Paired-samples t-test

Create two paired samples, X.A and X.B. The fact that they're paired means

X.A[1] goes with X.B[1] (e.g., these are two measurements from the same subject), etc. We will do a paired-samples t-test to test whether these samples could likely have come from populations with the same mean.

First, compute the difference scores. Each difference score equals X.A minus X.B for one subject. We can get all the difference scores at once simply by subtracting the two vectors.
> X = X.A – X.B

Now we do a single-sample t-test on the difference scores, using mu = 0 as the null hypothesis. Compute the mean and standard deviation of the difference scores, and then insert everything into the formula for t:
> t = M/(s/sqrt(n))

Be careful to use the right value for n: the number of difference scores (e.g., the number of subjects), not the total number of observations. t tells you how large the mean difference score is, relative to how big it could be expected to be by chance. To decide whether t is consistent with the null hypothesis, find the p-value. We'll do a two-tailed test, which means we need the probability of an outcome greater than |t| plus the probability of an outcome less than -|t|. There are a few different ways to do this, because those two probabilities are the same, but here's an easy command (computing twice the probability of an outcome less than -|t|).
> 2*pt(-abs(t),df)

Remember that the degrees of freedom for a paired-samples t-test equals n–1, because the Mean Square is based on n difference scores, and one degree of freedom disappears because M is in the formula.

Finally, look at the p-value and decide whether it's consistent with the null hypothesis. If p≤.05, then the samples probably couldn't have come from populations with equal means. If p>.05, the samples could have come from populations with equal means. Compare your result to R's built-in t-test function.

> t.test(X)

<u>Paired- vs. independent-samples t-tests</u>

Paired-samples t-tests are used for repeated-measures experiments, which are experiments in which each subject is measured multiple times (twice). Repeated-measures experiments can be contrasted with between-subjects experiments, in which the subjects are split into two groups and measured once each. The main advantage of a repeated-measures design is that it lets us eliminate variance due to individual differences. The following example illustrates what this means.

Load the data in TTest.txt on the course website. Display the data (by typing whatever variable name you gave it).

Notice that the scores are highly variable, but they are consistent between the two samples. The first subject has a very low score on both measures, the second subject has high scores, and so on. This is what we mean by individual differences – the subjects vary among each other in a way that affects *both* scores for each subject. This is impossible for independent samples, because by definition there's no relationship between the scores in the two samples.

Let's look at how the individual differences affect the mean squared error. The Mean Square is an estimate of the population variance. If we ran a between-subjects experiment, we'd have gotten data similar to what we have here, but the samples wouldn't be paired. This means we'd have to estimate the variance of the raw scores. Try it. Because the samples are the same size, the formula for independent-samples MS reduces to the average of the two sample variances:
> MS.independent = (var(X.A) + var(X.B))/2

Now compute the difference scores, by subtracting X.B from X.A. Notice how the difference scores don't vary nearly as much as the raw scores. This is because the individual differences are gone. When you subtract one score from the other, the overall level of each subject cancels out. This means that the MS of the difference scores will be smaller, because it doesn't include the variance between subjects.
> MS.paired = var(X)

The fact that the MS is smaller for the paired-samples test means that t is larger, and therefore we have more power to reject the null hypothesis. Use the t.test() function to do an independent-samples t-test on the separate samples and a single-sample t-test on the difference scores. To do an independent-samples t-test, give the function both samples and tell it to assume the variances are homogeneous:
> t.test(X.A, X.B, var.equal=TRUE)

Notice that only the paired-samples test (i.e., on the difference scores) shows a reliable difference. Also, the confidence interval for the difference between population means is much narrower for the paired-samples test.

Confidence Intervals

Make a single sample. We'll compute a 95% confidence interval for the mean of the sample.

First, compute the mean and the standard error of the mean. Recall that the standard error equals the sample standard deviation divided by sqrt(n).

> SE = sqrt(var(X)/n)

Next, find the critical t value. Since we're computing a 95% confidence interval, we want the critical value for a two-tailed test with alpha = .05. Put differently, we want the value of t.crit such that the probability between –t.crit and +t.crit equals 95%. The means there's a 2.5% probability below –t.crit and 2.5% above +t.crit, which in turn means there's a 97.5% probability below t.crit.

> t.crit = qt(.975,df)

(Make sure you remember to define df as one less than the sample size.)

Now compute the confidence interval.
> CI.upper = M + SE*t.crit
> CI.lower = M - SE*t.crit
The confidence interval contains all potential values of the population mean, mu, that would not be rejected as the null hypothesis (by a two-tailed test with alpha = .05). Test this out by running t-tests using various values of mu.0.
> t.test(X,mu=_)

Notice how you get a result of p<.05 for any value of mu.0 outside the confidence interval, and a result of p>.05 for any value inside the interval. If you use a value of mu exactly equal to CI.upper or CI.lower, you'll get p=.05 and t = t.crit.

Assignment

In this assignment, you'll create two paired samples, do a t-test, and find the confidence interval for the mean difference.

1. Compute the difference scores.

2. Compute the mean, the standard error of the mean, and the degrees of freedom.

3. Compute t and find the two-tailed p-value.

4. Find the critical t value using alpha = .05. Is the t from your samples more extreme than the critical value? Write how the answer to this question relates to the result from Step 3.

5. Use the mean, standard error, and critical t to compute the confidence interval.

6. Is zero in the confidence interval? Write how the answer to this question relates to the result from Step 3.

7. Use t.test() to verify the results of your hypothesis test and confidence interval.

Answers

1.

> X = X.A – X.B

2.

> M = mean(X)

> SE = sqrt(var(X)/length(X))

> df = length(X)-1

3.

> t = M/SE

> p = 2*pt(-|t|,df)

4.

> t.crit = qt(.975,df)

Possibility 1: |t| > t.crit, which means t is in the critical region. This means I reject the null hypothesis, which is consistent with the result above that p < .05.
Possibility 2: |t| < t.crit, which means t is not in the critical region. This means I retain the null hypothesis, which is consistent with the result above that p > .05.

5.

> CI.upper = M + SE*t.crit

> CI.lower = M – SE*t.crit

6.

Possibility 1: 0 is not in the confidence interval. This means the null hypothesis mu = 0 is rejected, which is consistent with the result above that p < .05.
Possibility 2: 0 is in the confidence interval. This means the null hypothesis mu = 0 is retained, which is consistent with the result above that p > .05.

7.

> t.test(X)