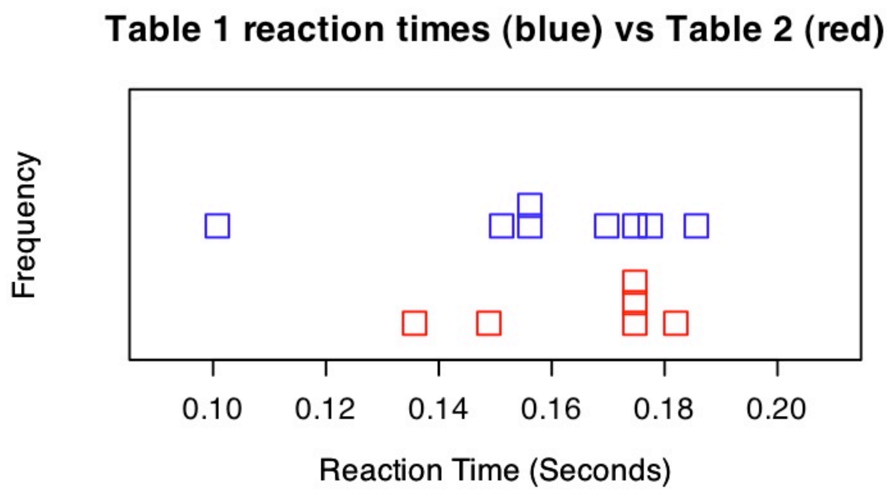
More Hypothesis Testing (Comparing Two Population Means)

Recap

First lets look at the reaction times again, separated by table.



It is obvious that there is no strong evidence of a difference in the "populations" underlying the students' reaction times. Nevertheless, if we go through the motions, here is what we would do:

First we summarize the two samples:

The mean and SD for table 1 was 0.159, 0.026
The mean and SD for table 2 was 0.165, 0.019

The question is whether the "null hypothesis" of no difference in population means is credible, in light of the data. The way we test the credibility of this null hypothesis is to see how the difference in sample means (i.e. 0.165-0.159 = .006 seconds) compares with its SD. But while we know how to estimate the SD of a sample mean, what we have here is the difference in two sample means, and so we need a little more theory to proceed.

Some Theory of Variability of Differences of Two Sample Means

When we were working out the SD of a sample mean, we did the following:

If X_1 and X_2 are independent,

$Var(X_1+X_2) = Var(X_1) + Var(X_2)$

If X_1 and X_2 are also Identically Distributed, Var (X_1+X_2) =2 Var (X_1) and similarly Var ($\sum_{i=1}^{n} X_i$)=nVar(X_1) and so Var($\overline{X}$) $= \frac{n * Var(X_1)}{n^2} = \frac{Var(X_1)}{n}$

In terms of SDs, SD($\overline{X}$)= $\frac{SD(X_1)}{\sqrt{n}}$ which we usually write as $\frac{\sigma}{\sqrt{n}}$

Note that our first step (i.e. Var (X_1+X_2) = Var (X_1) + Var(X_2))

can also prove that, when X_1 and X_2 are independent

Var (X_1-X_2) = Var (X_1) + Var(X_2)   (Note the minus sign)

since $(-1)^2 = 1$

Think what this means. **(X_1+X_2) and (X_1-X_2) have the same variability.**

Now consider the two random variables $\overline{X}_1$ and $\overline{X}_2$ where the subscripts now refer to two different samples. $\overline{X}_1$ and $\overline{X}_2$ are independent, and so the difference $(\overline{X_1} - \overline{X_2})$ has a Variance that is

Var ($\overline{X}_1$) + Var($\overline{X}_2$)

and an SD that is

$$\sqrt{Var(\overline{X_1}) + Var(\overline{X_2})}$$

which we usually write as

$$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

This is what we need:

$(\overline{\mathbf{X_1}} - \overline{\mathbf{X_2}})$ **has SD =** $\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

Now we usually do not know σ_1 or σ_2 . So we estimate with s_1 and s_2 .

So we compare $(\overline{\mathbf{X_1}} - \overline{\mathbf{X_2}})$ with $\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$ to see how far, in a statistical sense, the difference $(\overline{\mathbf{X_1}} - \overline{\mathbf{X_2}})$ is from 0. More specifically, we look at

$(\overline{\mathbf{X_1}} - \overline{\mathbf{X_2}}) / \sqrt{s_1^2/n_1 + s_2^2/n_2}$ and compare it to its assumed distribution – the distribution assumed under the assumption H_0 that $\mu_1 - \mu_2 = 0$ (which is the same as $\mu_1 = \mu_2$).

Now lets return to our comparison of the reaction time samples from the two tables:

We had $(\overline{\mathbf{X_1}} - \overline{\mathbf{X_2}})$ **= .006**

and

$s_1^2 = 0.026$, $s_2^2 = 0.019$ and $n_1 = 8$, $n_2 = 6$

So the crucial statistic is $.006/\sqrt{.026^2/8 + .019^2/6} = 0.500$

But what distribution do we use to decide if this is an unusual value (a small P-value)?

The answer is "a t-distribution with 7+5 = 12 df." Note the t-dist has mean 0, in line with our null hypothesis.

Using Table A.5 we can see right away that 0.500 is a fairly central value of t_{12} and so we do not reject the hypothesis that $\mu_1 = \mu_2$. Another way to get the same conclusion is to look at Table A.8 and infer that about 31.3% of the distribution of our statistics is to the right of 0.500 and so the P-value would be .313. If it made a difference we would even double this since we would have accepted a statistic that was too small as well as too large, so our P-value is more correctly .626. It certainly is not "small" like < .05 or < .01. So we Accept the hypothesis that $\mu_1 = \mu_2$ (again). In other words, there is no evidence that Table 1 has different reaction times, on average, than Table 2. It is reasonable to call it a "tie".

This comparison of two samples (Ch 9) uses the same logic as the tests we have been doing from Ch 8. The difference is that our statistic is now based on a difference of sample means, rather than a single sample mean, and so the SD that we compare it with must be the SD of the difference in sample means, rather than the SD of a sample mean.

Note: The procedure described here is very similar to the one on page 373 (in Ch 9): The difference is in the calculation of df: a good approximation is $(n_1\text{-}1) + (n_2\text{-}1)$ which is how we got the 12 df above. The book formula is technically more accurate but in hand calculations seldom used. But software like R uses it as shown on the next page:

R version of the t test of difference in reaction times from the two tables.

```
> t.test(reacs.sec[grp==1],reacs.sec[grp==2])

        Welch Two Sample t-test

data:  reacs.sec[grp == 1] and reacs.sec[grp == 2]
t = -0.5331, df = 11.98, p-value = 0.6037
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.03271012  0.01985104
sample estimates:
mean of x mean of y
0.1589928 0.1654224
```