

Two sample tests – Cases I, II, III

I. Overview

A. So far, we have concentrated on 1-sample tests. Often, however, we are interested in comparing 2 or more groups

EXAMPLES.

- ✓ Do women have lower wages than men?
- ✓ Do smokers have higher cancer rates than non-smokers?
- ✓ Do students in Catholic schools learn more than students in public schools?
- ✓ Do those who receive an experimental treatment behave differently than those who do not receive it?

B. We are therefore interested in testing hypotheses about μ_1 and μ_2 ; typically, we want to test

$$H_0: \mu_1 - \mu_2 = 0 \quad (\text{or, } \mu_1 = \mu_2)$$

versus

$$H_A: \mu_1 - \mu_2 \neq 0 \quad (\text{or, } \mu_1 \neq \mu_2)$$

or

$$H_A: \mu_1 - \mu_2 > 0 \quad (\text{or, } \mu_1 > \mu_2)$$

or

$$H_A: \mu_1 - \mu_2 < 0 \quad (\text{or, } \mu_1 < \mu_2)$$

C. To test these hypotheses, we usually take independent samples from both groups - and we estimate $\mu_1 - \mu_2$ by $\hat{\mu}_1 - \hat{\mu}_2$.

D. What is the mean, variance, and standard error of $\hat{\mu}_1 - \hat{\mu}_2$? Recall that $E(X + Y) = E(X) + E(Y)$ (Expectations Rule 8) and If X and Y are independent, $V(X \pm Y) = V(X) + V(Y)$ (Expectations Rule 16). Therefore, since these are independent samples,

$$E(\hat{\mu}_1 - \hat{\mu}_2) = E(\hat{\mu}_1) - E(\hat{\mu}_2) = \mu_1 - \mu_2,$$

$$V(\hat{\mu}_1 - \hat{\mu}_2) = V(\hat{\mu}_1) + V(\hat{\mu}_2) = \frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2},$$

$$\text{True S.E.}(\hat{\mu}_1 - \hat{\mu}_2) = \sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}$$

II. Testing hypotheses. As usual, we want to be able to construct a z or t score. Note that, if H_0 is $\mu_1 - \mu_2 = 0$, we are left with

$$z = \frac{\hat{\mu}_1 - \hat{\mu}_2}{\text{True s.e. of the point estimator}}, \text{ or else } t_v = \frac{\hat{\mu}_1 - \hat{\mu}_2}{\text{Estimated s.e.}}$$

As usual, how we test depends on the nature of the data and the information available.

More generally, suppose we are interested in some parameter Θ (in this case, $\Theta = \mu_1 - \mu_2$). Then, depending on the information available, the test statistic will often be of the form

$$z = \frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}}, \text{ or else } t_v = \frac{\hat{\theta} - \theta_0}{\hat{\sigma}_{\hat{\theta}}}$$

In other words, compute the difference between what was empirically observed ($\hat{\theta}$) and what the null hypothesis predicted (θ_0), and then divide by either the True Standard Error of the Estimate ($\sigma_{\hat{\theta}}$) or the Estimated Standard Error ($\hat{\sigma}_{\hat{\theta}}$). If the difference between what was predicted and what was empirically observed was small, we will not reject the null, but if it is large we will reject.

A. Case I: σ_1 and σ_2 are known. This is the simplest case, but also the least common.

1. True Standard error: As noted above,

$$\sigma_{(\hat{\mu}_1 - \hat{\mu}_2)} = \sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}$$

2. Test statistic:

$$z = \frac{\hat{\mu}_1 - \hat{\mu}_2 - (\mu_1 - \mu_2)_0}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}}$$

If the null hypothesis is $H_0: \mu_1 = \mu_2$, which is usually (albeit not always) the case, this simplifies to

$$z = \frac{\hat{\mu}_1 - \hat{\mu}_2 - (\mu_1 - \mu_2)_0}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}} = \frac{\hat{\mu}_1 - \hat{\mu}_2}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}}$$

3. *Confidence interval:*

$$\hat{\mu}_1 - \hat{\mu}_2 \pm z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}$$

4. *2-tailed acceptance region (variables not standardized):*

$$(\mu_1 - \mu_2)_0 \pm z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}$$

5. *Examples:*

1. Indiana University (population 1) claims that it has a lower crime rate than Ohio State University (population 2). A random sample of crime rates for 12 different months is drawn for each school, yielding $\hat{\mu}_1 = 370$ and $\hat{\mu}_2 = 400$. It is known that $\sigma_1^2 = 400$ and $\sigma_2^2 = 800$. Test Indiana's claim at the .02 level of significance.

SOLUTION:

Step 1:

$$\begin{aligned} H_0: & E(X)_1 - E(X)_2 = 0 && \text{or, } E(X)_1 = E(X)_2 \\ H_A: & E(X)_1 - E(X)_2 < 0 && \text{or, } E(X)_1 < E(X)_2 \end{aligned}$$

Step 2:

Since the population variances are known, the appropriate test statistic is

$$z = \frac{\hat{\mu}_1 - \hat{\mu}_2}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}} = \frac{\hat{\mu}_1 - \hat{\mu}_2}{\sqrt{\frac{400}{12} + \frac{800}{12}}} = \frac{\hat{\mu}_1 - \hat{\mu}_2}{10}$$

Step 3:

At the .02 level of significance, accept H_0 if $z \geq -2.05$, or, equivalently, if $\hat{\mu}_1 - \hat{\mu}_2 \geq -20.5$.

Step 4:

$$z = \frac{\hat{\mu}_1 - \hat{\mu}_2}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}} = \frac{\hat{\mu}_1 - \hat{\mu}_2}{\sqrt{\frac{400}{12} + \frac{800}{12}}} = \frac{\hat{\mu}_1 - \hat{\mu}_2}{10} = \frac{370 - 400}{10} = -3.0$$

Step 5. Reject H_0 ; the computed value of the test statistic falls outside of the acceptance region.

2. Use the same data as in problem 1. (a) Construct the 99% c.i. for $E(X)_1 - E(X)_2$ (b) Test, at the .01 level of significance,

$$\begin{aligned} H_0: & E(X)_1 - E(X)_2 = 0 && \text{or, } E(X)_1 = E(X)_2 \\ H_A: & E(X)_1 - E(X)_2 \neq 0 && \text{or, } E(X)_1 \neq E(X)_2 \end{aligned}$$

Solution.

(a) The 99% confidence interval is

$$\hat{\mu}_1 - \hat{\mu}_2 \pm z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}, \text{ i.e. } 370 - 400 \pm 2.58 \sqrt{\frac{400}{12} + \frac{800}{12}}, \text{ i.e.}$$

$$-30 - (2.58 * 10) \leq \mu_1 - \mu_2 \leq -30 + (2.58 * 10), \text{ i.e.}$$

$$-55.8 \leq \mu_1 - \mu_2 \leq -4.2$$

(b) Since the value specified in H_0 does not fall in the 99% c.i. (i.e. 0 is not between -55.8 and -4.2), we know that we can reject H_0 . Alternatively, we can answer (b) using our usual hypothesis testing procedures. The appropriate test statistic is the same as in problem 1. The acceptance region is $-2.58 \leq Z \leq 2.58$, or, equivalently, $-25.8 \leq \hat{\mu}_1 - \hat{\mu}_2 \leq 25.8$. Since $z_c = -3.0$ (the same as in problem 1) (or, equivalently, since the observed difference of -30 is not between -25.8 and 25.8) we reject H_0 .

B. Case II: σ_1 and σ_2 are unknown but assumed to be equal.

1. *True Standard error of $\hat{\mu}_1 - \hat{\mu}_2$ when $\sigma_1 = \sigma_2 = \sigma$:*

$$\sigma_{(\hat{\mu}_1 - \hat{\mu}_2)} = \sqrt{\frac{\sigma^2}{N_1} + \frac{\sigma^2}{N_2}} = \sqrt{\frac{\sigma^2 N_2 + \sigma^2 N_1}{N_1 N_2}} = \sqrt{\frac{\sigma^2 (N_1 + N_2)}{N_1 N_2}} = \sqrt{\sigma^2 \frac{(N_1 + N_2)}{N_1 N_2}}$$

2. *Estimated standard error.* We don't know σ^2 . We have to estimate it. We have two estimates though - s_1^2 and s_2^2 . Our best estimate of σ^2 is a weighted estimate of s_1^2 and s_2^2 , specifically

$$s^2 = \frac{DF_1 s_1^2 + DF_2 s_2^2}{DF_1 + DF_2} = \frac{(N_1 - 1)s_1^2 + (N_2 - 1)s_2^2}{N_1 + N_2 - 2}$$

Note that, the larger the sample size for a particular population is, the more heavily weighted its standard deviation is in the calculations. Substituting s^2 for σ^2 in our previous equation, we see that the *estimated standard error* is

$$\hat{\sigma}_{\hat{\mu}_1 - \hat{\mu}_2} = \sqrt{s^2 \left(\frac{N_1 + N_2}{N_1 N_2} \right)} = \sqrt{\left(\frac{(N_1 - 1)s_1^2 + (N_2 - 1)s_2^2}{N_1 + N_2 - 2} \right) \left(\frac{N_1 + N_2}{N_1 N_2} \right)}$$

3. *Test statistic:*

$$t = \frac{\hat{\mu}_1 - \hat{\mu}_2 - (\mu_1 - \mu_2)_0}{\sqrt{\left(\frac{(N_1 - 1)s_1^2 + (N_2 - 1)s_2^2}{N_1 + N_2 - 2} \right) \left(\frac{N_1 + N_2}{N_1 N_2} \right)}}$$

When H_0 is true, the test statistic has a T distribution with $N_1 + N_2 - 2$ degrees of freedom.

4. *Confidence interval:*

$$\hat{\mu}_1 - \hat{\mu}_2 \pm T_{\alpha/2, v} \sqrt{\left(\frac{(N_1 - 1)s_1^2 + (N_2 - 1)s_2^2}{N_1 + N_2 - 2} \right) \left(\frac{N_1 + N_2}{N_1 N_2} \right)}$$

5. *2-tailed acceptance region (Variables not standardized):*

$$(\mu_1 - \mu_2)_0 \pm T_{\alpha/2, v} \sqrt{\left(\frac{(N_1 - 1)s_1^2 + (N_2 - 1)s_2^2}{N_1 + N_2 - 2} \right) \left(\frac{N_1 + N_2}{N_1 N_2} \right)}$$

6. *Examples:*

1. A professor believes that women do better on her exams than men do. A sample of 8 women ($N_1 = 8$) and 10 men ($N_2 = 10$) yields $\hat{\mu}_1 = 7$, $\hat{\mu}_2 = 5.5$, $s_1^2 = 1$, $s_2^2 = 1.7$. Using $\alpha = .01$, test whether the female mean is greater than the male mean. Assume that $\sigma_1 = \sigma_2 = \sigma$.

Step 1:

H_0 : $E(X)_1 - E(X)_2 = 0$ or, $E(X)_1 = E(X)_2$

H_A : $E(X)_1 - E(X)_2 > 0$ or, $E(X)_1 > E(X)_2$

Step 2. The appropriate test statistic is

$$\begin{aligned}
t &= \frac{\hat{\mu}_1 - \hat{\mu}_2 - (\mu_1 - \mu_2)_0}{\sqrt{\left(\frac{(N_1 - 1)s_1^2 + (N_2 - 1)s_2^2}{N_1 + N_2 - 2}\right)\left(\frac{N_1 + N_2}{N_1 N_2}\right)}} = \frac{\hat{\mu}_1 - \hat{\mu}_2}{\sqrt{\left(\frac{7s_1^2 + 9s_2^2}{8 + 10 - 2}\right)\left(\frac{8 + 10}{8 * 10}\right)}} \\
&= \frac{\hat{\mu}_1 - \hat{\mu}_2}{\sqrt{\left(\frac{7 * 1 + 9 * 1.7}{8 + 10 - 2}\right)\left(\frac{8 + 10}{8 * 10}\right)}} = \frac{\hat{\mu}_1 - \hat{\mu}_2}{\sqrt{\left(\frac{22.3}{16}\right)\left(\frac{18}{80}\right)}} = \frac{\hat{\mu}_1 - \hat{\mu}_2}{\sqrt{\frac{401.4}{1280}}} = \frac{\hat{\mu}_1 - \hat{\mu}_2}{\sqrt{.31359}} \\
&= \frac{\hat{\mu}_1 - \hat{\mu}_2}{.56}
\end{aligned}$$

(The value of the denominator, of course, could not be determined until after we knew what the two sample variances were.)

Step 3. Since $n_1 + n_2 - 2 = 16$, we will accept H_0 if $T_{16} \leq 2.583$ (since for a T with $v = 16$, $F(2.583) = .99$), or, equivalently, accept H_0 if $\hat{\mu}_1 - \hat{\mu}_2 \leq 2.583 * .56$, i.e. $\hat{\mu}_1 - \hat{\mu}_2 \leq 1.446$.

Step 4.

$$t = \frac{\hat{\mu}_1 - \hat{\mu}_2}{.56} = \frac{1.5}{.56} = 2.679$$

Step 5. We should reject H_0 , the computed T value is outside the acceptance region. Or, equivalently, the observed difference of 1.5 is greater than 1.446, so reject H_0 .

2. Using the same data as in problem 1, compute (a) the 99% c.i. for $E(X)_1 - E(X)_2$ (b) test, at the .01 level of significance,

$$\begin{aligned}
H_0: & E(X)_1 - E(X)_2 = 0 && \text{or, } E(X)_1 = E(X)_2 \\
H_A: & E(X)_1 - E(X)_2 \neq 0 && \text{or, } E(X)_1 \neq E(X)_2
\end{aligned}$$

SOLUTION. (a) Note that, for $v = 16$, $F(2.921) = 1 - \alpha/2 = .995$. Since $\hat{\mu}_1 - \hat{\mu}_2 = 1.5$, and the estimated standard error of the mean is .56 (see above), the 99% c.i. is $1.5 \pm (2.921 * .56)$, i.e.

$$-0.136 \leq E(X)_1 - E(X)_2 \leq 3.136$$

(b) Note that the hypothesized value of $E(X)_1 - E(X)_2 = 0$ does fall within the 99% c.i.; hence, we will not reject H_0 . To double-check, we can use our hypothesis testing procedures.

The appropriate test statistic is the same as in problem 1. The acceptance region is $-2.921 \leq T_{16} \leq 2.921$, or, equivalently, $-1.636 \leq \hat{\mu}_1 - \hat{\mu}_2 \leq 1.636$ (since $2.921 * .56 = 1.636$).

As we saw in problem 1, $t_c = 1.5/.56 = 2.679$. This falls within the acceptance region of $-2.921 \leq T_{16} \leq 2.921$. Or, equivalently, note that $\hat{\mu}_1 - \hat{\mu}_2 = 1.5$, and this value falls in the region $-1.636 \leq \hat{\mu}_1 - \hat{\mu}_2 \leq 1.636$. Hence, do not reject H_0 .

C. Case III: σ_1 and σ_2 are not known and are not assumed to be equal. This case is quite common; for example we might expect men's and women's average wages and variability in wages to differ.

1. *Estimated standard error:*

$$\hat{\sigma}_{\hat{\mu}_1 - \hat{\mu}_2} = \sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}$$

2. *Test statistic:*

$$t = \frac{\hat{\mu}_1 - \hat{\mu}_2 - (\mu_1 - \mu_2)_0}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}}$$

When H_0 is true, the test statistic has a T distribution; but, what are the degrees of freedom equal to??? Obviously, the d.f. are

$$v = \frac{\left(\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}\right)^2}{\left(\frac{s_1^2}{N_1}\right)^2 \left(\frac{1}{N_1 - 1}\right) + \left(\frac{s_2^2}{N_2}\right)^2 \left(\frac{1}{N_2 - 1}\right)}$$

Stata refers to this as *Satterthwaite's degrees of freedom*. Note that this does not have to equal an even integer, so round off to the nearest whole number.

HISTORICAL FOOTNOTE/ALTERNATIVE FORMULA: Prior to 2003, I always used this equally intuitive formula:

$$v = \frac{\left(\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}\right)^2}{\left(\frac{s_1^2}{N_1}\right)^2 \left(\frac{1}{N_1+1}\right) + \left(\frac{s_2^2}{N_2}\right)^2 \left(\frac{1}{N_2+1}\right)} - 2$$

Stata refers to this as *Welch's degrees of freedom*. This formula can be found in Hays, Blalock, and probably other major texts. However, the default for both Stata and SPSS is *Satterthwaite's degrees of freedom*. So, I will bow to their judgment from now on (but don't be surprised if you see inconsistencies across sources or in my notes!!!)

3. *Confidence interval:*

$$\hat{\mu}_1 - \hat{\mu}_2 \pm t_{\alpha/2, v} \sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}$$

4. *2-tailed acceptance region (Variables not standardized):*

$$(\mu_1 - \mu_2)_0 \pm t_{\alpha/2, v} \sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}$$

5. *Special Case - Case II vs. Case III when Ns are equal:* For case II and Case III, consider the special cases where $N_1 = N_2 = N$. For case II, the estimated standard error is

$$\sqrt{\left(\frac{(N-1)s_1^2 + (N-1)s_2^2}{N+N-2}\right)\left(\frac{N+N}{NN}\right)} = \sqrt{\left(\frac{(N-1)(s_1^2 + s_2^2)}{2N-2}\right)\left(\frac{2N}{NN}\right)} = \sqrt{\left(\frac{(N-1)(s_1^2 + s_2^2)}{2(N-1)}\right)\left(\frac{2}{N}\right)} = \sqrt{\frac{s_1^2 + s_2^2}{N}}$$

and for case III the estimated standard error is

$$\sqrt{\frac{s_1^2}{N} + \frac{s_2^2}{N}} = \sqrt{\frac{s_1^2 + s_2^2}{N}}$$

Thus, when the sample sizes are equal, we get the same estimate of t_c regardless of whether we use case II or III, but the appropriate degrees of freedom will likely slightly differ. When the sample sizes are unequal, both the value of the test statistic and the degrees of freedom will likely differ.

Why is this special case of equal Ns important?

(1) If in doubt whether Case II or Case III applies, you minimize the possible damage of a mistake if you can choose your sample sizes so that $n_1 = n_2$. This is because the test statistic has the same value under both cases. The degrees of freedom slightly differ, which leads to slightly different acceptance regions, but only rarely will this make a difference. That is partially why, in experimental research, researchers prefer to have groups of equal size for each treatment condition.

(2) If both sample sizes are large, it really doesn't matter that much whether you assume Case II or Case III or whether the sample sizes are equal; either way you will tend to reach the same conclusions.

6. Examples:

1. Rework the previous Indiana vs. Ohio state problem, this time assuming σ_1 and σ_2 are unknown and unequal. Thus, $s_1^2 = 400$ and $s_2^2 = 800$, $n_1 = n_2 = 12$, $\hat{\mu}_1 = 370$, $\hat{\mu}_2 = 400$. Test the first university's claim that it has a lower crime rate using the .01 level of significance.

Solution. The appropriate test statistic is

$$t = \frac{\hat{\mu}_1 - \hat{\mu}_2 - (\mu_1 - \mu_2)_0}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}} = \frac{\hat{\mu}_1 - \hat{\mu}_2}{\sqrt{\frac{400}{12} + \frac{800}{12}}} = \frac{\hat{\mu}_1 - \hat{\mu}_2}{10}$$

Note that the computation of the test statistic is the same as for Case I, except that the sample variances are substituted for the population variances, and the test statistic has a T distribution rather than a Z distribution.

We must now figure out what v equals. For Case III, $v =$

$$\begin{aligned} v &= \frac{\left(\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}\right)^2}{\left(\frac{s_1^2}{N_1}\right)^2 \left(\frac{1}{N_1 - 1}\right) + \left(\frac{s_2^2}{N_2}\right)^2 \left(\frac{1}{N_2 - 1}\right)} \\ &= \frac{\left(\frac{1200}{12}\right)^2}{\left(\frac{400}{12}\right)^2 \left(\frac{1}{12 - 1}\right) + \left(\frac{800}{12}\right)^2 \left(\frac{1}{12 - 1}\right)} = 19.8 = \text{about } 20 \end{aligned}$$

Hence, for $\alpha = .01$, accept H_0 if $T_{20} \geq -2.528$, or, equivalently, accept H_0 if $\hat{\mu}_1 - \hat{\mu}_2 \geq -25.28$.

Since $t_c = -3.0$ (note that the calculations are the same as in Problem 1), reject H_0 .

Note. If you use Welch's formula, you get these d.f.:

$$v = \frac{\left(\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}\right)^2}{\left(\frac{s_1^2}{N_1}\right)^2 \left(\frac{1}{N_1+1}\right) + \left(\frac{s_2^2}{N_2}\right)^2 \left(\frac{1}{N_2+1}\right)} - 2 =$$

$$\frac{\left(\frac{1200}{12}\right)^2}{\left(\frac{400}{12}\right)^2 \left(\frac{1}{12+1}\right) + \left(\frac{800}{12}\right)^2 \left(\frac{1}{12+1}\right)} - 2 = 21.4 = \text{about } 21$$

2. Again work this problem: A professor believes that women do better on her exams than men do. A sample of 8 women ($N_1 = 8$) and 10 men ($N_2 = 10$) yields $\hat{\mu}_1 = 7$, $\hat{\mu}_2 = 5.5$, $s_1^2 = 1$, $s_2^2 = 1.7$. Using $\alpha = .01$, Test whether the female mean is greater than the male mean. DO NOT ASSUME $\sigma_1 = \sigma_2$.

Solution.

The appropriate test statistic is

$$t = \frac{\hat{\mu}_1 - \hat{\mu}_2 - (\mu_1 - \mu_2)_0}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}} = \frac{\hat{\mu}_1 - \hat{\mu}_2}{\sqrt{\frac{1}{8} + \frac{1.7}{10}}} = \frac{.543}{.543}$$

Without going through all the details, the degrees of freedom work out to be 15.9877, or about 16. Hence, accept H_0 if $T_{16} \leq 2.583$, or, equivalently, accept H_0 if $\hat{\mu}_1 - \hat{\mu}_2 \leq 1.403$ since $1.386 = 2.583 * .543$. The computed test stat is

$$t_c = (7 - 5.5) / .543 = 1.5 / .543 = 2.762.$$

The computed T value is too high, so reject H_0 .