Hypothesis testing for the difference of means

Now that we know how to build a confidence interval around the difference of means, let's work through the entire hypothesis testing procedure when we want to use the difference of sample means to make an inference about the difference of population means.

Building hypothesis statements

The null and alternative hypotheses will always be formulated in terms of the difference between the two population means, $\mu_1 - \mu_2$, and we can have three different scenarios.

In a two-tailed test, the null hypothesis will state that the means don't differ, whereas the alternative hypothesis states that there *is* a difference between means. So we write the hypothesis statements for a two-tailed test as

$$H_0 : \mu_1 - \mu_2 = 0$$

$$H_a : \mu_1 - \mu_2 \neq 0$$

or

$$H_0 : \mu_1 = \mu_2$$

$$H_a : \mu_1 \neq \mu_2$$

In an upper-tailed test, the alternative hypothesis states that the difference in means is positive, so we write

$$H_0 : \mu_1 - \mu_2 \leq 0$$

$$H_a : \mu_1 - \mu_2 > 0$$

or

$$H_0 : \mu_1 \leq \mu_2$$

$$H_a : \mu_1 > \mu_2$$

In a lower-tailed test, the alternative hypothesis states that the difference in means is negative, so we write

$$H_0 : \mu_1 - \mu_2 \geq 0$$

$$H_a : \mu_1 - \mu_2 < 0$$

or

$$H_0 : \mu_1 \geq \mu_2$$

$$H_a : \mu_1 < \mu_2$$

Calculating the test statistic

Large samples, unequal population variances

If the independent random samples we take from each population are both large enough, $n_1, n_2 \geq 30$, and the population variances are unequal, then the test statistic formula we'll use is

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

When our hypothesis statements only test for a difference of means, then $\mu_1 - \mu_2 = 0$, and the test statistic formula simplifies to

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Large samples, equal population variances

On the other hand, if the independent random samples we take from each population are both large enough, $n_1, n_2 \geq 30$, but our population variances are reasonably equal, then we use the formula for pooled variance,

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

and the formula for the test statistic becomes

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

or when $\mu_1 - \mu_2 = 0$, the formula simplifies to

$$z = \frac{\bar{x}_1 - \bar{x}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Small sample(s), unequal population variances

If one or both of our independent random samples is/are small, $n_1 < 30$ and/or $n_2 < 30$, and the population variances are unequal, then the test statistic formula we'll use is

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

or when $\mu_1 - \mu_2 = 0$, the formula simplifies to

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

When we calculate degrees of freedom, we'll use

$$\text{df} = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{s_2^2}{n_2}\right)^2}$$

If we find a non-integer value for degrees of freedom, we should always round down to the next lowest integer so that the estimate is more conservative.

Small sample(s), equal population variances

On the other hand, if one or both of the independent random samples we take from each population is/are small, $n_1 < 30$ and/or $n_2 < 30$, and the population variances are reasonably equal, then we use the formula for pooled variance,

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

and the formula for the test statistic becomes

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

or when $\mu_1 - \mu_2 = 0$, the formula simplifies to

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

When we calculate degrees of freedom, we'll use $\text{df} = n_1 + n_2 - 2$.

Making a conclusion

Once we've calculated a test statistic, we'll use the significance level α, along with either the p-value approach or the critical value approach, to determine whether or not we can reject the null hypothesis.

Let's work through an example so that we can see how this works.

Example

Suppose we want to determine whether the mean height of men is significantly higher than the mean height of women in a certain city, so we randomly sample 100 men and 100 women. Given the mean and standard deviation of both samples below, use the critical value approach to say whether men are significantly taller than women at a 1% level of significance.

Men	**Women**
$n_1 = 100$	$n_2 = 100$
$\bar{x}_1 = 69.5$ inches	$\bar{x}_2 = 67.8$ inches
$s_1 = 1.25$ inches	$s_2 = 1.12$ inches

Since we want to test the claim the the mean height of men is higher than the mean height of women, our hypothesis statements will be

$$H_0 : \mu_M - \mu_W \leq 0$$

$$H_a : \mu_M - \mu_W > 0$$

Because the sample standard deviations are $s_1 = 1.25$ and $s_2 = 1.12$, the sample variances are $s_1^2 = 1.25^2 = 1.5625$ and $s_2^2 = 1.12^2 = 1.2544$. The sample variance 1.5625 isn't more than twice the sample variance 1.2544, which means we can assume that the sample variances are reasonably equal, and therefore that the population variances are reasonably equal, so we'll use pooled variance.

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

$$s_p = \sqrt{\frac{(100 - 1)1.25^2 + (100 - 1)1.12^2}{100 + 100 - 2}}$$

$$s_p = \sqrt{\frac{99(1.5625) + 99(1.2544)}{198}}$$

$$s_p = \sqrt{\frac{1.5625 + 1.2544}{2}}$$

$$s_p = \sqrt{\frac{2.8169}{2}}$$

$$s_p \approx 1.187$$

Now calculate the z-test statistic.

$$z = \frac{\bar{x}_1 - \bar{x}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$z = \frac{69.5 - 67.8}{1.187\sqrt{\frac{1}{100} + \frac{1}{100}}}$$

$$z = \frac{1.7}{1.187 \cdot \frac{\sqrt{2}}{10}}$$

$$z \approx 10.13$$

Because we want to test at a significance level of 1%, our confidence level is $1 - \alpha = 1 - 0.01 = 0.99$. We're using a right-tailed test, so we need to use the z-table to find the z-score that corresponds to the probability 0.99. In a z-table, we find $z = 2.33$.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936

Therefore, if the z-test statistic that we've found is larger than 2.33, we would reject the null hypothesis. Since $10.13 > 2.33$, we can reject the null hypothesis at $\alpha = 0.01$ and conclude that the mean height of men is greater than the mean height of women.

Let's use another example to illustrate the use of a t-test for the difference of means when the variances are unequal.

Example

A company believes its new light bulb will last at least 30 days longer than its old bulb. They take random samples of 20 old bulbs and 20 new bulbs, and find the following:

New bulb	Old bulb
$n_1 = 20$	$n_2 = 20$
$\bar{x}_1 = 254$ days	$\bar{x}_2 = 205$ days

$s_1 = 5$ days $\qquad\qquad\qquad\qquad$ $s_2 = 13$ days

At a 0.05 level of significance, test the claim that the new bulb lasts at least 30 days longer than the old bulb.

Given μ_1 as the life expectancy of the new bulb, and μ_2 as the life expectancy of the old bulb, our hypothesis statements for the right-tailed test will be

$$H_0 : \mu_1 - \mu_2 \leq 30$$

$$H_a : \mu_1 - \mu_2 > 30$$

Because the sample standard deviations are $s_1 = 5$ and $s_2 = 13$, the sample variances are $s_1^2 = 5^2 = 25$ and $s_2^2 = 13^2 = 169$. We can see that s_2^2 is more than double s_1^2, so we can assume the sample variances are unequal, and therefore that the population variances are unequal. Because of this, and the fact that we have small samples, n_1, $n_2 < 30$, the test statistic will be

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

where $\mu_1 - \mu_2 = 30$. Substitute what we know to calculate the t-test statistic.

$$t = \frac{(254 - 205) - 30}{\sqrt{\frac{5^2}{20} + \frac{13^2}{20}}}$$

$$t = \frac{19}{\sqrt{\frac{25}{20} + \frac{169}{20}}}$$

$$t = \frac{19}{\sqrt{\frac{194}{20}}}$$

$$t = 38\sqrt{\frac{5}{194}}$$

$$t \approx 6.101$$

The degrees of freedom will be

$$\text{df} = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{s_2^2}{n_2}\right)^2}$$

$$\text{df} = \frac{\left(\frac{5^2}{20} + \frac{13^2}{20}\right)^2}{\frac{1}{20 - 1}\left(\frac{5^2}{20}\right)^2 + \frac{1}{20 - 1}\left(\frac{13^2}{20}\right)^2}$$

$$\text{df} = \frac{\left(\frac{25}{20} + \frac{169}{20}\right)^2}{\frac{1}{19}\left(\frac{25}{20}\right)^2 + \frac{1}{19}\left(\frac{169}{20}\right)^2}$$

$$\text{df} = \frac{\frac{37{,}636}{400}}{\frac{625}{7{,}600} + \frac{28{,}561}{7{,}600}}$$

$$\text{df} = \frac{37{,}636}{400}\left(\frac{7{,}600}{29{,}186}\right)$$

$$\text{df} = \frac{715{,}084}{29{,}186}$$

$$\text{df} \approx 24.501$$

Always round down for a more conservative estimate.

$$\text{df} \approx 24$$

Now find the critical t-value from the t-table using $1 - \alpha = 0.95$ and $\text{df} = 24$. Because we're running an upper-tailed test, the whole region of rejection is consolidated into the upper tail, which means we're looking at upper-tail probability of 0.05.

	Upper-tail probability p									
df	**0.25**	**0.20**	**0.15**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**	**0.001**	**0.0005**
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.685	0.857	1.059	1.318	**1.711**	2.064	2.492	2.797	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
	50%	**60%**	**70%**	**80%**	**90%**	**95%**	**98%**	**99%**	**99.8%**	**99.9%**
	Confidence level C									

The critical t-value is 1.711. Using the critical value approach, we compare 6.101 to 1.711. Because $6.101 > 1.711$, we reject the null hypothesis and conclude that the new bulb lasts at least 30 days longer than the old bulb.