L14 Relation of test and confidence region

1. Confidence region and tests
   Consider a population $N(\mu, \Sigma)$.

   (1) Confidence region
   $1-\alpha$ confidence region for $\mu \in R^p$ is the collection of all $\mu \in R^p$ satisfying

$$(\mu - \overline{X})' \left(\frac{S}{n}\right)^{-1} (\mu - \overline{X}) \le T^2_\alpha(p, n-1)$$

   (2) α-level Test

$H_0 : \mu = \mu_0$ vs $H_a : \mu \ne \mu_0$ Test Statistic: $T^2 = (\overline{X} - \mu_0)' \left(\frac{S}{n}\right)^{-1} (\overline{X} - \mu_0)$ Reject H_0 if $T^2 > T^2_\alpha(p, n-1)$

   (3) Test by p-value

$H_0 : \mu = \mu_0$ versus $H_a : \mu \ne \mu_0$ Test Statistic $T^2 = (\overline{X} - \mu_0)' \left(\frac{S}{n}\right)^{-1} (\overline{X} - \mu_0)$ p-value: $P(T^2(p, n-1) > T^2_{ob})$

2. Relations

   (1) Determing if μ_0 in the $1-\alpha$ confidence region for μ

$$\begin{aligned} & \mu_0 \text{ is in a } 1-\alpha \text{ confidence region for } \mu \\ \iff & \mu = \mu_0 \text{ satisfies } (\mu - \overline{X})' \left(\tfrac{S}{n}\right)^{-1} (\mu - \overline{X}) \le T^2_\alpha(p, n-1) \\ \iff & T^2 = (\overline{X} - \mu_0)' \left(\tfrac{S}{n}\right)^{-1} (\overline{X} - \mu_0) \le T_\alpha(p, n-1) \\ \iff & \alpha\text{-level test fails to reject } H_0 : \mu = \mu_0 \\ \iff & p\text{-value} > \alpha \end{aligned}$$

   Ex1: To test $H_0 : \mu = \begin{pmatrix} 6 \\ 9 \end{pmatrix}$ one ran SAS and obtained

Statistics	Value	F-value	Num DF	Den DF	Pr>F
Wilks' Lambda	0.7200	0.19	2	1	0.8485

   Is $\begin{pmatrix} 9 \\ 6 \end{pmatrix}$ in an 80% confidence region for μ?

   Note that $\begin{pmatrix} 6 \\ 9 \end{pmatrix}$ is in an 80% confidence region for $\mu \iff p$-value > 0.20

   But p-value$= 0.8485 > 0.20$. So yes. $\begin{pmatrix} 6 \\ 9 \end{pmatrix}$ is in the 80% confidence region for μ.

   (2) Find smallest confidence coefficient such that μ_0 is in the confidence region

$$\begin{aligned} \mu_0 \text{ is in a } 1-\alpha \text{ confidence region for } \mu & \iff p\text{-value} > \alpha \\ & \iff 1-\alpha > 1-(p\text{-value}). \end{aligned}$$

   Ex2: In Ex1 find the smallest confidence coefficient such that $\begin{pmatrix} 6 \\ 9 \end{pmatrix}$ is in the confidence region.

   With p-value$= 0.8485$,

$$\begin{aligned} & \mu_0 \text{ is in a } 1-\alpha \text{ confidence region for } \mu \iff p\text{-value} > \alpha \\ \iff & 1-\alpha > 1-(p\text{-value}) = 0.1515 = 15.15\% \end{aligned}$$

The smallest confidence coefficient such that $\begin{pmatrix} 6 \\ 9 \end{pmatrix}$ is in the confidence region is 15.15%.

3. Univariate two-sample problems

   (1) Populations, samples and basic statistics
   Populations: $N(\mu_1, \sigma^2)$ and $N(\mu_2, \sigma^2)$.
   Samples: $X_1, ..., X_{n_1}$ from Populations and $Y_1, .., Y_{n_2}$ from population 2
   Basic statistics: n_1, $\overline{X}$, $\text{CSS}_1 = \sum_i (X_i - \overline{X})^2$ and $s_1^2 = \frac{\text{CSS}_1}{n_1-1}$
   n_2, $\overline{Y}$, $\text{CSS}_2 = \sum_i (Y_i - \overline{Y})^2$ and $s_2^2 = \frac{\text{CSS}_2}{n_2-1}$
   $n = n_1 + n_2$, $\text{CSS} = \text{CSS}_1 + \text{CSS}_2$, $s_p^2 = \frac{\text{CSS}}{n-2} = \frac{n_1-1}{n-2} s_1^2 + \frac{n_2-1}{n-2} s_2^2$.

   (2) Point estimators and confidence intervals
   μ_1 has MLE $\overline{X}$ that is an UE;
   μ_2 has MLE $\overline{Y}$ that is an UE;
   σ^2 has MLE $\frac{\text{CSS}}{n}$ and UE s_p^2.
   $\left(\frac{\text{CSS}}{\chi^2_{\alpha/2}(n-2)}, \frac{\text{CSS}}{\chi^2_{1-\alpha/2}(n-2)} \right)$ is a $1 - \alpha$ CI for σ^2
   $\overline{X} - \overline{Y} \pm t_{\alpha/2}(n-2) \sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$ is a $1 - \alpha$ CI for $\mu_1 - \mu_2$.

   (3) Tests

$H_0 : \mu_1 - \mu_2 = \delta_0$ vs $H_a : \mu_1 - \mu_2 \neq \delta_0$
Test statistic: $f = \frac{\overline{X} - \overline{Y} - \delta_0}{\sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$
Reject H_0 if $t < -t_{\alpha/2}(n-2)$ or $t > t_{\alpha/2}(n-2)$

$H_0 : \mu_1 - \mu_2 = \delta_0$ vs $H_a : \mu_1 - \mu_2 \neq \delta_0$
Test statistic: $f = \frac{\overline{X} - \overline{Y} - \delta_0}{\sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$
p-value: $2P(t(n-2) > \lvert t_{ob} \rvert)$

   (4) Two multivariate populations

   Two populations $N_p(\mu_x, \Sigma)$ and $N_p(\mu_y, \Sigma)$ can be expressed as

$$N_{p \times 2}(\mu, \Sigma, I_2) \text{ where } \mu = (\mu_x, \mu_y) \in R^{p \times 2}$$

   which is equivalent to $N \left(\begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \begin{pmatrix} \Sigma & 0 \\ 0 & \Sigma \end{pmatrix} \right)$.

   Two samples $X \in R^{p \times n_1}$ from $N_p(\mu_x, \Sigma)$ and $Y \sim R^{p \times n_2}$ from $N_p(\mu_y, \Sigma)$ can be individually expressed as

$$X \sim N_{p \times n_1}(\mu_x 1'_{n_1}, \Sigma, I_{n_1}) \text{ and } Y \sim N_{p \times n_2}(\mu_y 1'_{n_1}, \Sigma, I_{n_2}),$$

   or expressed together by

$$(X, Y) \sim N_{p \times n}(\mu J', \Sigma, I_n)$$

   where $n = n_1 + n_2$, $\mu = (\mu_x, \mu_y) \in R^{p \times 2}$ and $J = \begin{pmatrix} 1_{n_1} & 0 \\ 0 & 1_{n_2} \end{pmatrix}$. The expression is equivalent to

$$\begin{pmatrix} X_1 \\ \vdots \\ X_{n_1} \\ Y_1 \\ \vdots \\ Y_{n_2} \end{pmatrix} \sim N \left(\begin{pmatrix} \mu_x \\ \vdots \\ \mu_x \\ \mu_y \\ \vdots \\ \mu_y \end{pmatrix}, \begin{pmatrix} \Sigma & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \Sigma \end{pmatrix} \right).$$