distribution calculator*

d.f.	t value	---->	probability
6	2.447	<----	0.95
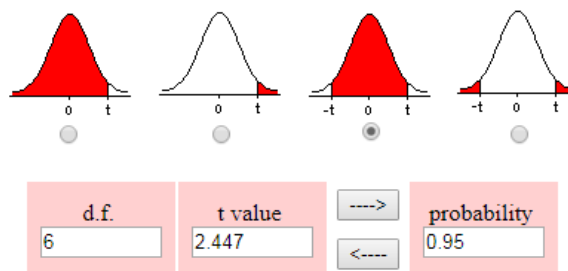
Margin of error is defined as: $t^* \times SE = t^*\dfrac{s_D}{\sqrt{n}} = 2.447 \times 1.93781 = 4.742$. Then, we can finally calculate the 95% confidence interval for the true (or population) difference between the average exam 1 scores and the average exam 2 scores.

$$\bar{x}_D \pm t^*\frac{s_D}{\sqrt{n}} = -4.43857 \pm 2.447 \times 1.93781 = -4.439 \pm 4.742 = (-9.18, 0.30)$$

This means that we are 95% certain/sure/confident that the interval (-9.18, 0.30) contains μ_D, population mean difference between Exam 1 scores and Exam 2 scores.

Example 4.3 *Hypothesis testing for independent (not paired) samples*
Mintaek is interested in finding out whether the average wingspan of Canada geese in the Julia Davis Park is the same with the average wingspan of Canada geese in the Ann Morrison Park. Mintaek captured 11 geese in the Julia Davis Park and another 11 geese in the Ann Morrison Park and measured wingspans of each one of them.

| Julia Davis Park | 56 | 57 | 57 | 58 | 59 | 59 | 60 | 60 | 61 | 62 | 63 |
| Ann Morrison Park | 57 | 58 | 58 | 59 | 60 | 61 | 63 | 64 | 64 | 65 | 65 |

We can find the summary information for the data by using software.

Julia Davis Park	$n_1 = 11$	$\bar{x}_1 = 59.272727$	$s_1 = 4.818181$
Ann Morrison Park	$n_2 = 11$	$\bar{x}_2 = 61.272727$	$s_2 = 9.218181$

You first need to recognize that this is **NOT** a paired sample, and they are two independent samples. Although we would expect geese from two parks to be similar, you can't match one geese captured from the Julia Davis Park with one from the Ann Morrison Park. In that case, we need to use the two-sample t-test.

Step 1: Since the problem did not explicitly specify how much of a difference we want to see, any difference larger than 0 would suffice here. Admittedly, μ_0 (null hypothesis value) would be 0 for most two sample problems.

- H_0: $\mu_1 = \mu_2$ (or $\mu_1 - \mu_2 = 0$), population mean wingspan of geese in the Julia Davis Park is equal to the population mean wingspan of geese in the Ann Morrison Park.

- H_A: $\mu_1 \neq \mu_2$ (or $\mu_1 - \mu_2 \neq 0$), population mean wingspan of geese in the Julia Davis Park is not equal to the population mean wingspan of geese in the Ann Morrison Park.

Note that I let the geese in Julia Davis Park to be the Group 1 and geese in Ann Morrison Park to be Group 2. You can number your group the other way as long as you are consistent.

Step 2: We are using the two-sample t-test here because we have two samples that are independent (not individually paired). We check all the underlying conditions (See workbook page 152 for description of conditions). For the significance level, I will continue to use $\alpha = 0.05$. Unless otherwise stated, you should use $\alpha = 0.05$ for this class.

We check the following conditions for the two-sample t-test for means:

- Since $n_1 + n_2 = 22 \geq 16$, we need to make sure that both sample graphs are not too skewed with no outliers. We can draw histograms of two samples.

  Looking at the histograms shown on the next page, I have no concerns here and conclude that both sample graphs are not too skewed and have no outliers.

- Since we are not given whether geese were randomly captured, we would need to assume that geese in both parks were captured randomly.
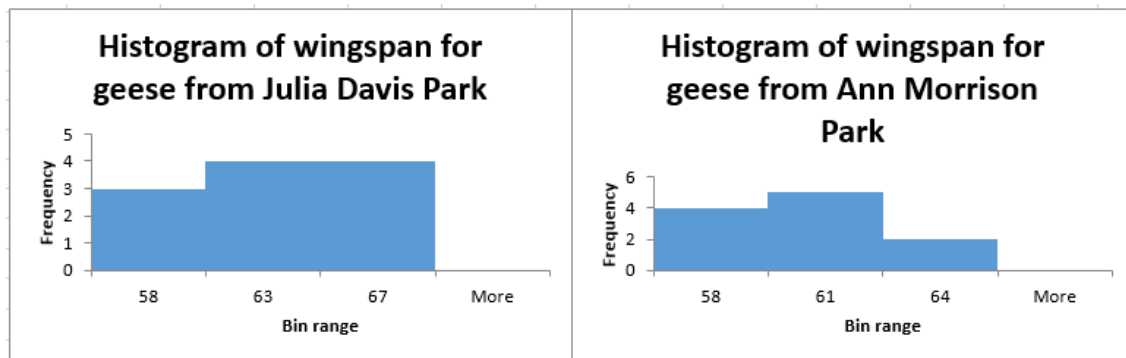
- Population is certainly larger than 10 times the sample size. It is clear that there are more than 220 ($= 10(n_1 + n_2) = 10(11 + 11) = 10 \times 22$) geese in Boise.

Since all conditions are met, we can use the two-sample t-test to make an inference on difference between μ_1 and μ_2.

Step 3: We can now find the t-test statistic value and corresponding p-values.

$$ t = \frac{(\bar{x}_1 - \bar{x}_2) - \mu_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(59.272727 - 61.272727) - 0}{\sqrt{\frac{4.818181^2}{11} + \frac{9.218181^2}{11}}} = \frac{-2}{3.136149} = -1.77 $$

where μ_0 (or $\mu_1 - \mu_2$) is the hypothesized population mean difference from the hypotheses (or "the null hypothesis value").

We find the *df* for the two-sample t-test by taking $n_1 - 1$ or $n_2 - 1$, whichever is smaller. Here, we have $n_1 - 1 = 10$ and $n_2 - 1 = 10$, so the *df* would be 10. Be careful, n_1 and n_2 can be different on other problems.

At the t-distribution table, look at the row corresponding to *df* $= 10$. We can see that the positive value of our t-test statistic $t = 1.77$ is between 1.372 and 1.812. Note that I am using the positive value of t because t-distribution is symmetric. It means the tail area to the right of $t = 1.77$ is going to be the same as the tail area to the left of $t = -1.77$.

Since the upper tail probability p corresponding to $t = 1.372$ and $t = 1.812$ are 0.10 and 0.05, respectively, we can guess that the **upper** tail probability p corresponding to our t-test statistic $t = 1.77$ is between 0.10 and 0.05. Therefore, the **lower** tail probability p corresponding to our t-test statistic $t = -1.77$ is also between 0.10 and 0.05.

Since this is a two-sided test, our P-value would be between 2×0.10 and 2×0.05, that is 0.2 and 0.1. We can express our p-value as: $0.1 <$ P-value < 0.2.

For homework, you can use `surfstat.anu.edu.au/surfstat-home/tables/t.php` to find the exact P-value. Make sure to choose the correct graph corresponding to the tail you are interested in. The software gave me the exact P-value of 0.1072 for the two-sided tests.

Step 4: We fail to reject the null hypothesis. It is because our P-value was larger than our pre-determined significance level $\alpha = 0.05$. We do not have a significant evidence to suggest that the population mean wingspan of geese in the Julia Davis Park is not equal to the population mean wingspan of geese in the Ann Morrison Park.

Interpretation of the p-value: The probability of obtaining a sample mean difference this far (or farther) away from 0 is 0.1072 if the population mean difference was 0.

IF THIS WAS A ONE-SIDED TEST WHERE $\mathbf{H_A} : \boldsymbol{\mu_1} < \boldsymbol{\mu_2}$ OR $\boldsymbol{\mu_1} - \boldsymbol{\mu_2} < \mathbf{0}$

Step 1: H_0 is the same, just H_A needs to be changed.

- H_0: $\mu_1 = \mu_2$ (or $\mu_1 - \mu_2 = 0$), population mean wingspan of geese in the Julia Davis Park is equal to the population mean wingspan of geese in the Ann Morrison Park.

- H_A: $\mu_1 < \mu_2$ (or $\mu_1 - \mu_2 < 0$), population mean wingspan of geese in the Julia Davis Park is less than the population mean wingspan of geese in the Ann Morrison Park.

Step 2: Same

Step 3: Since $t = -1.77$ for $df = 10$, we find the **lower** tail area of between 0.10 and 0.05. We are only interested in the lower tail. We do not consider the upper tail area. Therefore, our P-value would just be $0.05 < $ p-value < 0.10.

Similarly, for the website, choose the second graph from the left instead to find the area of only one tail. You should find the exact P-value of 0.0536.

Step 4: We reject the null hypothesis. It is because our P-value was smaller than our pre-determined significance level $\alpha = 0.05$. We have significant evidence to suggest that the population mean wingspan of geese in the Julia Davis Park is **less than** the population mean wingspan of geese in the Ann Morrison Park.

Interpretation of the p-value: The probability of obtaining a sample mean difference of -2 ($= \bar{x}_1 - \bar{x}_2$) or lower is 0.0536 if the population mean difference was 0.

Example 4.4 *Confidence intervals for independent (not paired) samples*
Mintaek now wants to estimate the true (or population) difference between the population mean wingspan of geese in the Julia Davis Park and the population mean wingspan of geese in the Ann Morrison Park with the 95% certainty.

Technically, you still would need to check all conditions for the two-sample t-confidence intervals. However, you may notice that they are identical to ones for the two-sample t-test. It will be omitted here for the sake of space.

We first find the standard error: $SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{4.818181^2}{11} + \frac{9.218181^2}{11}} = 3.136149$.

We need to find the t critical value, t^*, to find the margin of error. Note that $df = 15$ and confidence level is 95%. So we look at the row corresponding to $df = 10$ on the t-distribution table. We then find a column corresponding to the confidence level 95% from the last row. We should find $t^* = 2.228$ as appropriate critical value.

Margin of error is defined as: $t^* \times SE = t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = 2.228 \times 3.136140 = 6.987$. Then we can finally calculate the 95% confidence interval for the true (or population) difference between the population mean wingspan of geese in the Julia Davis Park and the population mean wingspan of geese in the Ann Morrison Park with the 95% certainty.

$$
\begin{aligned}
(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} &= -2 \pm 2.228 \times 3.136140 \\
&= -2 \pm 6.987 \\
&= (-2 - 6.987, -2 + 6.987) \\
&= (-8.99, 4.99)
\end{aligned}
$$

This means that we are 95% certain/sure/confident that the interval (-8.99, 4.99) contains the difference between the population mean wingspan of geese in the Julia Davis Park and the population mean wingspan of geese in the Ann Morrison Park (which is just $(\mu_1 - \mu_2)$ in statistical terms).