Section 8.2 Difference Between Two Means: Independent Samples

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 (1 of 2)

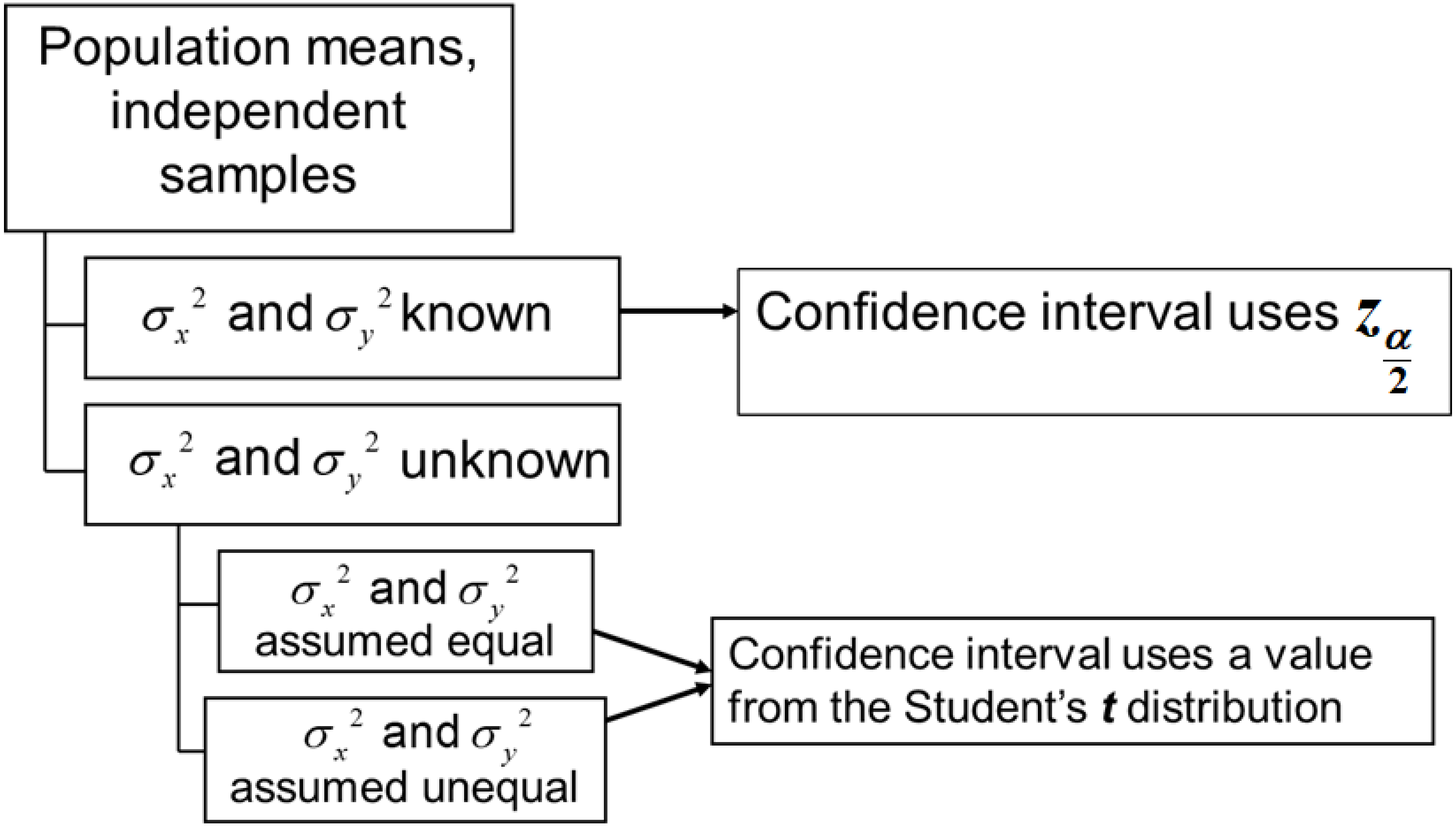
Population means, independent samples

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

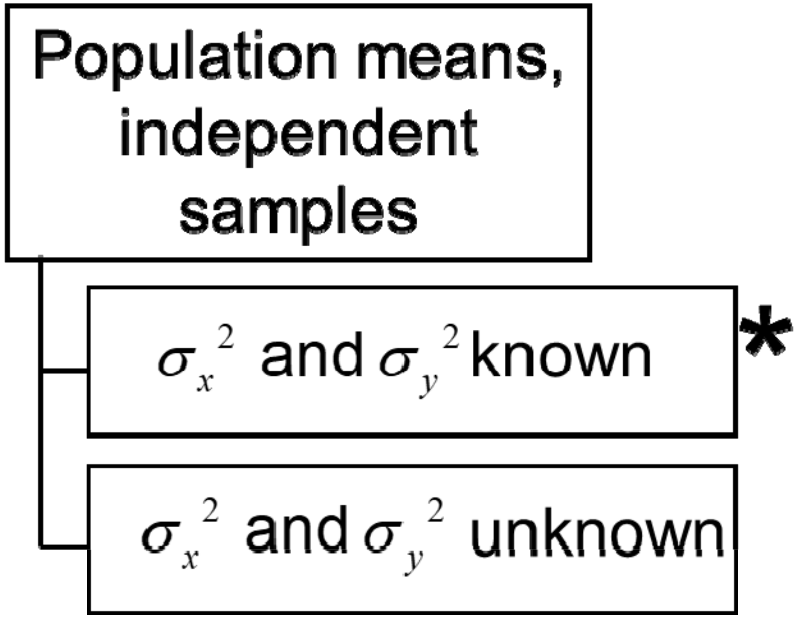
- Different data sources
  - Unrelated
  - Independent
    - Sample selected from one population has no effect on the sample selected from the other population
- The point estimate is the difference between the two sample means:

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

Difference Between Two Means: Independent Samples (2 of 2)



Sigma Sub *x* Squared and Sigma Sub *y* Squared Known (1 of 2)



Assumptions:

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

Sigma Sub *x* Squared and Sigma Sub *y* Squared Known (2 of 2)

- 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

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

has a standard normal distribution

Confidence Interval, Sigma Sub *x* Squared and Sigma Sub *y* Squared Known

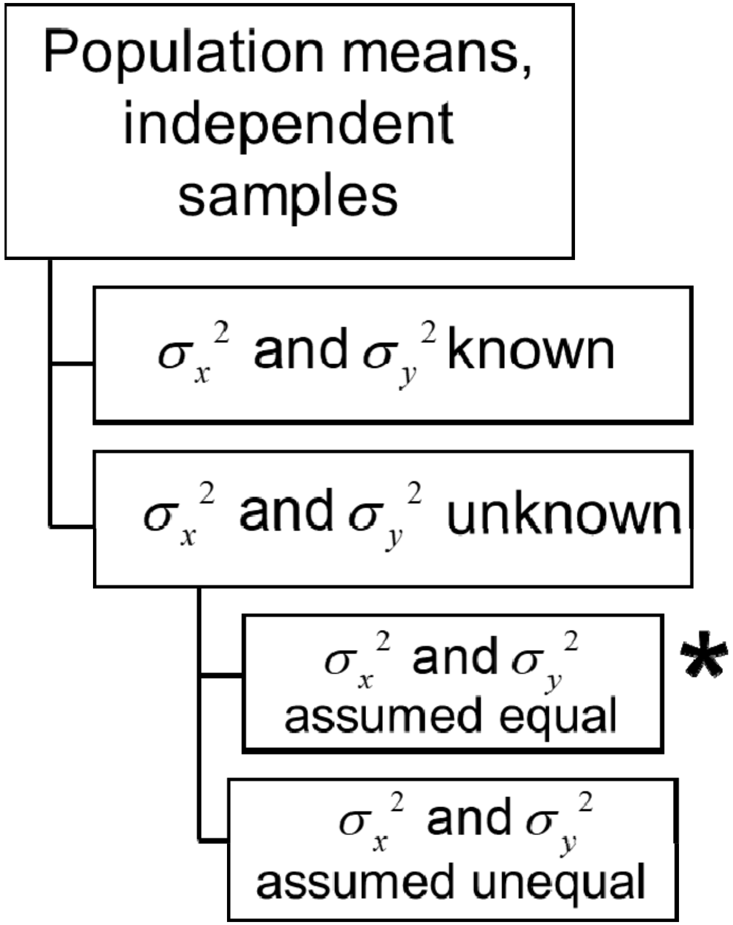
Population means, independent samples

- 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:

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

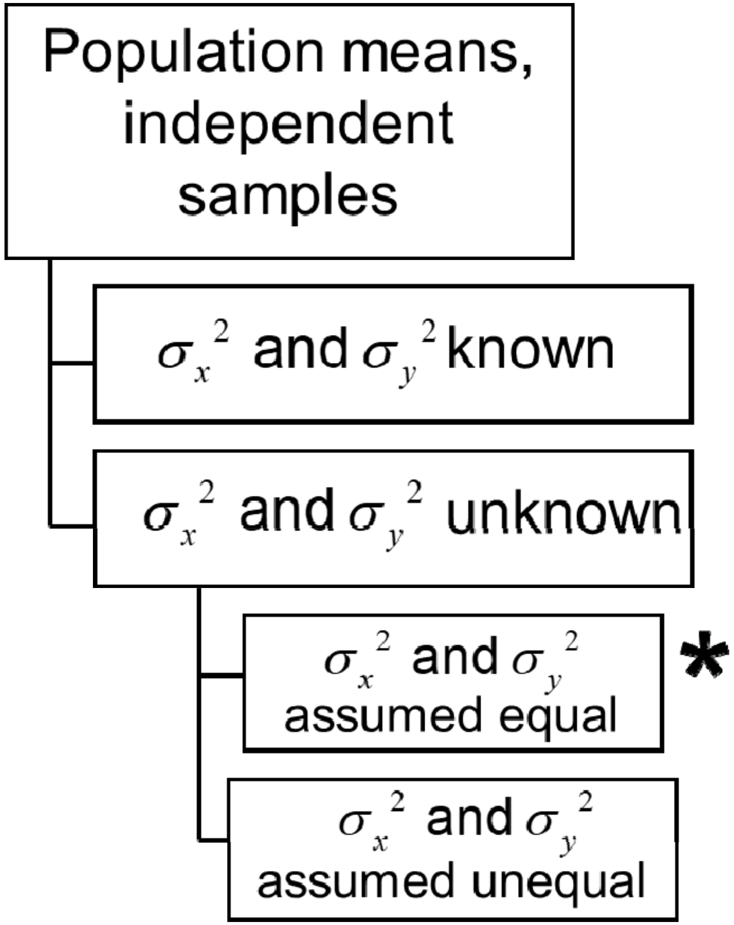
Sigma Sub *x* Squared and Sigma Sub *y* Squared Unknown, Assumed Equal (1 of 3)



Assumptions:

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

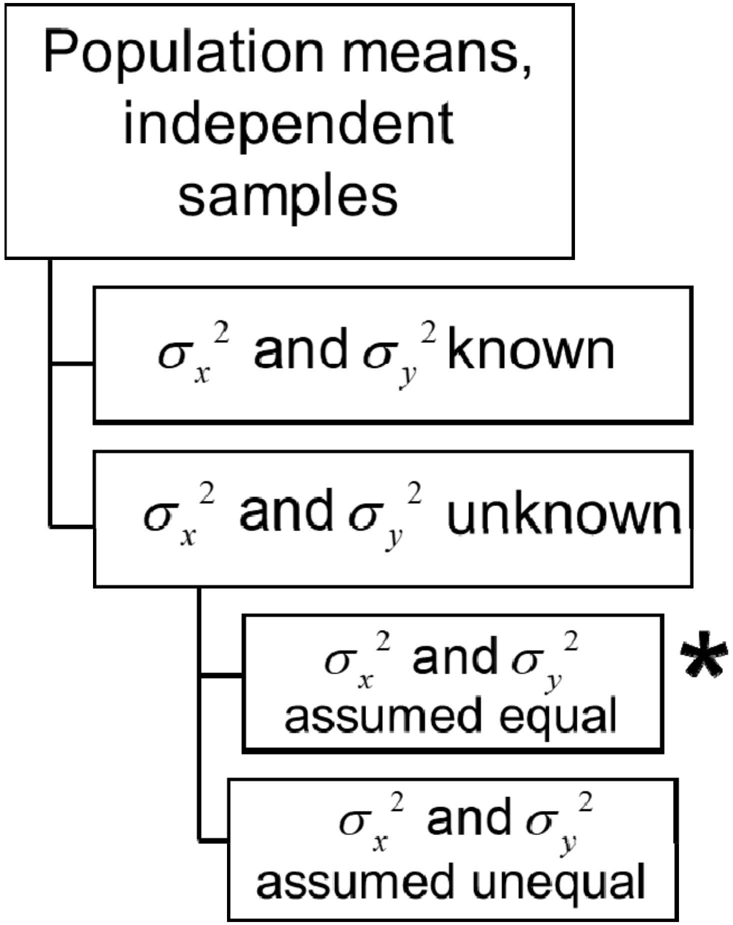
Sigma Sub *x* Squared and Sigma Sub *y* Squared Unknown, Assumed Equal (2 of 3)



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

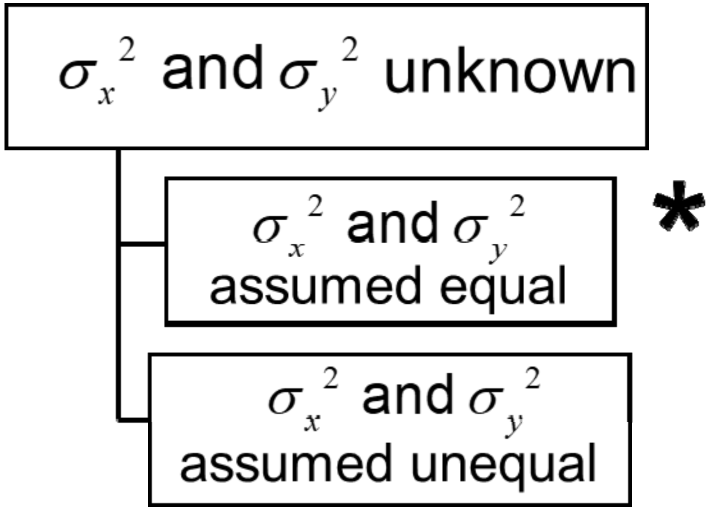
Sigma Sub *x* Squared and Sigma Sub *y* Squared Unknown, Assumed Equal (3 of 3)



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, Sigma Sub *x* Squared and Sigma Sub *y* Squared Unknown, Equal



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

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

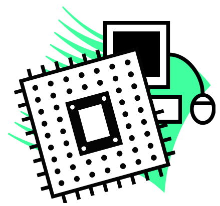
Where $$s_p^2 = \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2}$$

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



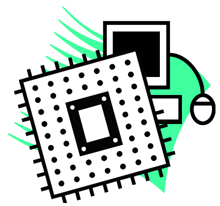
Calculating the Pooled Variance

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:

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



Calculating the Confidence Limits

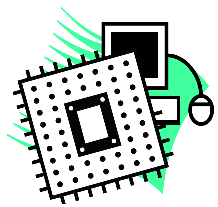
- The 95% confidence interval is

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

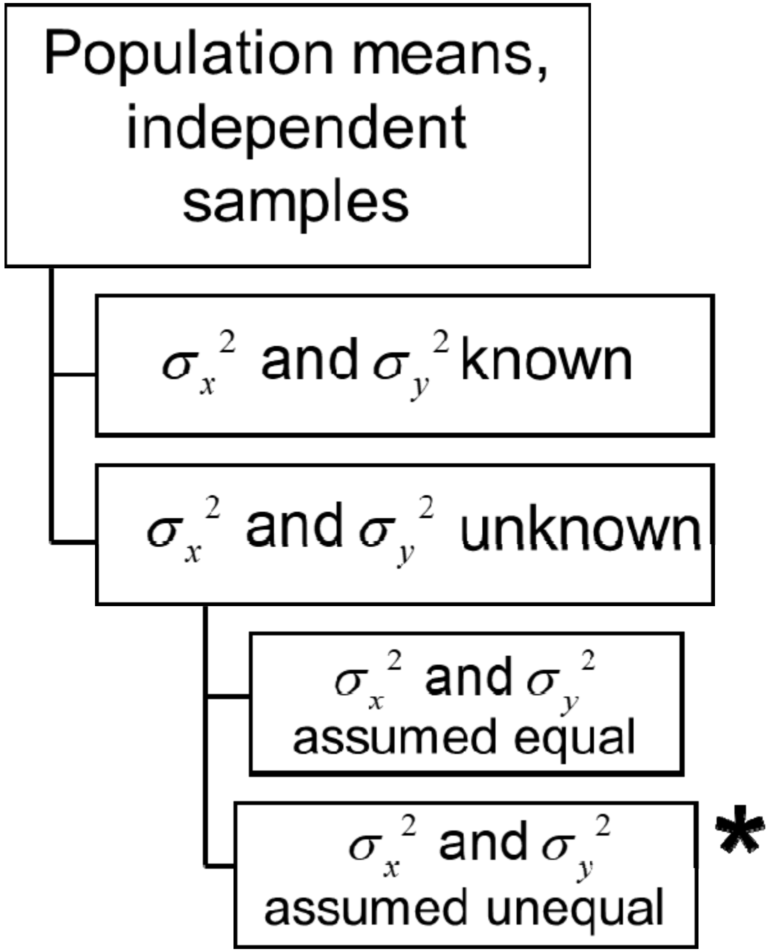
$$(3004-2538) \pm (2.054)\sqrt{\frac{4427.03}{17}+\frac{4427.03}{14}}$$

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

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



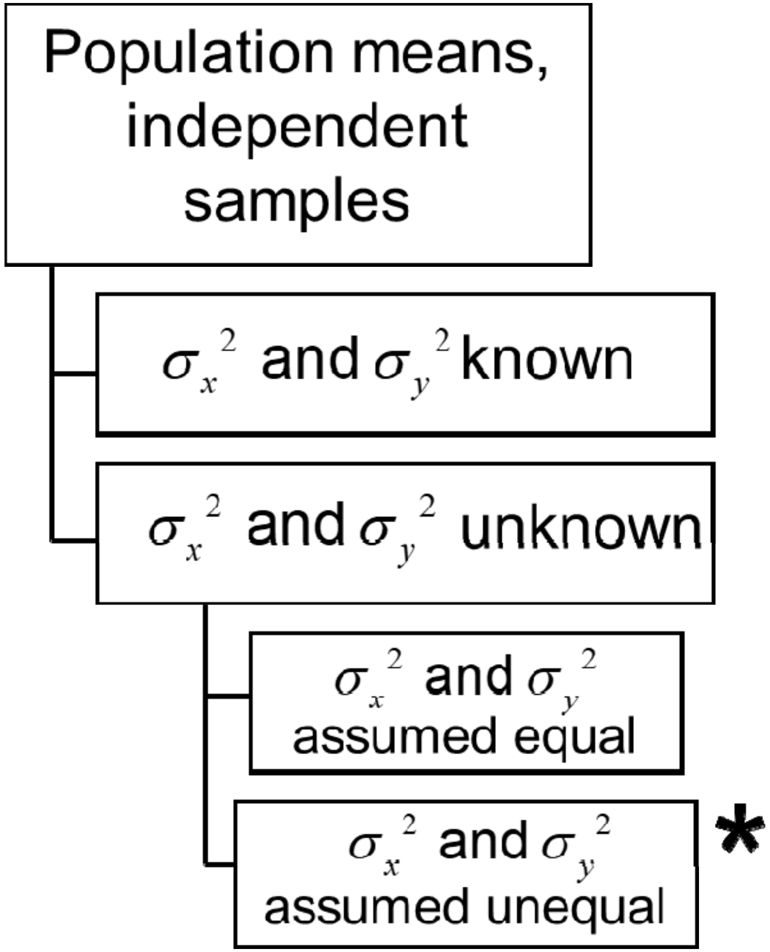
Sigma Sub *x* Squared and Sigma Sub *y* Squared Unknown, Assumed Unequal (1 of 2)



Assumptions:

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

Sigma Sub *x* Squared and Sigma Sub *y* Squared Unknown, Assumed Unequal (2 of 2)



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

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

Confidence Interval, Sigma Sub *x* Squared and Sigma Sub *y* Squared 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:

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

Where

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

Section 8.3 Two Population Proportions (1 of 2)

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, $P_x - P_y$