

6.5: Independent samples t-test — The return of the t-test?

If you've been following the Star Wars references, we are on last movie (of the original trilogy)... the independent t-test. This is were basically the same story plays out as before, only slightly different.

Remember there are different t -tests for different kinds of research designs. When your design is a between-subjects design, you use an independent samples t-test. Between-subjects design involve different people or subjects in each experimental condition. If there are two conditions, and 10 people in each, then there are 20 total people. And, there are no paired scores, because every single person is measured once, not twice, no repeated measures. Because there are no repeated measures we can't look at the difference scores between conditions one and two. The scores are not paired in any meaningful way, to it doesn't make sense to subtract them. So what do we do?

The logic of the independent samples t-test is the very same as the other t -tests. We calculated the means for each group, then we find the difference. That goes into the numerator of the t formula. Then we get an estimate of the variation for the denominator. We divide the mean difference by the estimate of the variation, and we get t . It's the same as before.

The only wrinkle here is what goes into the denominator? How should we calculate the estimate of the variance? It would be nice if we could do something very straightforward like this, say for an experiment with two groups A and B:

$$t = \frac{\bar{X}_A - \bar{X}_B}{\sqrt{\frac{SEM_A^2 + SEM_B^2}{2}}}$$

In plain language, this is just:

1. Find the mean difference for the top part
2. Compute the SEM (standard error of the mean) for each group, and average them together to make a single estimate, pooling over both samples.

This would be nice, but unfortunately, it turns out that finding the average of two standard errors of the mean is not the best way to do it. This would create a biased estimator of the variation for the hypothesized distribution of no differences. We won't go into the math here, but instead of the above formula, we can use a different one that gives us an unbiased estimate of the pooled standard error of the sample mean. Our new and improved t formula would look like this:

$$t = \frac{\bar{X}_A - \bar{X}_B}{s_p \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}}$$

and, s_p , which is the pooled sample standard deviation is defined as, note the s 's in the formula are variances:

$$s_p = \sqrt{\frac{(n_A - 1)s_A^2 + (n_B - 1)s_B^2}{n_A + n_B - 2}}$$

Believe you me, that is so much more formula than I wanted to type out. Shall we do one independent t -test example by hand, just to see the computations? Let's do it...but in a slightly different way than you expect. I show the steps using R. I made some fake scores for groups A and B. Then, I followed all of the steps from the formula, but made R do each of the calculations. This shows you the needed steps by following the code. At the end, I print the t -test values I computed "by hand", and then the t -test value that the R software outputs using the t -test function. You should be able to get the same values for t , if you were brave enough to compute t by hand.

```
## By "hand" using R r code
a <- c(1,2,3,4,5)
b <- c(3,5,4,7,9)

mean_difference <- mean(a)-mean(b) # compute mean difference

variance_a <- var(a) # compute variance for A
variance_b <- var(b) # compute variance for B

# Compute top part and bottom part of sp formula

sp_numerator <- (4*variance_a + 4* variance_b)
sp_denominator <- 5+5-2
sp <- sqrt(sp_numerator/sp_denominator) # compute sp

# compute t following formulat

t <- mean_difference / ( sp * sqrt( (1/5) +(1/5) ) )
```

```
t # print results
```

runrestartrestart & run all

```
-2.01799136683647
```

```
a <- c(1,2,3,4,5)
```

```
b <- c(3,5,4,7,9)
```

```
# using the R function t.test
```

```
t.test(a,b, paired=FALSE, var.equal = TRUE)
```

runrestartrestart & run all

Two Sample t-test

```
data: a and b
```

```
t = -2.018, df = 8, p-value = 0.0783
```

```
alternative hypothesis: true difference in means is not equal to 0
```

```
95 percent confidence interval:
```

```
-5.5710785  0.3710785
```

```
sample estimates:
```

```
mean of x mean of y
```

```
3.0
```

```
5.6
```

This page titled 6.5: Independent samples t-test — The return of the t-test? is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Matthew J. C. Crump via source content that was edited to the style and standards of the LibreTexts platform.