Statistics for Business and Economics

8th Edition

Chapter 8

Estimation: Additional Topics

Chapter Goals

After completing this chapter, you should be able to:

- Form confidence intervals for the difference between two means from dependent samples
- Form confidence intervals for the difference between two independent population means (standard deviations known or unknown)
- Compute confidence interval limits for the difference between two independent population proportions

Estimation: Additional Topics

Chapter Topics

Confidence Intervals

- Population Means, Dependent Samples
- Population Means, Independent Samples
- Population Proportions

Examples:

- Same group before vs. after treatment
- Group 1 vs. independent Group 2
- Proportion 1 vs. Proportion 2

8.1

Dependent Samples

Dependent samples

Confidence Interval Estimation of the Difference Between Two Normal Population Means: Dependent Samples

Tests Means of 2 Related Populations

- Paired or matched samples
- Repeated measures (before/after)
- Use difference between paired values:

$$d_i = x_i - y_i$$

- Eliminates Variation Among Subjects
- Assumptions:
  - Both Populations Are Normally Distributed

Mean Difference

Dependent samples

The i^{th} paired difference is d_i , where

$$d_i = x_i - y_i$$

The point estimate for the population mean paired difference is $\bar{d}$:

$$\bar{d} = \frac{\sum_{i=1}^{n} d_i}{n}$$

The sample standard deviation is:

$$S_d = \sqrt{\frac{\sum_{i=1}^{n} (d_i - \bar{d})^2}{n-1}}$$

n is the number of matched pairs in the sample

Confidence Interval for Mean Difference

Dependent samples

The confidence interval for the difference between two population means, μ_d , is

$$\bar{d} \pm t_{n-1,\alpha/2} \frac{S_d}{\sqrt{n}}$$

Where
n = the sample size
(number of matched pairs in the paired sample)

Confidence Interval for Mean Difference

(continued)

Dependent samples

- The margin of error is

$$ME = t_{n-1,\alpha/2} \frac{s_d}{\sqrt{n}}$$

- $t_{n-1,\alpha/2}$ is the value from the Student's t distribution with (n – 1) degrees of freedom for which

$$P(t_{n-1} > t_{n-1,\alpha/2}) = \frac{\alpha}{2}$$

Paired Samples Example

Dependent samples

- Six people sign up for a weight loss program. You collect the following data:

Person	Weight: Before (x)	Weight: After (y)	Difference, d_i
1	136	125	11
2	205	195	10
3	157	150	7
4	138	140	- 2
5	175	165	10
6	166	160	6
			42

$$\bar{d} = \frac{\sum d_i}{n}$$

$$= 7.0$$

$$S_d = \sqrt{\frac{\sum (d_i - \bar{d})^2}{n-1}}$$

$$= 4.82$$

Paired Samples Example

(continued)

Dependent samples

- For a 95% confidence level, the appropriate t value is $t_{n-1,\alpha/2} = t_{5,.025} = 2.571$
- The 95% confidence interval for the difference between means, μ_d , is

$$\bar{d} \pm t_{n-1,\alpha/2} \frac{S_d}{\sqrt{n}}$$

$$7 \pm (2.571) \frac{4.82}{\sqrt{6}}$$

$$-1.94 < \mu_d < 12.06$$

Since this interval contains zero, we cannot be 95% confident, given this limited data, that the weight loss program helps people lose weight

Difference Between Two Means: Independent Samples

8.2

Population means, independent samples

Confidence Interval Estimation of the Difference Between Two Normal Population Means: Independent Samples

Goal: Form a confidence interval for the difference between two population means, $\mu_x - \mu_y$

Difference Between Two Means: Independent Samples

(continued)

Population means, independent samples

Goal: Form a confidence interval for the difference between two population means, $\mu_x - \mu_y$

- Different data sources
  - Unrelated
  - Independent
    - Sample selected from one population has no effect on the sample selected from the other population
- The point estimate is the difference between the two sample means:

$$\bar{x} - \bar{y}$$

Difference Between Two Means: Independent Samples

(continued)

Population means, independent samples

- σ_x^2 and σ_y^2 known → Confidence interval uses $z_{\alpha/2}$
- σ_x^2 and σ_y^2 unknown
  - σ_x^2 and σ_y^2 assumed equal
  - σ_x^2 and σ_y^2 assumed unequal

  → Confidence interval uses a value from the Student's **t** distribution

σ_x^2 and σ_y^2 Known

Population means, independent samples

- σ_x^2 and σ_y^2 known *
- σ_x^2 and σ_y^2 unknown

Assumptions:

- Samples are randomly and independently drawn
- both population distributions are normal
- Population variances are known

σ_x^2 and σ_y^2 Known

(continued)

Population means, independent samples

- σ_x^2 and σ_y^2 known *
- σ_x^2 and σ_y^2 unknown

When σ_x and σ_y are known and both populations are normal, the variance of $\bar{X} - \bar{Y}$ is

$$\sigma^2_{\bar{X}-\bar{Y}} = \frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}$$

…and the random variable

$$Z = \frac{(\bar{x} - \bar{y}) - (\mu_X - \mu_Y)}{\sqrt{\dfrac{\sigma_x^2}{n_X} + \dfrac{\sigma_y^2}{n_Y}}}$$

has a standard normal distribution

Confidence Interval, σ_x^2 and σ_y^2 Known

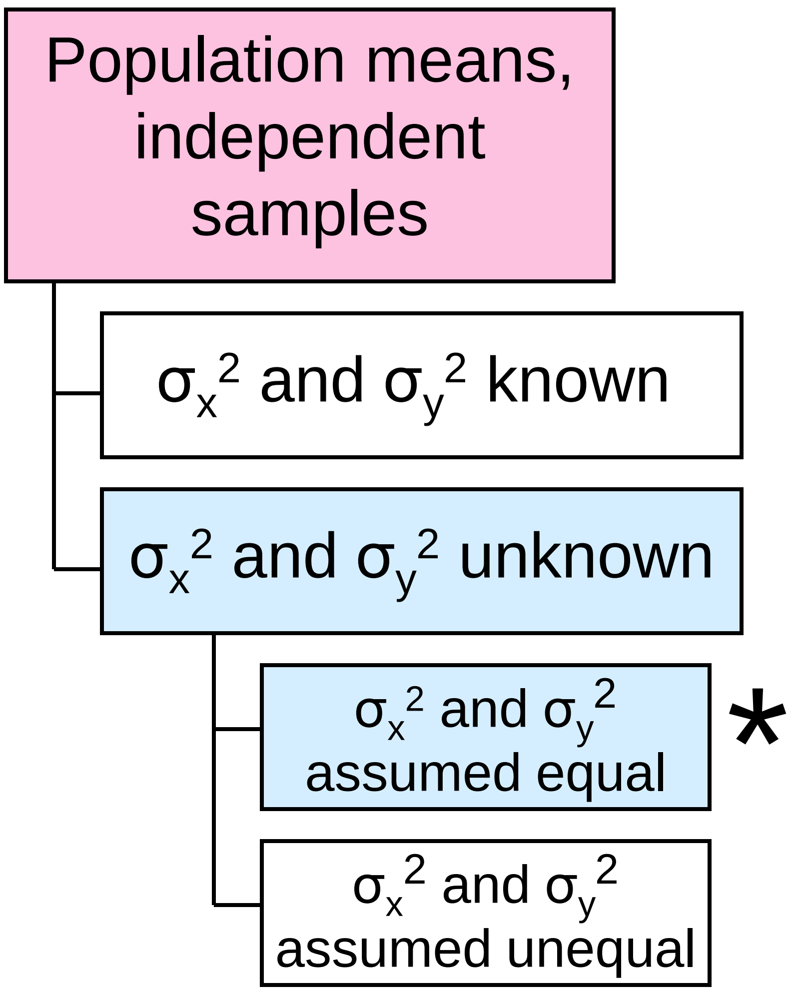
Population means, independent samples

- σ_x^2 and σ_y^2 known
- σ_x^2 and σ_y^2 unknown

* The confidence interval for $\mu_x - \mu_y$ is:

$$(\bar{x} - \bar{y}) \pm z_{\alpha/2} \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}$$

σ_x^2 and σ_y^2 Unknown, Assumed Equal



Assumptions:

- Samples are randomly and independently drawn
- Populations are normally distributed
- Population variances are unknown but assumed equal

σ_x^2 and σ_y^2 Unknown, Assumed Equal

(continued)

- Population means, independent samples
  - σ_x^2 and σ_y^2 known
  - σ_x^2 and σ_y^2 unknown
    - σ_x^2 and σ_y^2 assumed equal *
    - σ_x^2 and σ_y^2 assumed unequal

Forming interval estimates:

- The population variances are assumed equal, so use the two sample standard deviations and pool them to estimate σ
- use a t value with $(n_x + n_y – 2)$ degrees of freedom

σ_x^2 and σ_y^2 Unknown, Assumed Equal

(continued)

- Population means, independent samples
  - σ_x^2 and σ_y^2 known
  - σ_x^2 and σ_y^2 unknown
    - σ_x^2 and σ_y^2 assumed equal *
    - σ_x^2 and σ_y^2 assumed unequal

The pooled variance is

$$s_p^2 = \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2}$$

Confidence Interval, σ_x^2 and σ_y^2 Unknown, Equal

- σ_x^2 and σ_y^2 unknown
  - σ_x^2 and σ_y^2 assumed equal *
  - σ_x^2 and σ_y^2 assumed unequal

The confidence interval for $\mu_1 - \mu_2$ is:

$$(\bar{x} - \bar{y}) \pm t_{n_x + n_y - 2, \alpha/2} \sqrt{\frac{s_p^2}{n_x} + \frac{s_p^2}{n_y}}$$

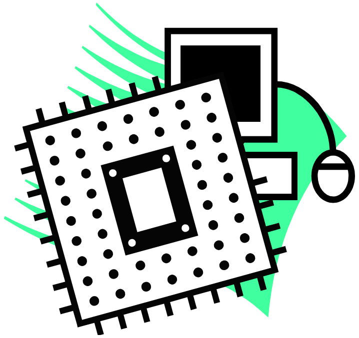
Where

$$s_p^2 = \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2}$$

Pooled Variance Example

You are testing two computer processors for speed. Form a confidence interval for the difference in CPU speed. You collect the following speed data (in Mhz):

	CPU_x	CPU_y
Number Tested	**17**	**14**
Sample mean	**3004**	**2538**
Sample std dev	**74**	**56**



Assume both populations are normal with equal variances, and use 95% confidence

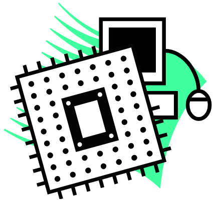
Calculating the Pooled Variance

The pooled variance is:

$$S_p^2 = \frac{(n_x - 1)S_x^{\ 2} + (n_y - 1)S_y^{\ 2}}{(n_x - 1) + (n_y - 1)} = \frac{(17 - 1)74^2 + (14 - 1)56^2}{(17 - 1) + (14 - 1)} = 4427.03$$

The t value for a 95% confidence interval is:

$$t_{n_x + n_y - 2,\, \alpha/2} = t_{29,\, 0.025} = 2.045$$



Calculating the Confidence Limits

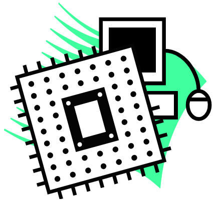
- The 95% confidence interval is

$$(\bar{x} - \bar{y}) \pm t_{n_x + n_y - 2, \alpha/2} \sqrt{\frac{s_p^2}{n_x} + \frac{s_p^2}{n_y}}$$

$$(3004 - 2538) \pm (2.054) \sqrt{\frac{4427.03}{17} + \frac{4427.03}{14}}$$

$$416.69 < \mu_X - \mu_Y < 515.31$$

We are 95% confident that the mean difference in CPU speed is between 416.69 and 515.31 Mhz.



σ_x^2 and σ_y^2 Unknown, Assumed Unequal

- Population means, independent samples
  - σ_x^2 and σ_y^2 known
  - σ_x^2 and σ_y^2 unknown
    - σ_x^2 and σ_y^2 assumed equal
    - σ_x^2 and σ_y^2 assumed unequal *

Assumptions:

- Samples are randomly and independently drawn
- Populations are normally distributed
- Population variances are unknown and assumed unequal

σ_x^2 and σ_y^2 Unknown, Assumed Unequal

(continued)

Population means, independent samples

- σ_x^2 and σ_y^2 known
- σ_x^2 and σ_y^2 unknown
  - σ_x^2 and σ_y^2 assumed equal
  - σ_x^2 and σ_y^2 assumed unequal *

Forming interval estimates:

- The population variances are assumed unequal, so a pooled variance is not appropriate
- use a t value with ν degrees of freedom, where

$$\nu = \frac{\left[\left(\frac{s_x^2}{n_x}\right) + \left(\frac{s_y^2}{n_y}\right)\right]^2}{\left(\frac{s_x^2}{n_x}\right)^2 / (n_x - 1) + \left(\frac{s_y^2}{n_y}\right)^2 / (n_y - 1)}$$

Confidence Interval, σ_x^2 and σ_y^2 Unknown, Unequal

- σ_x^2 and σ_y^2 unknown
  - σ_x^2 and σ_y^2 assumed equal
  - σ_x^2 and σ_y^2 assumed unequal *

The confidence interval for $\mu_1 - \mu_2$ is:

$$(\bar{x} - \bar{y}) \pm t_{\nu,\alpha/2} \sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}$$

Where

$$\nu = \frac{\left[\left(\frac{s_x^2}{n_x}\right) + \left(\frac{s_y^2}{n_y}\right)\right]^2}{\left(\frac{s_x^2}{n_x}\right)^2 / (n_x - 1) + \left(\frac{s_y^2}{n_y}\right)^2 / (n_y - 1)}$$

8.3

Two Population Proportions

Population proportions

Confidence Interval Estimation of the Difference Between Two Population Proportions (Large Samples)

Goal: Form a confidence interval for the difference between two population proportions, $P_x - P_y$

Two Population Proportions

Population proportions

Goal: Form a confidence interval for the difference between two population proportions, $P_x - P_y$

Assumptions:

Both sample sizes are large (generally at least 40 observations in each sample)

The point estimate for the difference is $\hat{p}_x - \hat{p}_y$

Two Population Proportions

(continued)

Population proportions

- The random variable

$$Z = \frac{(\hat{p}_x - \hat{p}_y) - (p_x - p_y)}{\sqrt{\frac{\hat{p}_x(1-\hat{p}_x)}{n_x} + \frac{\hat{p}_y(1-\hat{p}_y)}{n_y}}}$$

is approximately normally distributed

Confidence Interval for Two Population Proportions

Population proportions

The confidence limits for $P_x - P_y$ are:

$$(\hat{p}_x - \hat{p}_y) \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_x(1-\hat{p}_x)}{n_x} + \frac{\hat{p}_y(1-\hat{p}_y)}{n_y}}$$

Example: Two Population Proportions

Form a 90% confidence interval for the difference between the proportion of men and the proportion of women who have college degrees.

- In a random sample, 26 of 50 men and 28 of 40 women had an earned college degree

Example: Two Population Proportions

(continued)

Men: $\hat{p}_x = \dfrac{26}{50} = 0.52$

Women: $\hat{p}_y = \dfrac{28}{40} = 0.70$

$$\sqrt{\frac{\hat{p}_x(1-\hat{p}_x)}{n_x} + \frac{\hat{p}_y(1-\hat{p}_y)}{n_y}} = \sqrt{\frac{0.52(0.48)}{50} + \frac{0.70(0.30)}{40}} = 0.1012$$

For 90% confidence, $Z_{\alpha/2} = 1.645$

Example: Two Population Proportions

(continued)

The confidence limits are:

$$(\hat{p}_x - \hat{p}_y) \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}_x(1-\hat{p}_x)}{n_x}+\frac{\hat{p}_y(1-\hat{p}_y)}{n_y}}$$

$$= (.52 - .70) \pm 1.645\,(0.1012)$$

so the confidence interval is

$$-0.3465 < P_x - P_y < -0.0135$$

Since this interval does not contain zero we are 90% confident that the two proportions are not equal

Chapter Summary

- Compared two dependent samples (paired samples)
  - Formed confidence intervals for the paired difference
- Compared two independent samples
  - Formed confidence intervals for the difference between two means, population variance known, using z
  - Formed confidence intervals for the differences between two means, population variance unknown, using t
- Formed confidence intervals for the differences between two population proportions