Inferences comparing two central values

TRUE FACT: if $X \sim$ normal$(\mu_1, {\sigma_1}^2)$ and $Y \sim$ normal$(\mu_2, {\sigma_2}^2)$, and X and Y are independent, then

$$X + Y \sim \text{normal}\left(\mu_1 + \mu_2, {\sigma_1}^2 + {\sigma_2}^2\right)$$

$$X - Y \sim \text{normal}\left(\mu_1 - \mu_2, {\sigma_1}^2 + {\sigma_2}^2\right)$$

Note: if X and Y have correlation ρ, then

$$X + Y \sim \text{normal}\left(\mu_1 + \mu_2, {\sigma_1}^2 + {\sigma_2}^2 + 2\rho\sigma_1\sigma_2\right)$$

$$X - Y \sim \text{normal}\left(\mu_1 - \mu_2, {\sigma_1}^2 + {\sigma_2}^2 - 2\rho\sigma_1\sigma_2\right)$$

Two populations:

distribution of $\overline{Y}_2$ distribution of $\overline{Y}_1$

Data: $\overline{y}_1, s_1^2, n_1;\ \overline{y}_2, s_2^2, n_2$

Use $\overline{y}_1 - \overline{y}_2$ as an estimate of $\mu_1 - \mu_2$

Distribution of $\overline{Y}_1 - \overline{Y}_2$ has a normal$\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$ distribution

Old fashioned assumption: $\sigma_1^2 = \sigma_2^2 = \sigma^2$

σ^2 estimated by "pooled" estimate:

$$s_p^2 = \frac{n_1 - 1}{n_1 + n_2 - 2} s_1^2 + \frac{n_2 - 1}{n_1 + n_2 - 2} s_2^2$$

Also $\overline{Y}_1 - \overline{Y}_2 \sim$ normal$\left(\mu_1 - \mu_2, \sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right)$, and

$$T = \frac{\overline{Y}_1 - \overline{Y}_2 - (\mu_1 - \mu_2)}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

has a Student's t distribution with $n_1 + n_2 - 2$ degrees of freedom

100(1 − α)% confidence interval for $\mu_1 - \mu_2$

$$\overline{y}_1 - \overline{y}_2 \pm t_{\alpha/2}\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

Hypothesis tests

H_0: $\mu_1 - \mu_2 = \delta_0$ (fixed constant, usually 0)

$$\mathrm{H_a}\colon\ \mu_1 - \mu_2 \left\{ \begin{array}{c} > \\ < \\ \neq \end{array} \right\} \delta_0$$

Test statistic:

$$t = \frac{\overline{y}_1 - \overline{y}_2 - (\delta_0)}{\sqrt{{s_{\mathrm{p}}}^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Rejection region:

$$\text{reject } \mathrm{H}_0 \text{ if} \left\{ \begin{array}{c} t \geq t_\alpha \\ t \leq -t_\alpha \\ |t| \geq t_{\alpha/2} \end{array} \right\}$$

Modern approach: $\sigma_1^{\,2} \neq \sigma_2^{\,2}$ (Behrens-Fisher problem)

One solution to constructing CI & t-test for $\mu_1 - \mu_2$ in the presence on unequal variances is based on Welsh's approximation:

$$T = \frac{\overline{Y}_1 - \overline{Y}_2 - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{S_1^{\,2}}{n_1} + \frac{S_2^{\,2}}{n_2}\right)}}$$

has an approximate Student's t distribution with

$$\text{df} = \left[\frac{(n_1 - 1)(n_2 - 1)}{(1 - c)^2(n_1 - 1) + c^2(n_2 - 1)}\right]$$

where

$$c = \frac{\frac{s_1^{\,2}}{n_1}}{\frac{s_1^{\,2}}{n_1} + \frac{s_2^{\,2}}{n_2}}$$

and $[a]$ denotes rounding a *down* to the nearest integer

$100(1 - \alpha)\%$ confidence interval for $\mu_1 - \mu_2$

$$\overline{y}_1 - \overline{y}_2 \pm t_{\alpha/2}\sqrt{\left(\frac{s_1^{\,2}}{n_1} + \frac{s_2^{\,2}}{n_2}\right)}$$

Hypothesis tests

H$_0$: $\mu_1 - \mu_2 = \delta_0$ (fixed constant, usually 0)

H$_a$: $\mu_1 - \mu_2 \left\{ \begin{array}{c} > \\ < \\ \neq \end{array} \right\} \delta_0$

Test statistic:

$$t = \frac{\overline{y}_1 - \overline{y}_2 - (\delta_0)}{\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}}$$

Rejection region:

$$\text{reject } H_0 \text{ if} \left\{ \begin{array}{c} t \geq t_\alpha \\ t \leq -t_\alpha \\ |t| \geq t_{\alpha/2} \end{array} \right\}$$

Paired observations

Measure X and Y from the same subject (before/after, right/left, etc.). We are interested if there is any change, that is, interested in estimating/testing the difference of means: $\mathrm{E}(X) - \mathrm{E}(Y) = \mu_1 - \mu_2$. The random variables X and Y are probably **dependent**. However, the **differences** $D = X - Y$ are independent from subject to subject if the subjects are picked at random.

Model: $D \sim \text{normal}(\mu_d, \sigma_d^2)$ where $\mu_d = \mu_1 - \mu_2$

Data:

case 1	x_1	y_1	$d_1 = x_1 - y_1$
case 2	x_2	y_2	$d_2 = x_2 - y_2$
	$\vdots$		
case n	x_n	y_n	$d_n = x_n - y_n$

Estimate $\mu_1 - \mu_2$ with $\overline{d}$; estimate σ_d^2 with s_d^2

Inferences reduce to ordinary one-sample CI & tests for μ_d using Student's t distribution (df $= n - 1$)

100$(1 - \alpha)$% confidence interval for $\mu_1 - \mu_2$

$$\overline{d} \pm t_{\alpha/2}\sqrt{\frac{s_d^2}{n}}$$

Hypothesis tests

$\mathrm{H_0}$: $\mu_d = \mu_1 - \mu_2 = \delta_0$ (usually 0)

$$\mathrm{H_a}\text{: } \mu_d = \mu_1 - \mu_2 \left\{ \begin{array}{c} > \\ < \\ \neq \end{array} \right\} \delta_0$$

Test statistic:

$$t = \frac{\overline{d} - (\delta_0)}{(s_d/\sqrt{n})}$$

Rejection region:

$$\text{reject } \mathrm{H_0} \text{ if} \left\{ \begin{array}{c} t \geq t_\alpha \\ t \leq -t_\alpha \\ |t| \geq t_{\alpha/2} \end{array} \right\}$$