Comparison of the two mean values

CI for the difference in means in the two-sample normal model with known variances

$X_1 \dots X_{n_1} \sim N(\mu_1, \sigma_1^2)$ ← independent random samples
$Y_1 \dots Y_{n_2} \sim N(\mu_2, \sigma_2^2)$ ←

σ_1^2, σ_2^2 – are known

$\bar{X} = \frac{1}{n_1}\sum_{i=1}^{n_1} X_i$; $\bar{Y} = \frac{1}{n_2}\sum_{i=1}^{n_2} Y_i$

$\bar{Z} = \bar{X} - \bar{Y}$

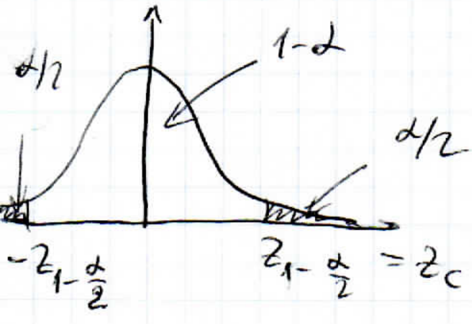
1) $\bar{Z}$ has normal distribution (as a sum of norm. dist. r.v.)

2) $\bar{X} \sim N(\mu_1, \frac{\sigma_1^2}{n_1})$; $\bar{Y} \sim N(\mu_2, \frac{\sigma_2^2}{n_2})$

$\bar{X}, \bar{Y}$ are independent

$\Rightarrow \bar{Z} \sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$

$\Rightarrow Z = \dfrac{\bar{Z} - (\mu_1 + \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$



$z_c = \text{qnorm}(1 - \frac{\alpha}{2})$

$P\{-z_c \le Z \le z_c\} = 1 - \alpha$

$P\left\{-z_c \le \dfrac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \le z_c\right\} = 1 - \alpha$

$$P\left\{(\bar{x}-\bar{y})-z_c\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\le \mu_1-\mu_2\le(\bar{x}-\bar{y})+z_c\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right\}=$$
$$=1-\alpha$$

$\Rightarrow$ CI for the difference in means in the two-sample normal model with known variances on confidence level $(1-\alpha)$:

$$\left[(\bar{x}-\bar{y})-z_c\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\,;\,(\bar{x}-\bar{y})+z_c\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right]$$

CI for the difference in means in the two-sample normal model with unknown equal variances

---

$X_1 \dots X_{n_1} \sim N(\mu_1,\sigma^2)$ ↖ independent random
$Y_1 \dots Y_{n_2} \sim N(\mu_2,\sigma^2)$ ↙ samples

σ^2 is unknown

$$\bar{x}=\frac{1}{n_1}\sum_{i=1}^{n_1}x_i\ ;\qquad \bar{y}=\frac{1}{n_2}\sum_{i=1}^{n_2}y_i$$

$$S_1^2=\frac{1}{n_1-1}\sum_{i=1}^{n_1}(x_i-\bar{x})^2\ ;\qquad S_2^2=\frac{1}{n_2-1}\sum_{i=1}^{n_2}(y_i-\bar{y})^2$$

$$\Rightarrow \underset{\substack{\uparrow\\ \text{pooled}}}{S_p^2}=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$$
– estimator of σ^2 based on two samples

↗ weighted mean

$$\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{(n_1-1)+(n_2-1)}$$

$$T=\frac{\bar{x}-\bar{y}-(\mu_1-\mu_2)}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t(n_1+n_2-2)$$

" " "

CI for the difference in means in the two-sample normal model with unknown equal variances on confidence level $(1-\alpha)$:

$$\left[(\bar{x}-\bar{y}) - t_{1-\frac{\alpha}{2}, n_1+n_2-2}\, S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \,;\, (\bar{x}-\bar{y}) + t_{1-\frac{\alpha}{2}, n_1+n_2-2}\, S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right]$$

CI for the difference in means in the two-sample normal model with unknown unequal variances

---

$$X_1 \ldots X_{n_1} \sim N(\mu_1, \sigma_1^2)$$
$$Y_1 \ldots Y_{n_2} \sim N(\mu_2, \sigma_2^2)$$

independent 2 samples

$$\bar{X} = \frac{1}{n_1}\sum_{i=1}^{n_1} X_i \,, \qquad \bar{Y} = \frac{1}{n_2}\sum_{i=1}^{n_2} Y_i$$

$$S_1^2 = \frac{1}{n_1-1}\sum_{i=1}^{n_1}(X_i-\bar{X})^2 ; \quad S_2^2 = \frac{1}{n_2-1}\sum_{i=1}^{n_2}(Y_i-\bar{Y})^2$$

Welch's t-test:

$$t = \frac{\bar{X}-\bar{Y}}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}} \sim t(\nu)$$

$$\nu \approx \frac{\left(\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}\right)^2}{\frac{S_1^4}{n_1^2(n_1-1)} + \frac{S_2^4}{n_2^2(n_2-1)}}$$

$$\left[\bar{x}-\bar{y} - t_{1-\frac{\alpha}{2},\nu}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}} \,;\, \bar{x}-\bar{y} + t_{1-\frac{\alpha}{2},\nu}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}\right]$$

CI for the difference in means in the two-sample normal model with unknown unequal variances on confidence level $(1-\alpha)$