Matched-pair hypothesis testing

We've recently been looking at hypothesis testing for the difference of means when we take two independent samples from one or two populations. Technically, we say that we have **independent samples** when there's no relationship between the observations we find for each sample.

But sometimes we'll want to run a hypothesis test on the difference of means between **dependent samples**, which are samples for which the observations from one sample are related to an observation from the other sample.

Matched-pair tests

When we do hypothesis testing with dependent samples, we often call it a **matched-pair test**, because each subject in the second sample matches with a particular subject in the first sample.

It's common to run a matched-pair test that compares some new technique or method to an old one, or looks at a before-and-after change.

For instance, a weight-loss study could define Population 1 as the set of starting weights for each participant, and Population 2 as the set of ending weights for each participant. Each participants starting and ending weights (from Populations 1 and 2, respectively) form a matched-pair for that individual.

In this example, there's an advantage to using a matched-pair test, instead of a difference of means test with independent samples. If we took the independent samples approach, sample 1 could be taken from the population before the weight loss study begins, and sample 2 could be taken from the population after the weight loss study ends. This approach introduces extra variability unnecessarily because we'll get different people in both samples.

But if we take the matched-pair approach, we keep the people the same across both samples, creating a matched-pair of each person's starting and ending weights.

In general, hypothesis testing with dependent samples will follow a really similar process as the one we've used for the difference of means with independent samples, except that we'll create one variable as the difference between the two samples, and we'll perform the hypothesis test with just this one variable, instead of with two variables.

Let's work through an example so that we can see how to use dependent samples in a matched-pair hypothesis test.

Example

A fast food restaurant is implementing new workplace policies with the goal of increasing employee satisfaction by 2 points on a scale of 1 to 10. The restaurant surveys 10 employees, asking them both before and after the policies are enacted to rate their workplace satisfaction on the $1-10$ scale, and records the results in the table below.

Employee	1	2	3	4	5	6	7	8	9	10
Before x_1	3	3	5	7	1	0	2	6	6	5
After x_2	3	6	9	7	3	5	5	5	9	9
Difference, d	0	3	4	0	2	5	3	-1	3	4
d^2	0	9	16	0	4	25	9	1	9	16

Can the restaurant say at 5% significance that the policies increased employee satisfaction by 2 points?

The restaurant will define the "before" responses as Population 1, and the "after" responses as Population 2. The samples are dependent because it's reasonable to see how an employee's "after" response could be affected by their "before" response.

Then their null and alternative hypotheses will be

$$H_0: \mu_2 - \mu_1 \leq 2$$

$$H_a: \mu_2 - \mu_1 > 2$$

where μ_1 is the mean employee satisfaction before the new workplace policies are implemented, and μ_2 is the mean employee satisfaction after the new workplace policies are implemented. And because $\mu_2 - \mu_1$ is the difference in employee ratings, the hypothesis statements could also be written as

$$H_0: \mu_d \leq 2$$

$$H_a : \mu_d > 2$$

where μ_d is the mean difference between the two populations.

To find the mean difference, we'll sum the differences and divide by the number of matched-pairs in our sample, $n = 10$.

$$\bar{d} = \frac{\sum_{i=1}^{n} d_i}{n} = \frac{0+3+4+0+2+5+3+(-1)+3+4}{10} = \frac{23}{10} = 2.3$$

So the sample mean tells us that employee satisfaction increases by about 2.3 on a scale of 1 to 10. Then the sample standard deviation is

$$s_d = \sqrt{\frac{\sum_{i=1}^{n} (d_i - \bar{d})^2}{n-1}}$$

To calculate this, we'll first find

$$\sum_{i=1}^{n} (d_i - \bar{d})^2$$

$$(0-2.3)^2 + (3-2.3)^2 + (4-2.3)^2 + (0-2.3)^2 + (2-2.3)^2$$

$$+(5-2.3)^2 + (3-2.3)^2 + (-1-2.3)^2 + (3-2.3)^2 + (4-2.3)^2$$

$$(-2.3)^2 + 0.7^2 + 1.7^2 + (-2.3)^2 + (-0.3)^2 + 2.7^2 + 0.7^2 + (-3.3)^2 + 0.7^2 + 1.7^2$$

$$5.29 + 0.49 + 2.89 + 5.29 + 0.09 + 7.29 + 0.49 + 10.89 + 0.49 + 2.89$$

$$36.1$$

Then the sample standard deviation is

$$s_d = \sqrt{\frac{36.1}{9}}$$

$$s_d \approx \sqrt{4.011}$$

$$s_d \approx 2.003$$

Because the population standard deviations are unknown, and/or because both sample sizes are small, n_1, $n_2 < 30$, the test statistic will be

$$t = \frac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{n}}}$$

$$t \approx \frac{2.3 - 2}{\frac{2.003}{\sqrt{10}}}$$

$$t \approx 0.3 \cdot \frac{\sqrt{10}}{2.003}$$

$$t \approx 0.474$$

and the degrees of freedom are

$$\text{df} = n - 1 = 10 - 1 = 9$$

At a significance level of 5% (a confidence level of 95%) for an upper-tail test, and df $= 9$, the t-table gives 1.833.

	Upper-tail probability p									
df	**0.25**	**0.20**	**0.15**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**	**0.001**	**0.0005**
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.703	0.883	1.100	1.383	**1.833**	2.262	2.821	3.250	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
	50%	**60%**	**70%**	**80%**	**90%**	**95%**	**98%**	**99%**	**99.8%**	**99.9%**
	Confidence level C									

The restaurant's t-test statistic $t \approx 0.474$ doesn't meet the threshold $t = 1.833$, so the critical value approach tells them that they can't reject the null hypothesis, and therefore can't conclude that the new workplace policies increased employee satisfaction by 2 points.

Confidence intervals for matched-pair tests

If the restaurant from the previous example had known the population standard deviation σ_d, they could have calculated a confidence interval around the difference $\bar{d}$ using

$$(a, b) = \bar{d} \pm z_{\alpha/2}\sigma_{\bar{d}}$$

$$(a, b) = \bar{d} \pm z_{\alpha/2}\frac{\sigma_d}{\sqrt{n}}$$

If, instead, the restaurant had an unknown population standard deviation σ_d and/or a small sample $n < 30$, to find a confidence interval around the difference $\bar{d}$ they would have used

$$(a,b) = \bar{d} \pm t_{\alpha/2}\frac{s_d}{\sqrt{n}} \text{ with df} = n - 1$$

Let's continue with the previous example in order to calculate the confidence interval.

Example (cont'd)

Find a 95% confidence interval around $\bar{d}$ using the information in the previous example.

From the previous example, we see that population standard deviation σ_d is unknown, and we have a small sample $n = 10 < 30$, so we'll calculate the confidence interval as

$$(a,b) = \bar{d} \pm t_{\alpha/2}\frac{s_d}{\sqrt{n}}$$

$$(a,b) \approx 2.3 \pm 2.262 \cdot \frac{2.003}{\sqrt{10}}$$

$$(a,b) \approx 2.3 \pm 1.433$$

So the margin of error is 1.433 and the confidence interval is

$$(a,b) \approx (2.3 - 1.433, 2.3 + 1.433)$$

$$(a,b) \approx (0.867, 3.733)$$

Therefore, there's a 95% chance that the change in employee satisfaction changes between 0.867 points and 3.733 points.