Two Sample Confidence Intervals
Section 21.1, 21.2, 21.3
Lecture 38

Robb T. Koether

Hampden-Sydney College

Wed, Mar 30, 2016

Outline

1. Two Independent Samples
2. The Probability Distribution of $\overline{x}_1 - \overline{x}_2$
   - The Mean
   - The Standard Deviation
   - The Shape
3. Confidence Intervals Concerning $\overline{x}_1 - \overline{x}_2$
4. Assignment

Outline

1. Two Independent Samples
2. The Probability Distribution of $\overline{x}_1 - \overline{x}_2$
   - The Mean
   - The Standard Deviation
   - The Shape
3. Confidence Intervals Concerning $\overline{x}_1 - \overline{x}_2$
4. Assignment

Two Independent Samples

Definition (Independent Samples)

Two samples are independent if there is no logical relationship between the members of one sample and the members of the other.

Two Independent Samples

- Typically, independent samples are used when we want to compare the mean (or any other parameter) of one population to the mean of another population.
- Recall that matched pairs (dependent samples) are used when we want to study the before and after effect of a treatment on individual subjects.

Outline

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

- We wish to study the difference in population means.
- Let Population #1 have mean μ_1 and standard deviation σ_1.
- Let Population #2 have mean μ_2 and standard deviation σ_2.
- So we wish to study $\mu_1 - \mu_2$.

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

- The estimator of $\mu_1 - \mu_2$ is $\overline{x}_1 - \overline{x}_2$.
- So we need to know the sampling distribution of $\overline{x}_1 - \overline{x}_2$.
- That is, we need to know
  - The mean of $\overline{x}_1 - \overline{x}_2$.
  - The standard deviation of $\overline{x}_1 - \overline{x}_2$.
  - The shape of the sampling distribution.

Outline

The Mean

- The mean of $\overline{x}_1 - \overline{x}_2$ is $\mu_1 - \mu_2$.

Outline

The Standard Deviation

- The standard deviation of $\overline{x}_1 - \overline{x}_2$ is more complicated.

The Standard Deviation

- The standard deviation of $\overline{x}_1 - \overline{x}_2$ is more complicated.
- According to the Central Limit theorem, their individual standard deviations are

$$\frac{\sigma_1}{\sqrt{n_1}} \text{ and } \frac{\sigma_2}{\sqrt{n_2}}.$$

The Standard Deviation

- The standard deviation of $\overline{x}_1 - \overline{x}_2$ is more complicated.
- According to the Central Limit theorem, their individual standard deviations are
$$\frac{\sigma_1}{\sqrt{n_1}} \text{ and } \frac{\sigma_2}{\sqrt{n_2}}.$$
- So their variances are
$$\frac{\sigma_1^2}{n_1} \text{ and } \frac{\sigma_2^2}{n_2}.$$

The Standard Deviation

- It turns out that, for any two random variables, the variance of their difference is the sum of their variances.

The Standard Deviation

- It turns out that, for any two random variables, the variance of their difference is the sum of their variances.
- Therefore, the variance of $\overline{x}_1 - \overline{x}_2$ is

$$\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}.$$

The Standard Deviation

- It turns out that, for any two random variables, the variance of their difference is the sum of their variances.
- Therefore, the variance of $\overline{x}_1 - \overline{x}_2$ is

$$\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}.$$

- And the standard deviation of $\overline{x}_1 - \overline{x}_2$ is

$$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}.$$

The Standard Error

- When we use s_1 and s_2 to approximate σ_1 and σ_2, we will calculate the standard error of $\overline{x}_1 - \overline{x}_2$:

$$\mathrm{SE}_{\overline{x}_1-\overline{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}.$$

Outline

The Shape

- If the two populations are normal or almost normal, then the distribution of $\overline{x}_1 - \overline{x}_2$ is normal or almost normal.
- If both sample sizes are large, say $n_1 \geq 100$ and $n_2 \geq 100$, then the Central Limit Theorem guarantees that $\overline{x}_1 - \overline{x}_2$ will be close enough to normal.

The Shape

- However, that is assuming that we know σ_1 and σ_2.
- In practice, we almost always have to use s_1 and s_2 to estimate σ_1 and σ_2.
- Therefore, we will need to use the *t* distribution under the same conditions as described in the previous lecture.
- To describe the shape of the *t* distribution, we must specify the number of degrees of freedom.

Degrees of Freedom

- The textbook relies on software at this point.
- The software presents us with "Option 1" and "Option 2."
- We will instead take the traditional approach.
- The number of degrees of freedom is the total of the degrees of freedom from the two samples.
- That is,

$$\begin{aligned} df &= df_1 + df_2 \\ &= (n_1 - 1) + (n_2 - 1) \\ &= n_1 + n_2 - 2. \end{aligned}$$

Outline

Confidence Intervals Concerning $\overline{x}_1 - \overline{x}_2$

- The general form of a confidence interval:

$$\text{the point estimate} \pm \text{a margin of error.}$$

- In this case, the margin of error is a number of standard errors (SE), depending on the level of confidence.
- Thus, the confidence interval is

$$(\overline{x}_1 - \overline{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}.$$

Example

Example (Confidence Interval – Difference of Means)

- I am trying to choose between two wood stoves.
- I want to estimate the difference in their emission rates.
- I collected 9 measures from Stove #1:

$$1.25, 0.85, 0.44, 1.49, 1.35, 1.50, 0.86, 1.17, 1.52$$

  and 7 measures from Stove #2:

$$1.36, 1.43, 1.24, 1.19, 1.24, 1.78, 1.54$$

- Find a 95% confidence interval for the difference of the means.

Example

Example (Confidence Interval – Difference of Means)

- We have

Sample 1	Sample 2
$\overline{x}_1 = 1.159$	$\overline{x}_2 = 1.397$
$s_1 = 0.3713$	$s_2 = 0.2989$
$n_1 = 9$	$n_2 = 7$

Example

Example (Confidence Interval – Difference of Means)

- The number of degrees of freedom is

$$(9-1)+(7-1)=14.$$

- So, $t^* = \texttt{invT(.025,14)} = -2.145.$
- The confidence interval is

$$\begin{aligned}(\overline{x}_1-\overline{x}_2)\pm t^*\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}} \\ &= (1.159-1.397)\pm(2.145)\sqrt{\frac{0.3713^2}{9}+\frac{0.2989^2}{7}} \\ &= -0.238\pm 0.359.\end{aligned}$$

Outline

1. Two Independent Samples
2. The Probability Distribution of $\overline{x}_1 - \overline{x}_2$
   - The Mean
   - The Standard Deviation
   - The Shape
3. Confidence Intervals Concerning $\overline{x}_1 - \overline{x}_2$
4. Assignment

Assignment

Assignment

- Read Section 21.1, 21.2, 21.3.
- Apply Your Knowledge: 1, 2, 3, 4.
- Check Your Skills: 18, 19, 20, 24.
- Exercises 35, 39.