Paired samples t test

Date: 2017-11-01

What it is

The paired samples t test is used to compare means under two conditions when observations from the two samples are related. Specifically, each observations in one sample is *paired* with exactly one observation in the second sample through the sampling procesure.

When to use it

The paired samples t test is used when individuals in each sample share a relationships (e.g., twins, brither/sister, husband/wife) or when two observations are taken from the same individual under different conditions (pre-test/post-test, control/treatment, etc.). We must have the same pairing for every observation in both samples. Since the test is run on the difference between observations within the pairs, all CLT conditions are relevant to the within-pair differences. In particular, the difference should be approximately normal of there should be a large sample (at least 30 pairs).

How to use it

Conceptually, the paired samples t test is performed by arranging the data so that x_1 and y_1 come from the same pair and so forth. Then you compute the difference $d_i = x_1 - y_1$ and take the average difference, $\bar{d}$. You then essentially run a one-sample t test on the differences.[1] The test statistic becomes

$$t = \frac{\bar{d} - 0}{s_d / \sqrt{n_d}}$$

with $df = n_d - 1$, where n_d is the number of differences from within the pairs.

In R, we perform the paired sample t test using the same `t.test()` function as we do for the independent samples t test. We enter the observations from condition 1 as `x`, the observations from contiion 2 as `y`, and an extra input `paired=TRUE`. We also include other inputs such as `alternative=` for a one-sided test, if needed.

The line of code would look something like[2]

```
t.test(x={Condition 1 data}, y={Condition 2 data}, paired=TRUE)
```

[1] We usually test $H_0 : \mu_d = 0$ against $H_A : \mu_d \neq 0$ (or a one-sided alternative). If we want the null to be a value different from 0, replace the 0 in the numerator of the test statistic with the value of μ_d from H_0.

[2] Again, with other inputs as needed.

Why it works

It is important to note that the average difference is equal to the difference of the averages. Mathematically, if $x_i - y_i = d_i$ and $\bar{d}$ is the average of the d_i, then $\bar{x} - \bar{y} = \bar{d}$ as well. So in some sense could perform an independent samples t test comparing $\bar{x} - \bar{y}$. So why have a paired test?

We expect paired data to be more similar (correlated) than unpaired data. When this is true, the standard error of the paired estimate is smaller than that of the independent samples estimate, meaning the paired estimate is more accurate.

In general, for random variables X and Y,

$$\begin{aligned} Var(X - Y) &= Var(X) + Var(Y) - 2Cov(X, Y) \\ &= Var(X) + Var(Y) - 2\sqrt{Var(X)Var(Y)} \cdot Cor(X, Y) \end{aligned}$$

So, a positive correlation between X and Y means the variance for the paired samples estimator will be less than $Var(X) + Var(Y)$, the variance for the independent samples estimator.[3]

[3] The same relationship holds for standard errors ($SE = \sqrt{Var}$).