Statistics for Business and Economics

Chapter 8

Estimation: Additional Topics

Difference Between Two Means: Independent Samples

8.2

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:

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

σ_x^2 and σ_y^2 Unknown, Assumed Equal

- 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 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:

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

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

Standardized tests are taken by students from large (x) and small (y) high schools. Form a confidence interval for the difference in scores. You collect the following data:

	Score_x	Score_y
Number Obs.	**9**	**15**
Sample mean	**81.31**	**78.61**
Sample var.	**60.76**	**48.24**

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

Calculating the Pooled Variance

The pooled variance is:

$$S_p^2 = \frac{(n_x - 1)S_x^2 + (n_y - 1)S_y^2}{(n_x - 1) + (n_y - 1)} = \frac{8 \times 60.76 + 14 \times 48.24}{22} = 52.79$$

The t value for a 95% confidence interval is:

$$t_{n_x + n_y - 2,\ \alpha/2} = t_{22,\ 0.025} = 2.074$$

Calculating the Confidence Limits

- The 95% confidence interval is

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

$$(81.31 - 78.61) \pm (2.074) \sqrt{\frac{52.79}{9} + \frac{52.79}{15}}$$

$$-3.65 < \mu_X - \mu_Y < 9.05$$

We are 95% confident that the mean difference in scores is between -3.65 and 9.05.

8.3

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

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

Two Population Proportions

(continued)

Population proportions

- The random variable

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

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

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

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

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