Inferences from Two Means

Matched Pairs and Independent Samples

We've seen that trying to compare two different qualitative samples isn't all that different from trying to draw conclusions about any one sample. If we want to estimate a parameter we use a confidence interval; if we want to test a claim we use a test of hypotheses. We need to calculate things slightly differently, but the processes are basically the same.

This is still true when the samples are quantitative, but we need an additional distinction. If the experiment is a matched pairs experiment, then the calculations are going to feel very familiar. If the samples cannot be matched because they're independent then we need yet another formula for the standard error and the test statistic.

DEPENDENT MEANS: MATCHED PAIRS EXPERIMENTS

Recall that a matched pairs experiment is a study with a design where observations from different samples belong together. The simplest example of such an experiment is a comparison between an observation before and after a treatment. If we take all the "before" data and mush it all together into one measure of centre, and the same with the "after" data, we lose the connection between the data points.

When using data from a matched pairs experiment, what we're most likely looking for is the difference between the after and before observations. In a weight loss trial, for example, we're far less interested in the starting weight, or even the final weight; we want to look at the amount of weight lost. We want to know whether that number is 0, representing no effect, or not 0, suggesting the weight loss treatment had an effect, good or bad.

The procedure for evaluating data like this really isn't two samples. Each subject generates two numbers (such as before and after), but in the end those two numbers come down to one number, their difference. The textbook calls the difference d, but otherwise it's the same calculations as we'd do for one mean:

CI: $\bar{d} \pm t_{\alpha/2} \cdot \dfrac{s_d}{\sqrt{n}}$ Test Statistic: $t = \dfrac{\bar{d} - \mu_d}{s_d / \sqrt{n}}$

The claimed population mean of the differences (μ_d) and the standard deviation of the differences (s_d) replace the usual symbols for those concepts, but otherwise the formulas and steps are identical to those in problems you've already done.

Example 1: Students were given a vocabulary test of 100 random scientific terms, then they attended a workshop on recognizing Greek and Latin roots, and were tested again with another 100 random words. The scores of the students for each test are listed in the table on the next page. Estimate the mean number of words a student taking this workshop can expect to improve by as a 90% confidence interval.

Authored by Darren Rigby

	A	B	C	D	E	F	G	H	I	J	K	L	M
Test 1	38	47	22	56	71	30	17	59	92	56	48	29	37
Test 2	55	68	36	77	83	48	34	85	96	53	63	45	56

Solution: Imagine that the testing department for the vocabulary program hadn't sent you the data table above, but instead just found the differences themselves, and submitted that data for you to build a confidence interval for:

	A	B	C	D	E	F	G	H	I	J	K	L	M
Test 1	38	47	22	56	71	30	17	59	92	56	48	29	37
Test 2	55	68	36	77	83	48	34	85	96	53	63	45	56
Δ	17	21	14	21	12	18	17	26	4	−3	15	16	19

(Notice that we've been asked to report the amount of improvement, and student J did not improve. The negative score reflects that. In cases like this we always subtract "after − before".)

If these new numbers were all the data we had, we'd find the estimate and standard deviation just like always, and that's the right procedure here:

$$n = 13;\ \bar{d} = 15.154\ldots;\ s_d = 7.5371\ldots;\ \text{d.f.} = 12;\ t_{0.05} = 1.782;\ t_{0.05} \cdot \frac{s_d}{\sqrt{n}} = 3.7251\ldots$$

Therefore the confidence interval is 15.154 ± 3.725 = (11.4, 18.9).

INDEPENDENT MEANS: SEPARATE SAMPLES

In most studies with two samples it isn't reasonable to pair up observations from those samples. If we're testing whether men or women have a greater tolerance for pain, there's no reason to pair any particular man with any particular woman. The samples might not even be the same size, making pairs impossible.

In such cases, we need another method of calculating the standard error, and from there, margin of error for a confidence interval and the test statistic for a test of hypotheses:

standard error: $\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$ CI: $(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$ $t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$

There are two methods of calculating the degrees of freedom for determining which row of Table A-3 to use. The more precise one is used by statistics software and is long and tedious to do by hand. The simpler method, sometimes called the conservative method, is to take n − 1 as usual, using the smaller of the two values of n.

Again, all the same steps and procedures apply for estimating or testing a claim about independent samples.

Example 2: Two models of cell phone are tested for overheating. Each phone has the same resource-intensive apps installed on it, they're left running for an hour, and their final temperature is recorded. For Model A, $\bar{x}$ = 50.3°C, s = 8.7°C, and n = 74, and for Model B, $\bar{x}$ = 57.3°C, s = 9.2°C, and n = 86. Estimate the difference in temperature between the two models, B minus A, with a 95% confidence interval.

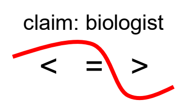
Solution: The estimate of the parameter is 57.3°C − 50.3°C = 7.0°C.

For the critical t value, we need the degrees of freedom. The smaller sample size is $n_A = 74$, so the conservative degrees of freedom is 73, which becomes row 70 in Table A-3. For a 95% CL, $t_{0.025}$ is 1.994.

The standard error comes out to 1.4167…. Putting it all together, the margin of error is 1.4167 · 1.994 = 2.8249…, and the confidence interval is 7.0 ± 2.8 or (4.2, 9.8).

Example 3: A biologist wants to know whether raccoons who live in the city have longer lifespans than those who live in rural areas. She tags raccoon kits (the young of a raccoon) of a certain size so their ages are known, and their lifespans are tracked. For the 83 urban raccoons, the mean lifespan was 3.2 years with a standard deviation of 0.8 years, and for the 61 rural raccoons, the mean lifespan was 2.7 years with a standard deviation of 0.4 years. Test the biologist's claim that raccoons in the city live longer at the 5% level.

Solution: We need a test of hypotheses. The null hypothesis is that the lifespans are equal. The biologist's claim is that the lifespans of urban raccoons are strictly greater than those of rural raccoons, so the test is a one-tailed test with the claim being the alternative hypothesis:



H_0: $\mu_u = \mu_r$
H_a: $\mu_u > \mu_r$

For the critical t value, we need the degrees of freedom. The smaller sample size is $n_r = 61$, so the conservative degrees of freedom is 60. For a one-tailed test with $\alpha = 0.05$, $t_{0.05}$ is 1.671.

Now we just need to calculate the test statistic.

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(3.2 - 2.7) - 0}{\sqrt{\frac{0.8^2}{83} + \frac{0.4^2}{61}}} = 4.9185\ldots$$

The test statistic is much farther away from the mean than the critical value is; the result is in the rejection zone. The test is therefore significant. We would say that we have sufficient evidence to support the claim that urban raccoons lead longer lives than rural ones.

EXERCISES

A. For Example 1, the author went to an online calculator to get the standard deviation of the test score improvements. It also gave confidence intervals for various CLs:

90%, $1.645s_{\bar{x}}$ = 15.1538 ±3.439; 95%, $1.960s_{\bar{x}}$ = 15.1538 ±4.097;
99%, $2.576s_{\bar{x}}$ = 15.1538 ±5.385

The results don't match the calculation provided in this worksheet. What is the website doing wrong?

B. It's well-known in swimming that the newer full-body technical swimsuits decrease drag and improve times, but do they do so equally for everyone? A researcher reads a report that in elite athletes the suits reduce times by 5%. The qualifications for being on the Canadian swim team for the 100 m freestyle is under 51 seconds, so she estimates this is an improvement of no more than 2.5 seconds for this event.

She finds swim clubs across Canada and finds adults who have only been swimming for 5 years or less. She has them do 100-m freestyle swims with and without the full-body suits on consecutive days. The results are in the table below. Test the claim that the suits reduce times for the 100 m freestyle for beginner swimmers by more than 2.5 seconds. Use $\alpha = 0.05$. Assume all requirements for the test are met.

	A	B	C	D	E	F	G	H
Tech suit	97.8	85.7	83.4	95.4	82.5	91.2	68.3	85.4
Reg. suit	101.2	89.1	86.4	98.9	85.2	93.0	70.9	86.8

C. Is the organic waste addition to the recycling program working for residents (as opposed to businesses)? Before the ban on food scraps as garbage began in January 2015, and again in November 2019, households across the Lower Mainland were randomly selected and their garbage output tracked. For the 252 households selected in 2014, the mean weight of the garbage was 2.78 kg ($s = 1.52$ kg), and for the 312 households selected in 2019, the mean weight was 1.82 kg ($s = 1.61$ kg).

1) The think tank that conducted the study published a 90% confidence interval for this data using the techniques described in this worksheet. What did they report?

2) The think tank came under harsh criticism over their technique in producing their report—the distribution of the weights in 2019 cannot be normal, otherwise approximately 13% of all households would be producing negative amounts of garbage. Does the criticism hold merit, and why?

D. The "lazy keto" diet severely restricts the amount of carbohydrates consumed, just like the regular keto diet does, but doesn't require any other specification on what goes into the diet, whereas the regular keto diet regulates all dietary intake.

A dietician has 51 clients on the regular keto diet, who lost an average of 7.8 kg in their first three weeks ($s = 2.2$ kg), and 13 clients on the lazy keto diet, who lost an average of 5.4 kg in their first three weeks ($s = 1.8$ kg). Test the claim that the lazy keto diet is at least as effective at rapid weight loss as the regular diet at $\alpha = 0.05$.

SOLUTIONS

A: By now, I'm sure you've seen 1.645, 1.96 and 2.57[5] often enough to recognize them as the critical z scores for those CLs. The sample size here isn't nearly large enough to justify using z scores instead of the t distribution.

B: H_0: $\mu_d = 2.5$; H_a: $\mu_d > 2.5$; $\bar{d} = 2.725$; $s_d = 0.77598\ldots$; $t = 0.820\ldots$; $t_{0.05} = 1.895$; Fail to reject: "The study does not constitute evidence to support the claim that…"

C: (1) $0.742 < \mu < 1.178$ (2) The Central Limit Theorem assures us that with sample sizes this large, the averaged data would behave as normal, with a much smaller value for s, which means that using the t distribution is appropriate. The technique is fine.

D: H_0: $\mu_L = \mu_R$; H_a: $\mu_L < \mu_R$; conserv. df = 12; $t_{0.05} = (-)1.782$; $t = -4.0912\ldots$; Reject: "The study constitutes evidence to warrant rejection of the claim that…"