Confidence Intervals for The Difference of Means of Two Independent Populations

Assumptions

1. The data are

$$Y_{1,1}, Y_{1,2}, \ldots, Y_{1,n}, \text{ where } Y_{1,j} = \mu_1 + \epsilon_{1,j}, \text{ (population 1)}$$
$$Y_{2,1}, Y_{2,2}, \ldots, Y_{2,n}, \text{ where } Y_{2,j} = \mu_2 + \epsilon_{2,j}, \text{ (population 2).}$$

2. The $\epsilon_{1,j}$ are independent with a $N(0, \sigma_1^2)$ distribution, and the $\epsilon_{2,j}$ are independent with a $N(0, \sigma_2^2)$ distribution.

3. The measurements are paired: $Y_{1,1}$ is paired with $Y_{2,1}$, $Y_{1,2}$ is paired with $Y_{2,2}$, and so on. This might happen if, for example, $Y_{1,1}, Y_{1,2}, \ldots, Y_{1,n}$ were blood pressure measurements taken on each of n people, and $Y_{2,1}, Y_{2,2}, \ldots, Y_{2,n}$ were blood pressure measurements taken on the same individuals after the administration of blood pressure medication.

Formulas

To estimate the difference in population means, $\mu_1 - \mu_2$, we take the differences $D_i = Y_{1,i} - Y_{2,i}$. Under the given assumptions, these differences have a $N(\mu_D, \sigma_D^2)$ distribution, where $\mu_D = \mu_1 - \mu_2$, and $\sigma_D^2 = \sigma_1^2/n_1 + \sigma_2^2/n_2$. We can therefore substitute D_i for Y_i in the formula for a confidence interval for the mean of a single population to obtain a confidence interval for $\mu_1 - \mu_2$.

Specifically, if $\overline{D}$ is the mean and S_D the standard deviation of the differences D_i, the formula for a level L confidence interval is $\left(\overline{D} - \hat{\sigma}(\overline{D}) t_{n-1,\frac{1+L}{2}},\ \overline{D} + \hat{\sigma}(\overline{D}) t_{n-1,\frac{1+L}{2}}\right)$, where $\hat{\sigma}(\overline{D}) = S_D/\sqrt{n}$, and $t_{n-1,\frac{1+L}{2}}$ may be obtained from a table of quantiles of the t distribution (click here).

Click Here for An Example