

STAT 414 / 415

Probability Theory and Mathematical Statistics

When Population Variances Are Equal

 [Printer-friendly version \(https://onlinecourses.science.psu.edu/stat414/print/book/export/html/274\)](https://onlinecourses.science.psu.edu/stat414/print/book/export/html/274)

Let's start with the good news, namely that we've already done the dirty theoretical work 😊 in developing a hypothesis test for the difference in two population means $\mu_1 - \mu_2$ when we developed a $(1-\alpha)100\%$ confidence interval for the difference in two population means. Recall that if you have two **independent samples** from two **normal distributions** with **equal variances**

$$\sigma_X^2 = \sigma_Y^2 = \sigma^2$$

, then:

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{S_p \sqrt{\frac{1}{n} + \frac{1}{m}}}$$

follows a t_{n+m-2} distribution where S_p^2 , the pooled sample variance:

$$S_p^2 = \frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2}$$

is an unbiased estimator of the common variance σ^2 . Therefore, if we're interested in testing the null hypothesis:

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

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:

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{S_p \sqrt{\frac{1}{n} + \frac{1}{m}}}$$

and follow the standard hypothesis testing procedures. Let's take a look at an example.

Example

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:

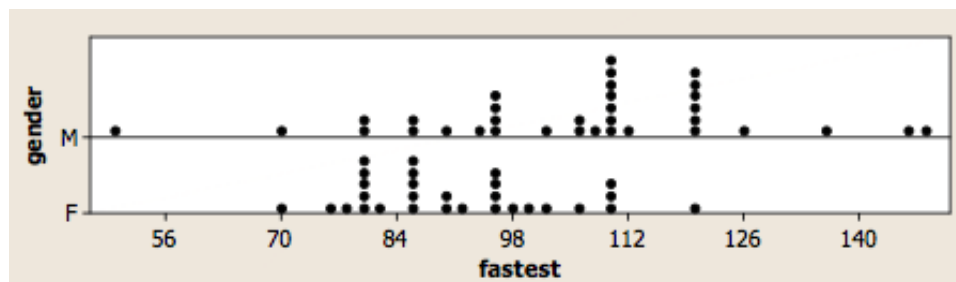


Is the mean fastest speed driven by male college students different than the mean fastest speed driven by female college students?

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)
$n = 34$	$m = 29$
$\bar{x} = 105.5$	$\bar{y} = 90.9$
$s_x = 20.1$	$s_y = 12.2$

and here is a graphical summary of the data in the form of a dotplot:



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. Because the observed standard deviations of the two samples are of similar magnitude, we'll assume that the population variances are equal. Let's also assume that the two populations of fastest speed driven for males and females are normally distributed. (We can confirm, or deny, such an assumption using a normal probability plot, but let's simplify our analysis for now.) The randomness of the two samples allows us to assume independence of the measurements as well.

Okay, assumptions all met, we can test the null hypothesis:

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

against the alternative hypothesis:

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

using the test statistic:

$$t = \frac{(105.5 - 90.9) - 0}{16.9 \sqrt{\frac{1}{34} + \frac{1}{29}}} = 3.42$$

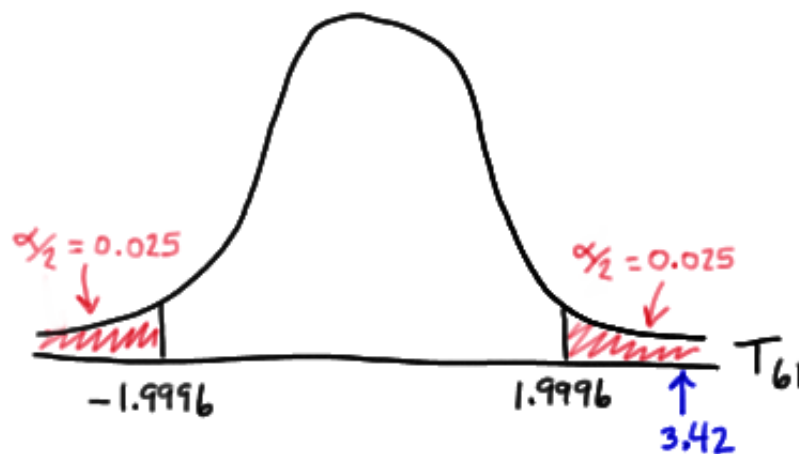
because, among other things, the pooled sample standard deviation is:

$$s_p = \sqrt{\frac{33(20.1^2) + 28(12.2^2)}{61}} = 16.9$$

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

$$|t| \geq t_{\alpha/2, n+m-2} = t_{0.025, 61} = 1.9996$$

We reject the null hypothesis because the test statistic ($t = 3.42$) falls in the rejection region:



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.

Not surprisingly, the decision is the same using the P -value approach. The P -value is 0.0012:

$$P = 2 \times P(T_{61} > 3.42) = 2(0.0006) = 0.0012$$

Therefore, because $P = 0.0012 \leq \alpha = 0.05$, 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.

By the way, we'll see how to tell Minitab to conduct a two-sample t -test in a bit here, but in the meantime, this is what the output would look like:

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 = μ (1) - μ (2)

Estimate for difference: 14.6085

95% CI for difference: (6.0630, 23.1540)

T-Test of difference = 0 (vs not =): T-Value = 3.42 P-Value = 0.001 DF = 61

Both use Pooled StDev = 16.9066

◀ Lesson 39: Tests of the Equality of Two
Means (/stat414/node/224)

up
(/stat414/node/224)

When Population Variances Are Not
Equal ▶ (/stat414/node/275)
