L20: Two-sample estimation

1. Two-sample t-intervals

   (1) A variable with t-distribution

   $$\frac{l'(\overline{x}-\overline{y})-l'(\mu_x-\mu_y)}{s_{l'(\overline{x}-\overline{y})}} \sim t(n-2) \text{ where } s^2_{l'(\overline{x}-\overline{y})}=\left(\frac{1}{n_1}+\frac{1}{n_2}\right)l'S_ul.$$

   Proof: $\overline{x}-\overline{y} \sim N\left(\mu_x-\mu_y,\ \left(\frac{1}{n_1}+\frac{1}{n_2}\right)\Sigma\right)$ and CSSCP $\sim W_{p\times p}(\Sigma,\, n-2)$ are independent. So $l'(\overline{x}-\overline{y}) \sim N\left(l'(\mu_x-\mu_y),\ \sigma^2_{l'(\overline{x}-\overline{y})}\right)$ where $\sigma^2_{l'(\overline{x}-\overline{y})}=\left(\frac{1}{n_1}+\frac{1}{n_2}\right)l'\Sigma l$ is estimated by $s^2_{l'(\overline{x}-\overline{y})}=\left(\frac{1}{n_1}+\frac{1}{n_2}\right)l'S_ul$ where $S_u=\frac{CSSCP}{n-2}$ and $\frac{l'\text{CSSCP}l}{l'\Sigma l} \sim W_{1\times 1}(1,\, n-2)=\chi^2(n-2)$ are independent. So $\frac{l'(\overline{x}-\overline{y})-l'(\mu_x-\mu_y)}{\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)l'\Sigma l}}\cdot\frac{1}{\sqrt{\frac{l'\text{CSSCP}l}{l'\Sigma l(n-2)}}} \sim t(n-2)$, i.e.,

   $$\frac{l'(\overline{x}-\overline{y})-l'(\mu_x-\mu_y)}{s_{l'(\overline{x}-\overline{y})}} \sim t(n-2).$$

   (2) t-interval for $\theta=l'(\mu_x-\mu_y)$
   $l'(\overline{x}-\overline{y}) \pm t_{\alpha/2}(n-2)s_{l'(\overline{x}-\overline{y})}$ is a $1-\alpha$ CI for θ.
   $(-\infty,\ l'(\overline{x}-\overline{y})+t_{\alpha}(n-2)s_{l'(\overline{x}-\overline{y})})$ is a $1-\alpha$ lower-sided CI for θ.
   $(l'(\overline{x}-\overline{y})-t_{\alpha}(n-2)s_{l'(\overline{x}-\overline{y})},\ \infty)$ is a $1-\alpha$ upper-sided CI for θ.

   (3) t-tests
   With t in (1) one can also test on hypothesis about $l'(\mu_x-\mu_y)$. For example

   $H_0 : l'(\mu_x-\mu_y)=\delta_0$ versus $H_a : l'(\mu_x-\mu_y)\neq\delta_0$
   Test statistic: $t=\frac{l'(\overline{x}-\overline{y})-\delta_0}{s_{l'(\overline{x}-\overline{y})}}$
   Reject H_0 if $t<-t_{\alpha/2}(n-2)$ or $t>t_{\alpha/2}(n-2)$
   (p-value: $2\times P(t(n-2)>|t_{ob}|)$).

2. T^2-Confidence region for $\delta=\mu_x-\mu_y$

   (1) Confidence region for $\delta=\mu_x-\mu_y$
   Because $(\overline{X}-\overline{Y}-\delta)'\left(\left(\frac{1}{n_1}+\frac{1}{n_2}\right)S_u\right)^{-1}(\overline{X}-\overline{Y}-\delta) \sim T^2(p,\, n-2)$,
   $P\left((\overline{X}-\overline{Y}-\delta)'\left(\left(\frac{1}{n_1}+\frac{1}{n_2}\right)S_u\right)^{-1}(\overline{X}-\overline{Y}-\delta) \le T^2_\alpha(p,\, n-2)\right)=1-\alpha.$
   Thus $\{\delta :\ (\overline{X}-\overline{Y}-\delta)'\left(\left(\frac{1}{n_1}+\frac{1}{n_2}\right)S_u\right)^{-1}(\overline{X}-\overline{Y}-\delta) \le T^2_\alpha(p,\, n-2)\}$ is a confidence region for δ with confidence coefficient $1-\alpha$.
   This confidence region is a ellipsoid in R^p with center $\overline{X}-\overline{Y}$.

   (2) A relation
   $$\begin{aligned} & \delta_0 \text{ is in the } 1-\alpha T^2\text{-confidence region for } \delta=\mu_x-\mu_y \\ \iff\ & \alpha\text{-level } T^2\text{-test on } H_0 : \mu_x-\mu_y=\delta_0 \text{ fails to reject } H_0 \end{aligned}$$

   (3) Smallest confidence coefficient
   $$\begin{aligned} & \delta_0 \text{ is in the } 1-\alpha T^2\text{-confidence region for } \delta=\mu_x-\mu_y \\ \iff\ & \alpha\text{-level } T^2\text{-test on } H_0 : \mu_x-\mu_y=\delta_0 \text{ fails to reject } H_0 \\ \iff\ & (p\text{-value of the tet}) > \alpha \iff 1-(\text{p-value of the test}) < 1-\alpha \\ \Longleftarrow\ & (\text{Coffidence coefficient}) > 1-(\text{p-value}). \end{aligned}$$

3. Simultaneous confidence intervals

   (1) Bonferroni simultaneous CIs for $l_i'(\mu_x - \mu_y)$.
   By Bonferroni method,

$$l_i'(\overline{x} - \overline{y}) \pm t_{\alpha/(2k)}(n-2) s_{l_i'(\overline{x}-\overline{y})}, \ \ i = 1, ..., k,$$

   are k simultaneous CIs for $l_i'(\mu_x - \mu_y)$, $i = 1, .., k$, with overall confidence coefficient $1 - \alpha$.

   (2) Scheffe's confidence intervals

$$l_i'(\overline{x} - \overline{y}) \pm \sqrt{T_\alpha^2(p,\, n-2)}\, s_{l_i'(\overline{x}-\overline{y})}, \ \ i = 1, 2, ...,$$

   are simultaneous CIs for $l_i'(\mu_x - \mu_y)$, $i = 1, 2, ...$, with overall confidence coefficient $1 - \alpha$.

   Proof By the extended Cauchy-Schwartz inequality

$$\begin{aligned} &\{l_i'[(\overline{x} - \overline{y}) - (\mu_x - \mu_y)]\}^2 \\ \le\ & l_i' \left[\left(\tfrac{1}{n_1} + \tfrac{1}{n_2} \right) S_u \right] l_i \, [(\overline{x} - \overline{y}) - (\mu_x - \mu_y)]' \left[\left(\tfrac{1}{n_1} + \tfrac{1}{n_2} \right) S_u \right]^{-1} [(\overline{x} - \overline{y}) - (\mu_x - \mu_y)]. \end{aligned}$$

   So $\dfrac{\{l_i'[(\overline{x} - \overline{y}) - (\mu_x - \mu_y)]\}^2}{l_i' \left[\left(\frac{1}{n_1} + \frac{1}{n_2} \right) S_u \right] l_i} \le [(\overline{x} - \overline{y}) - (\mu_x - \mu_y)]' \left[\left(\dfrac{1}{n_1} + \dfrac{1}{n_2} \right) S_u \right]^{-1} [(\overline{x} - \overline{y}) - (\mu_x - \mu_y)].$

   Let $A = \left\{ [(\overline{x} - \overline{y}) - (\mu_x - \mu_y)]' \left[\left(\frac{1}{n_1} + \frac{1}{n_2} \right) S_u \right]^{-1} [(\overline{x} - \overline{y}) - (\mu_x - \mu_y)] \le T_\alpha^2(p,\, n-2) \right\}$ and
   $I_i = \left\{ \frac{\{l_i'[(\overline{x} - \overline{y}) - (\mu_x - \mu_y)]\}^2}{l_i' \left[\left(\frac{1}{n_1} + \frac{1}{n_2} \right) S_u \right] l_i} \le T_\alpha^2(p,\, n-2). \right\}$.
   Then $A \subset I_i$ for all $i \Longrightarrow A \subset \cap_i I_i \Longrightarrow P(A) \le P(\cap_i I_i) \Longrightarrow P(\cap_i I_i) \ge 1 - \alpha$.
   But $I_i = \left\{ l_i'(\mu_x - \mu_y) \in l_i'(\overline{x} - \overline{y}) \pm \sqrt{T_\alpha^2(p,\, n-2)}\, s_{l_i'(\overline{x}-\overline{y})} \right\}$.