Independent Samples: Comparing Means

Lecture 38
Section 11.4

Robb T. Koether

Hampden-Sydney College

Tue, Apr 3, 2012

Outline

Outline

The Sum of Two Distributions

- Run the program `Distribution of Difference.exe`.
- Note the mean, standard deviation, and variance of
  - x_1
  - x_2
  - $x_1 - x_2$
- Do you see any patterns?

Outline

The Difference of Two Distributions

- It turns out that

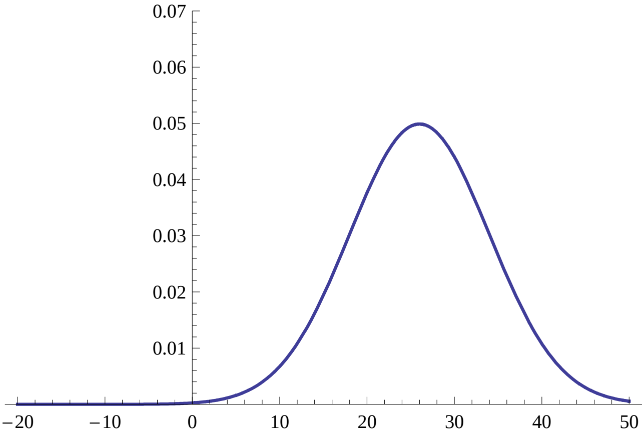
$$\mu_{x_1 - x_2} = \mu_1 - \mu_2$$
$$\sigma^2_{x_1 - x_2} = \sigma_1^2 + \sigma_2^2$$
$$\sigma_{x_1 - x_2} = \sqrt{\sigma_1^2 + \sigma_2^2}$$

- Furthermore, if x_1 and x_2 are both normal, then $x_1 - x_2$ is also normal.

The Sum of Two Distributions

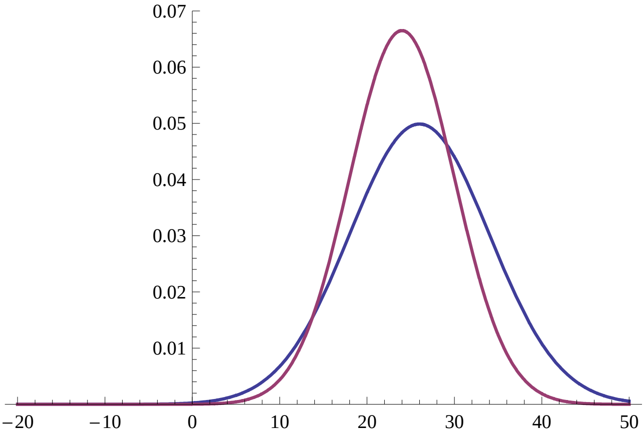
- Consider the average age at which a male first marries (x_1) vs. the average age at which a female first marries (x_2). (μ_1 vs. μ_2.)
- Suppose that x_1 is $N(26, 8)$ and that x_2 is $N(24, 6)$.

The Distribution of x_1



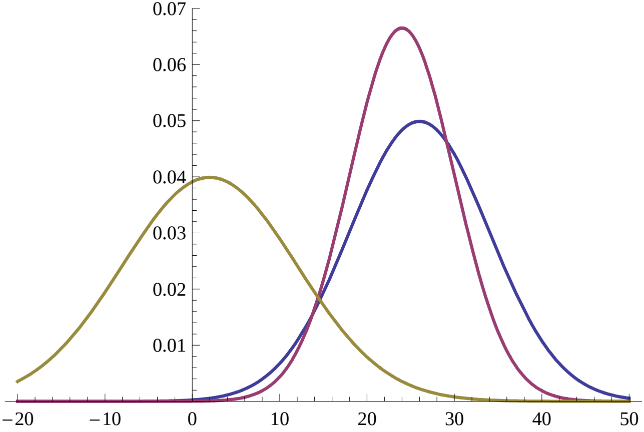
- Age when males first marry (x_1) is $N(26, 8)$.

The Distribution of x_2



- Age when females first marry (x_2) is $N(24, 6)$.

The Distribution of $x_2 - x_2$



- Difference between ages $(x_1 - x_2)$ is $N(2, 10)$.

Outline

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

- Now let's consider two populations.
- Population 1 has mean μ_1 and standard deviation σ_1.
- Population 2 has mean μ_2 and standard deviation σ_2.
- We wish to compare μ_1 and μ_2.
- We do so by taking samples and comparing sample means $\overline{x}_1$ and $\overline{x}_2$.
- This means that we need to know the distribution of $\overline{x}_1 - \overline{x}_2$.

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

- For large sample sizes, we know that

$$\overline{x}_1 \text{ is } N\left(\mu_1, \frac{\sigma_1}{\sqrt{n_1}}\right)$$

and

$$\overline{x}_2 \text{ is } N\left(\mu_2, \frac{\sigma_2}{\sqrt{n_2}}\right)$$

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

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

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

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

- Also, $\overline{x_1} - \overline{x_2}$ is normal, so

$$\overline{x}_1 - \overline{x}_1 \text{ is } N\left(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right).$$

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

- Therefore (for large samples), the test statistic will be

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

Whether to use z or t

Which Statistic?

- If the sample size is large, then it does not matter whether we use z or t.
- However, the best rule to follow is to use z when σ_1 and σ_2 are known and use t when σ_1 and σ_2 are unknown.

Whether to use z or t

Which Statistic?

- If the sample size is small, then it does matter and it is necessary to know that both populations are normal.
- Use z if σ_1 and σ_2 are known and use t if σ_1 and σ_2 are unknown.

Outline

Example

Example (Testing hypotheses concerning $\mu_1 - \mu_2$)

- I bought a book of Sudoku puzzles.

Example

Example (Testing hypotheses concerning $\mu_1 - \mu_2$)

- I bought a book of Sudoku puzzles.
- The puzzles were divided into 4 levels of difficulty.

Example

Example (Testing hypotheses concerning $\mu_1 - \mu_2$)

- I bought a book of Sudoku puzzles.
- The puzzles were divided into 4 levels of difficulty.
- Were the Level 2 puzzles really harder than the Level 1 puzzles?

Example

Example (Testing hypotheses concerning $\mu_1 - \mu_2$)

- I bought a book of Sudoku puzzles.
- The puzzles were divided into 4 levels of difficulty.
- Were the Level 2 puzzles really harder than the Level 1 puzzles?
- My average times for Levels 1 and 2 were

Level 1	Level 2
$\overline{x}_1 = 3.02$	$\overline{x}_2 = 4.70$
$n_1 = 30$	$n_2 = 30$

Example

Example (Testing hypotheses concerning $\mu_1 - \mu_2$)

- I bought a book of Sudoku puzzles.
- The puzzles were divided into 4 levels of difficulty.
- Were the Level 2 puzzles really harder than the Level 1 puzzles?
- My average times for Levels 1 and 2 were

Level 1	Level 2
$\overline{x}_1 = 3.02$	$\overline{x}_2 = 4.70$
$n_1 = 30$	$n_2 = 30$

- Assume that $\sigma_1 = 0.85$ and $\sigma_2 = 0.95$.

Example

Example (Testing hypotheses concerning $\mu_1 - \mu_2$)

- I bought a book of Sudoku puzzles.
- The puzzles were divided into 4 levels of difficulty.
- Were the Level 2 puzzles really harder than the Level 1 puzzles?
- My average times for Levels 1 and 2 were

Level 1	Level 2
$\overline{x}_1 = 3.02$	$\overline{x}_2 = 4.70$
$n_1 = 30$	$n_2 = 30$

- Assume that $\sigma_1 = 0.85$ and $\sigma_2 = 0.95$.
- Test the hypotheses at the 5% level of significance.

Example

Example (Testing hypotheses concerning $\mu_1 - \mu_2$)

(1) $\mu_1 =$ average time to solve a level 1 puzzle.
$\mu_2 =$ average time to solve a level 2 puzzle.
$H_0 : \mu_1 = \mu_2$.
$H_1 : \mu_1 < \mu_2$.

Example

Example (Testing hypotheses concerning $\mu_1 - \mu_2$)

(1) $\mu_1 =$ average time to solve a level 1 puzzle.

$\mu_2 =$ average time to solve a level 2 puzzle.

$H_0 : \mu_1 = \mu_2$.

$H_1 : \mu_1 < \mu_2$.

(2) $\alpha = 0.05$.

Example

Example (Testing hypotheses concerning $\mu_1 - \mu_2$)

(1) $\mu_1 =$ average time to solve a level 1 puzzle.
$\mu_2 =$ average time to solve a level 2 puzzle.
$H_0 : \mu_1 = \mu_2$.
$H_1 : \mu_1 < \mu_2$.

(2) $\alpha = 0.05$.

(3) The test statistic:

$$z = \frac{(\overline{x_1} - \overline{x_2}) - 0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}.$$

Example

Example (Testing hypotheses concerning $\mu_1 - \mu_2$)

(4) Compute z:

$$z = \frac{(3.02 - 4.70) - 0}{\sqrt{\frac{0.85^2}{30} + \frac{0.95^2}{30}}} = \frac{-1.68}{0.2327} = -7.256.$$

Example

Example (Testing hypotheses concerning $\mu_1 - \mu_2$)

(4) Compute z:

$$\begin{aligned} z &= \frac{(3.02 - 4.70) - 0}{\sqrt{\frac{0.85^2}{30} + \frac{0.95^2}{30}}} \\ &= \frac{-1.68}{0.2327} \\ &= -7.256. \end{aligned}$$

(5) p-value $=$ `normalcdf(-E99,-7.218)` $= 0$.

Example

Example (Testing hypotheses concerning $\mu_1 - \mu_2$)

(6) Reject H_0.

Example

Example (Testing hypotheses concerning $\mu_1 - \mu_2$)

(6) Reject H_0.

(7) The average time to solve a Level 1 puzzle is less than the average time to solve a Level 2 puzzle. That is, the Level 2 puzzles are harder than the Level 1 puzzles.

Example

Example (Testing hypotheses concerning $\mu_1 - \mu_2$)

- Is Level 3 harder than Level 2?

Example

Example (Testing hypotheses concerning $\mu_1 - \mu_2$)

- Is Level 3 harder than Level 2?
- My average times for Levels 2 and 3 were

Level 2	Level 3
$\overline{x}_1 = 4.70$	$\overline{x}_2 = 6.61$
$n_1 = 30$	$n_2 = 30$

Example

Example (Testing hypotheses concerning $\mu_1 - \mu_2$)

- Is Level 3 harder than Level 2?
- My average times for Levels 2 and 3 were

Level 2	Level 3
$\overline{x}_1 = 4.70$	$\overline{x}_2 = 6.61$
$n_1 = 30$	$n_2 = 30$

Example

Example (Testing hypotheses concerning $\mu_1 - \mu_2$)

- Is Level 3 harder than Level 2?
- My average times for Levels 2 and 3 were

Level 2	Level 3
$\overline{x}_1 = 4.70$	$\overline{x}_2 = 6.61$
$n_1 = 30$	$n_2 = 30$

Example

Example (Testing hypotheses concerning $\mu_1 - \mu_2$)

- Is Level 3 harder than Level 2?
- My average times for Levels 2 and 3 were

Level 2	Level 3
$\overline{x}_1 = 4.70$	$\overline{x}_2 = 6.61$
$n_1 = 30$	$n_2 = 30$

- Assume that $\sigma_1 = 0.95$ and $\sigma_2 = 1.70$.

Example

Example (Testing hypotheses concerning $\mu_1 - \mu_2$)

- Is Level 3 harder than Level 2?
- My average times for Levels 2 and 3 were

Level 2	Level 3
$\overline{x}_1 = 4.70$	$\overline{x}_2 = 6.61$
$n_1 = 30$	$n_2 = 30$

- Assume that $\sigma_1 = 0.95$ and $\sigma_2 = 1.70$.
- Test the hypotheses at the 5% level of significance.

Example

Example (Testing hypotheses concerning $\mu_1 - \mu_2$)

- A new drug is introduced. Is it better than the old drug?
- A group of 40 patients was given the new drug and a group of 60 patients was given the old drug.
- Time until recovery (in days) was measured for each patient.

New Drug (# 1)	Old Drug (# 2)
$n_1 = 40$	$n_2 = 60$
$\overline{x}_1 = 5.4$	$\overline{x}_2 = 6.8$

- Assume that $\sigma_1 = 1.8$ and $\sigma_2 = 1.3$.
- Test the hypotheses at the 5% level of significance.

Example

Example (Testing hypotheses concerning $\mu_1 - \mu_2$)

(1) $\mu_1 =$ average time to recovery for the new drug.
$\mu_2 =$ average time to recovery for the old drug.
$H_0 : \mu_1 - \mu_2 = 0.$
$H_1 : \mu_1 - \mu_2 < 0.$

(2) $\alpha = 0.05.$

(3) The test statistic:

$$z = \frac{(\overline{x_1} - \overline{x_2}) - 0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}.$$

Example

Example (Testing hypotheses concerning $\mu_1 - \mu_2$)

(4) Compute z:

$$z = \frac{(5.4 - 6.8) - 0}{\sqrt{\frac{1.8^2}{40} + \frac{1.3^2}{60}}} = \frac{-1.4}{0.3304} = -4.237.$$

(5) p-value $=$ `normalcdf(-E99,-4.237)` $= 1.132 \times 10^{-5}$.

Example

Example (Testing hypotheses concerning $\mu_1 - \mu_2$)

(6) Reject H_0.

(7) The average time to recovery for the new drug is less than it is for the old drug. That is, the new drug is more effective than the old drug.

Outline

The TI-83 - Means of Independent Samples

TI-83 Two-sample z-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 either `2-SampZTest`.
- Choose `Data` or `Stats`.

The TI-83 - Means of Independent Samples

TI-83 Two-sample z or t test

- Provide the information that is requested.
- 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.

The TI-83 - Means of Independent Samples

TI-83 Two-sample z-test

- The display shows, among other things, the value of the test statistic z and the p-value.

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 postponed to Monday.