

COMPARING TWO MEANS

- We often want to know whether one group is different from another
- * We can analyze this by comparing their means
- × Is \overline{y}_1 different from \overline{y}_2 is equivalent to asking whether $\overline{y}_1 \overline{y}_2$ is different from zero.

COMPARING TWO MEANS

The standard deviation of the difference between two sample means is

 $SD(\overline{y}_1 - \overline{y}_2) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

 We still don't know the true standard deviations of the two groups, so we need to estimate and use the standard error

 $SE\left(\overline{y}_{1}-\overline{y}_{2}\right) = \sqrt{\frac{s_{1}^{2}}{n_{1}} + \frac{s_{2}^{2}}{n_{2}}}$

COMPARING TWO MEANS

 The sampling model is a Student's t.
+ The confidence interval we build is called a twosample t-interval (for the difference in means).
+ The corresponding hypothesis test is called a twosample hypothesis test.

SAMPLING DISTRIBUTION FOR THE DIFFERENCE BETWEEN TWO MEANS

 When the conditions are met, the standardized sample difference between the means of two independent groups

$$t = \frac{\left(\overline{y}_1 - \overline{y}_2\right) - \left(\mu_1 - \mu_2\right)}{SE\left(\overline{y}_1 - \overline{y}_2\right)}$$

can be modeled by a Student's t-model with a number of degrees of freedom found with a special formula. We estimate the standard error with

 $SE(\overline{y}_1 - \overline{y}_2) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

ASSUMPTIONS AND CONDITIONS

- Independence Assumption (Each condition needs to be checked for both groups.):
 - + Randomization Condition: Were the data collected with suitable randomization (representative random samples or a randomized experiment)?
 - + 10% Condition: We don't usually check this condition for differences of means. We will check it for means only if we have a very small population or an extremely large sample.

ASSUMPTIONS AND CONDITIONS

Normal Population Assumption:
+ Nearly Normal Condition: This must be checked for *both* groups. A violation by either one violates the condition.

 Independent Groups Assumption: The two groups we are comparing must be independent of each other. (See Chapter 25 if the groups are not independent of one another...)

TWO-SAMPLE T-INTERVAL

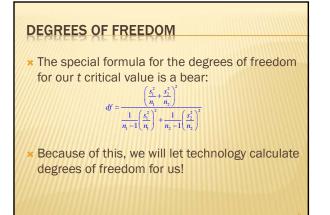
When the conditions are met, we are ready to find the confidence interval for the difference between means of two independent groups.

The confidence interval is

 $\left(\overline{y}_1 - \overline{y}_2\right) \pm t_{df}^* \times SE\left(\overline{y}_1 - \overline{y}_2\right)$

where the standard error of the difference of the means is $SE\left(\overline{y}_1 - \overline{y}_2\right) = \sqrt{\frac{s_1^2}{n_1^2} + \frac{s_2^2}{n_2^2}}$

The critical value depends on the particular confidence level, C, that you specify and on the number of degrees of freedom, which we get from the sample sizes and a special formula.



TESTING THE DIFFERENCE BETWEEN TWO MEANS

- * The hypothesis test we use is the two-sample ttest for means.
- * The conditions for the two-sample *t*-test for the difference between the means of two independent groups are the same as for the two-sample *t*-interval.

TESTING THE DIFFERENCE BETWEEN TWO MEANS

We test the hypothesis H₀: $\mu_1 - \mu_2 = \Delta_0$, where the hypothesized difference, Δ_0 , is almost always 0, using the statistic

$$t = \frac{\left(\overline{y}_1 - \overline{y}_2\right) - \Delta_0}{SE\left(\overline{y}_1 - \overline{y}_2\right)}$$

The standard error is

$$SE(\overline{y}_1 - \overline{y}_2) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

When the conditions are met and the null hypothesis is true, this statistic can be closely modeled by a Student's t-model with a number of degrees of freedom given by a special formula. We use that model to obtain a P-value.

POOLED T-TEST

- × If we are willing to assume that the variances of two means are equal, we can pool the data from two groups to estimate the common variance and make the degrees of freedom formula much simpler.
- We are still estimating the pooled standard deviation from the data, so we use Student's tmodel, and the test is called a pooled t-test.