Independent Samples: Comparing Means–Confidence Intervals

Lecture 39
Section 11.5

Robb T. Koether

Hampden-Sydney College

Wed, Apr 4, 2012

Outline

Outline

The Distribution of $\overline{x}_1 - \overline{x}_2$

- Recall that, $\overline{x}_1 - \overline{x}_2$ has mean and standard deviation

$$\mu_{\overline{x}_1 - \overline{x}_2} = \mu_1 - \mu_2,$$

$$\sigma_{\overline{x}_1 - \overline{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}.$$

- If the samples sizes are large, then $\overline{x_1} - \overline{x_2}$ is normal, in which case

$$z = \frac{(\overline{x}_1 - \overline{x}_1) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}.$$

Outline

When σ_1 and σ_2 are Unknown

- When σ_1 and σ_2 are unknown, we substitute s_1 and s_2 for them.
- Whenever we use s_1 and s_2 instead of σ and the populations are normal, we must use the t distribution instead of the standard normal distribution, *unless the sample sizes are large*.

When σ_1 and σ_2 are Unknown

- The formula for t that we end up with will be *much* simpler if we make one more assumption:

$$\boxed{\text{Assume that } \sigma_1 = \sigma_2.}$$

- To make this assumption, we need evidence.
- Sufficient evidence will be that
  - The values of s_1 and s_2 are close to each other.
  - The histograms have similar widths.
  - The boxplots have similar widths.
- If $\sigma_1 = \sigma_2$, then we will just call the value σ.

Estimating σ

- Under this assumption, we can simplify the formula for $\sigma_{\overline{x}_1-\overline{x}_2}$.

$$\begin{aligned}
\sigma_{\overline{x}_1-\overline{x}_2} &= \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} \\
&= \sqrt{\frac{\sigma^2}{n_1}+\frac{\sigma^2}{n_2}} \\
&= \sqrt{\sigma^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)} \\
&= \sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}
\end{aligned}$$

Estimating σ

- That's fine, except we still do not know σ.
- So...

Estimating σ

- s_1 and s_2 are each estimators for σ.
- We can get a better estimate for σ than either one of these if we "pool" s_1 and s_2 together.
- The pooled estimate for σ is

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

$\overline{x}_1 - \overline{x}_2$ and the t Distribution

- So the test statistic is

$$t = \frac{(\overline{x}_1 - \overline{x}_2) - 0}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

with

$$df = df_1 + df_2 = n_1 + n_2 - 2$$

degrees of freedom.

Outline

Example

Example (Testing $\mu_1 = \mu_2$ using t)

- Suppose we test the new drug on only 8 patients and the old drug on 16 patients.
- We record the number of days until recovery for each individual.
- The results are

New Drug (#1)	Old Drug (#2)
$\overline{x}_1 = 5.2$	$\overline{x}_2 = 6.4$
$s_1 = 1.8$	$s_2 = 2.0$
$n_1 = 8$	$n_2 = 16$

Example

Example (Testing $\mu_1 = \mu_2$ using t)

- The researchers made QQ plots, which indicated that both populations are normal.
- Also, s_1 and s_2 are very close in value, so we may assume that $\sigma_1 = \sigma_2$.
- Test the hypothesis that the new drug is better, at the 1% significance level.

Example

Example (Testing $\mu_1 = \mu_2$ using t)

(1) Let $\mu_1 =$ average time to recovery for the new drug.
Let $\mu_2 =$ average time to recovery for the old drug.
$H_0 : \mu_1 = \mu_2$
$H_1 : \mu_1 < \mu_2$

(2) $\alpha = 0.01$.

Example

Example (Testing $\mu_1 = \mu_2$ using t)

(3) The test statistic is

$$t = \frac{(\overline{x}_1 - \overline{x}_2) - 0}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

with $df = n_1 + n_2 - 2$.

Example

Example (Testing $\mu_1 = \mu_2$ using t)

(4) Compute

$$s_p = \sqrt{\frac{7s_1^2 + 15s_2^2}{22}} = 1.939$$

and

$$t = \frac{(5.2 - 6.4) - 0}{1.939\sqrt{\frac{1}{8} + \frac{1}{16}}} = \frac{-1.2}{0.8394} = -1.430.$$

Example (Testing $\mu_1 = \mu_2$ using t)

(5) The number of degrees of freedom is $df = df_1 + df_2 = 22$, so the p-value is

$$\begin{aligned} p\text{-value} &= P(t_{22} < -1.430) \\ &= \texttt{tcdf(-E99,-1.430,22)} \\ &= 0.0834. \end{aligned}$$

Example

Example (Testing $\mu_1 = \mu_2$ using t)

(6) The p-value is larger than α, so we accept H_0.

(7) The new drug is no more effective than the old drug.

Outline

TI-83 Two-sample *t* test

- Enter the data from the first sample into L_1.
- Enter the data from the second sample into L_2.
- Press `STAT > TESTS`.
- Choose `2-SampTTest`.
- (For large samples, `2-SampZTest` is ok.)
- Choose `Data` or `Stats`.

The TI-83 - Means of Independent Samples

TI-83 Two-sample t test

- Provide the information that is requested.
- `2-SampTTest` will ask whether to use a pooled estimate of σ. Answer yes.
- Select `Calculate` and press `ENTER`.

- Note that you are not asked for the hypothetical difference between μ_1 and μ_2.
- The TI-83 assumes that the null hypothesis is $H_0 : \mu_1 = \mu_2$.
- That is, the hypothetical difference is always 0.

TI-83 Two-sample *t* test

- The display shows, among other things, the value of the test statistic (z or t) and the p-value.
- It also shows, for the t test, the value of the pooled estimate s_p.

An Example

Practice

- Rework the previous examples using the TI-83.

Outline

Assignment

Homework

- Read Section 11.4, pages 695 - 712 (skip confidence intervals).
- Let's Do It! 11.6.
- Exercises 26(omit c), 27(omit d), 28, 29, 31, 32, page 713.