Independent Samples: Comparing Means

Lecture 37
Sections 11.1, 11.2, 11.4

Robb T. Koether

Hampden-Sydney College

Mon, Mar 31, 2008

Outline

Introduction

- When two samples are taken from two different populations, they may be taken independently or not independently.
- When they are not independent, the data are usually paired and we study the difference between the pairs.
- When they are independent, the best we can do is study the difference between the averages of the samples.
- We will study only the independent samples.
- In this lecture, we will learn how to test a hypothesis concerning the difference between the population means.
- We will also learn how to perform the test on the TI-83.

Dependent Samples

- In a paired study, two observations are made on each subject, producing one sample of bivariate data.
- Or we could think of it as two samples of paired data.
- Paired data are often “before" and “after" observations.
- By comparing the mean before treatment to the mean after treatment, we can determine whether the treatment had an effect.

Independent Samples

- On the other hand, with independent samples, there is no logical way to “pair" the data.
- One sample might be from a population of males and the other from a population of (unrelated) females.
  - Of course, males and females could be paired if they were twins or husband and wife.
- Or one might be the treatment group and the other the control group.
- Furthermore, the independent samples could be of different sizes.
  - Paired samples must be of the same size.

The Estimator of $\mu_1 - \mu_2$

- We start with 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 $\overline{x}_2$.

The Estimator of $\mu_1 - \mu_2$

- We will use $\overline{x}_1 - \overline{x}_2$ as an estimator of $\mu_1 - \mu_2$.
- If we want to know whether $\mu_1 = \mu_2$, we test to see whether $\mu_1 - \mu_2 = 0$ by computing $\overline{x}_1 - \overline{x}_2$ and comparing it to 0.

The Distributions of $\overline{x}_1$ and $\overline{x}_2$

- Let n_1 and n_2 be the sample sizes.
- If the samples are large, then $\overline{x}_1$ and $\overline{x}_1$ have (approx.) normal distributions.
- However, if either sample is small, then we will need an additional assumption:

  The population of the small sample(s) is normal.

  in order to use the t distribution.

Further Assumption

- We will also assume that the two populations have the same standard deviation.
- Call it σ.
- That is, $\sigma = \sigma_1 = \sigma_2$.
- If this assumption is not supported by the evidence, then it should not be made.
- If this assumption is not made, then the formulas become *much* more complicated. See p. 658.

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

- If the sample sizes are large enough (or the populations are normal), then according to the Central Limit Theorem,
- $\overline{x}_1$ has a normal distribution with mean μ_1 and standard deviation $\dfrac{\sigma_1}{\sqrt{n_1}}$.
- $\overline{x}_2$ has a normal distribution with mean μ_2 and standard deviation $\dfrac{\sigma_2}{\sqrt{n_2}}$.

Some Statistical Facts

1. For any two random variables X and Y

$$\begin{aligned}\mu_{X+Y} &= \mu_X + \mu_Y \\ \sigma^2_{X+Y} &= \sigma^2_X + \sigma^2_Y \\ \sigma_{X+Y} &= \sqrt{\sigma^2_X + \sigma^2_Y}\end{aligned}$$

2. If X and Y are both normal $X + Y$ is also normal.

Some More Statistical Facts

1. For the difference $X - Y$, the situation is very similar.
2. For any two random variables X and Y

$$\begin{aligned} \mu_{X-Y} &= \mu_X - \mu_Y \\ \sigma^2_{X-Y} &= \sigma^2_X + \sigma^2_Y \\ \sigma_{X-Y} &= \sqrt{\sigma^2_X + \sigma^2_Y} \end{aligned}$$

3. If X and Y are both normal $X - Y$ is also normal.

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

- It follows from theory that $\overline{x}_1 - \overline{x}_2$ is normal with
  - Mean

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

  - Variance

$$\sigma^2_{\overline{x}_1 - \overline{x}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

  - Standard deviation

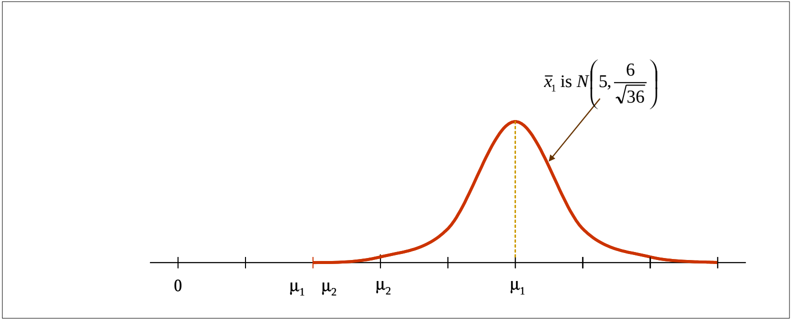
$$\sigma_{\overline{x}_1 - \overline{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

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

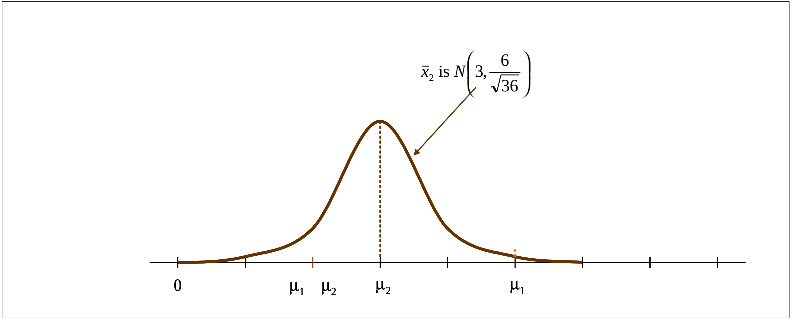
- If we assume that $\sigma_1 = \sigma_2$, (call it σ), then the standard deviation may be simplified to

$$\sigma_{\overline{x}_1 - \overline{x}_2} = \sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}} = \sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

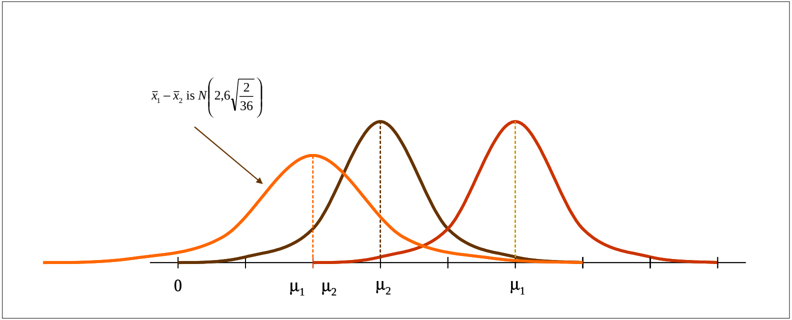
The Distribution of $\overline{x}_1$



The Distribution of $\overline{x}_2$



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



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

- If $\overline{x}_1 - \overline{x}_2$ is normal with mean

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

and standard deviation

$$\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}},$$

then it follows that

$$Z = \frac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Example

- Work exercise 11.32 on page 716 under the assumption that $\sigma = 6$ for both populations.
- Which route to work is shorter, Route 1 or Route 2?

Route 1	Route 2
$n_1 = 40$ $\overline{x}_1 = 31.945$	$n_2 = 40$ $\overline{x}_2 = 28.105$

- Assume that $\sigma = 6$.
- Test hypotheses at 5% level.

Summary

Summary

- In dependent samples, the data are usually paired and we study the difference between the pairs.
- In independent samples, we study the difference between the sample means.
- The statistic $\overline{x}_1 - \overline{x}_2$ has a normal distribution if the populations are normal or if the sample sizes are large enough.
- Under the simplest circumstances, the statistic is

$$Z = \frac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}.$$