STAT 414 / 415

Probability Theory and Mathematical Statistics

When Population Variances Are Not Equal

🖨 Printer-friendly version (https://onlinecourses.science.psu.edu/stat414/print/book/export/html/275)

Let's again start with the good news that we've already done the dirty theoretical work here. Recall that if you have two **independent samples** from two **normal distributions** with **unequal variances**

$$\sigma_X^2 \neq \sigma_Y^2$$

, then:

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{\sqrt{\dfrac{S_X^2}{n} + \dfrac{S_Y^2}{m}}}$$

follows, at least approximately, a t_r distribution where r, the adjusted degrees of freedom is determined by the equation:

$$r = \frac{\left(\dfrac{s_X^2}{n} + \dfrac{s_Y^2}{m}\right)^2}{\dfrac{(s_X^2/n)^2}{n-1} + \dfrac{(s_Y^2/m)^2}{m-1}}$$

If r doesn't equal an integer, as it usually doesn't, then we take the integer portion of r. That is, we use $[r]$ if necessary.

With that now being recalled, if we're interested in testing the null hypothesis:

$$H_0 : \mu_X - \mu_Y = 0 \text{ (or equivalently } H_0 : \mu_X = \mu_Y)$$

against any of the alternative hypotheses:

$$H_A : \mu_X - \mu_Y \neq 0, \quad H_A : \mu_X - \mu_Y < 0, \text{ or } H_A : \mu_X - \mu_Y > 0$$

we can use the test statistic:

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{\sqrt{\dfrac{S_X^2}{n} + \dfrac{S_Y^2}{m}}}$$

and follow the standard hypothesis testing procedures. Let's return to our fastest speed driven example.

Example (continued)

A psychologist was interested in exploring whether or not male and female college students have different driving behaviors. There were a number of ways that she could quantify driving behaviors. She opted to focus on the fastest speed ever driven by an individual. Therefore, the particular statistical question she framed was as follows:

> Is the mean fastest speed driven by male college students different than the mean fastest speed driven by female college students?

She conducted a survey of a random $n = 34$ male college students and a random $m = 29$ female college students. Here is a descriptive summary of the results of her survey:

Males (X)	Females (Y)
$n = 34$	$m = 29$
$\bar{x} = 105.5$	$\bar{y} = 90.9$
$s_x = 20.1$	$s_y = 12.2$

Is there sufficient evidence at the $\alpha = 0.05$ level to conclude that the mean fastest speed driven by male college students differs from the mean fastest speed driven by female college students?

> **Solution.** This time let's not assume that the population variances are equal. Then, we'll see if we arrive at a different conclusion. Let's still assume though that the two populations of fastest speed driven for males and females are normally distributed. And, we'll again permit the randomness of the two samples to allow us to assume independence of the measurements as well.
>
> That said, then we can test the null hypothesis:

$$H_0 : \mu_M - \mu_F = 0$$

against the alternative hypothesis:

$$H_A : \mu_M - \mu_F \neq 0$$

comparing the test statistic:

$$t = \frac{(105.5 - 90.9) - 0}{\sqrt{\dfrac{20.1^2}{34} + \dfrac{12.2^2}{29}}} = 3.54$$

to a *T* distribution with *r* degrees of freedom, where:

$$r = \frac{\left(\dfrac{12.2^2}{29} + \dfrac{20.1^2}{34}\right)^2}{\left(\dfrac{1}{28}\right)\left(\dfrac{12.2^2}{29}\right)^2 + \left(\dfrac{1}{33}\right)\left(\dfrac{20.1^2}{34}\right)^2} = 55.5$$

Oops... that's not an integer, so we're going to need to take the greatest integer portion of that *r*. That is, we take the degrees of freedom to be [*r*] = [55.5] = 55.

Then, the **critical value approach** tells us to reject the null hypothesis in favor of the alternative hypothesis if:

$$t > t_{0.025,55} = 2.004$$

We reject the null hypothesis because the test statistic ($t = 3.54$) falls in the rejection region:
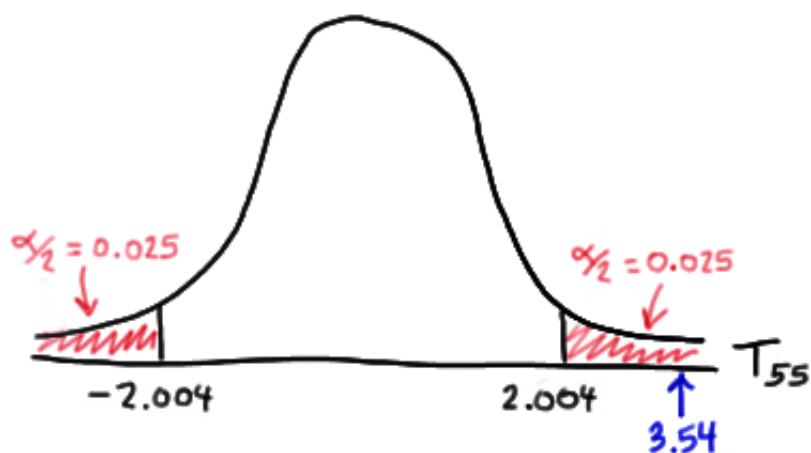
There is (again!) sufficient evidence at the $\alpha = 0.05$ level to conclude that the average fastest speed driven by the population of male college students differs from the average fastest speed driven by the population of female college students.

And again, the decision is the same using the *P*-value approach. The *P*-value is 0.0008:

$$P = 2 \times P(T_{55} > 3.54) = 2(0.0004) = 0.0008$$

Therefore, because $P = 0.0008 \leq \alpha = 0.05$, we reject the null hypothesis in favor of the alternative hypothesis. Again, we conclude that there is sufficient evidence at the $\alpha = 0.05$ level to conclude that the average fastest speed driven by the population of male college students differs from the average fastest speed driven by the population of female college students.

At any rate, we see that in this case, our conclusion is the same regardless of whether or

not we assume equality of the population variances.

And, just in case you're interested... we'll see how to tell Minitab to conduct a Welch's *t*-test very soon, but in the meantime, this is what the output would look like for this example:

```
Two-sample T for fastest

gender   N   Mean  StDev  SE Mean
1       34  105.5   20.1     3.4
2       29   90.9   12.2     2.3


Difference = mu (1) - mu (2)
Estimate for difference:  14.6085
95% CI for difference:  (6.3575, 22.8596)
T-Test of difference = 0 (vs not =): T-Value = 3.55  P-Value = 0.001  DF = 55
```

‹ When Population Variances Are Equal   up   Using Minitab › (/stat414/node/302)
(/stat414/node/274)   (/stat414/node/224)