Confidence Intervals for The Difference of Means of Two Independent Populations

In all cases we assume:

1. The data are
$$Y_{1,1}, Y_{1,2}, \ldots, Y_{1,n_1}, \text{ where } Y_{1,j} = \mu_1 + \epsilon_{1,j}, \text{ (population 1)}$$
$$Y_{2,1}, Y_{2,2}, \ldots, Y_{2,n_2}, \text{ where } Y_{2,j} = \mu_2 + \epsilon_{2,j}, \text{ (population 2).}$$

2. The two populations are independent.

Case 1: Known Variances

Assumptions

1. Either
   (a) n_1 and n_2 are large, or
   (b) The $\epsilon_{1,j}$ are from a $N(0, \sigma_1^2)$ population, and the $\epsilon_{2,j}$ are from a $N(0, \sigma_2^2)$ population.
2. σ_1^2 and σ_2^2 are known.

Formulas

A level L confidence interval for $\mu_1 - \mu_2$ is $\left(\overline{Y}_1 - \overline{Y}_2 - \sigma(\overline{Y}_1 - \overline{Y}_2) z_{\frac{1+L}{2}},\ \overline{Y}_1 - \overline{Y}_2 + \sigma(\overline{Y}_1 - \overline{Y}_2) z_{\frac{1+L}{2}}\right)$, where $\sigma(\overline{Y}_1 - \overline{Y}_2) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$, and $z_{\frac{1+L}{2}}$ may be obtained from a table of normal quantiles (click here).

Case 2: Unknown Variances, n_1 and n_2 Large

This case is treated exactly as Case 1 except that $\sigma(\overline{Y}_1 - \overline{Y}_2)$ is replaced by $\hat{\sigma}(\overline{Y}_1 - \overline{Y}_2) = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$, where S_1 and S_2 are the sample standard deviations computed from the data from populations 1 and 2 respectively.

Case 3: Variances Unknown, but Assumed to Be Equal, n_1 and n_2 Not Both Large

Assumptions

1. The $\epsilon_{1,j}$ are from a $N(0, \sigma_1^2)$ population, and the $\epsilon_{2,j}$ are from a $N(0, \sigma_2^2)$ population.
2. σ_1^2 and σ_2^2 are unknown, but are assumed to be equal.

Formulas

A level L confidence interval for $\mu_1 - \mu_2$ is $\left(\overline{Y}_1 - \overline{Y}_2 - \hat{\sigma}_p(\overline{Y}_1 - \overline{Y}_2) t_{\nu,\frac{1+L}{2}},\ \overline{Y}_1 - \overline{Y}_2 + \hat{\sigma}_p(\overline{Y}_1 - \overline{Y}_2) t_{\nu,\frac{1+L}{2}}\right)$, where $\hat{\sigma}_p(\overline{Y}_1 - \overline{Y}_2) = \sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$, where $S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$, S_1 and S_2 are the sample standard deviations computed from the data from populations 1 and 2 respectively, $\nu = n_1 + n_2 - 2$, and $t_{\nu,\frac{1+L}{2}}$ may be obtained from a table of quantiles of the t distribution (click here).

Case 4: Variances Unknown, and Not Assumed to Be Equal, n_1 and n_2 Not Both Large

Assumptions

1. The $\epsilon_{1,j}$ are from a $N(0, \sigma_1^2)$ population, and the $\epsilon_{2,j}$ are from a $N(0, \sigma_2^2)$ population.
2. σ_1^2 and σ_2^2 are unknown, and are not assumed to be equal.

Formulas

A level L confidence interval for $\mu_1 - \mu_2$ is $\left(\overline{Y}_1 - \overline{Y}_2 - \hat{\sigma}(\overline{Y}_1 - \overline{Y}_2)t_{\nu,\frac{1+L}{2}},\ \overline{Y}_1 - \overline{Y}_2 + \hat{\sigma}(\overline{Y}_1 - \overline{Y}_2)t_{\nu,\frac{1+L}{2}}\right)$, where $\hat{\sigma}(\overline{Y}_1 - \overline{Y}_2) = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$, S_1 and S_2 are the sample standard deviations computed from the data from populations 1 and 2 respectively, the degrees of freedom ν is taken as the largest integer less than or equal to $\left[\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right]^2 \Big/ \left[\frac{(S_1^2/n_1)^2}{n_1-1} + \frac{(S_2^2/n_2)^2}{n_2-1}\right]$, and $t_{\nu,\frac{1+L}{2}}$ may be obtained from a table of quantiles of the t distribution (click here).

Click Here for An Example