Unit II (Multivariate Analysis)

The generalised T^2 statistic

In the univariate case, a random sample of size n observations $x_1, x_2 \cdots x_n$ are drawn from a normal population with mean μ and unknown variance σ^2, the test statistic under the hypothesis

$H_0 : \mu = \mu_0$ is $t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \sim t(\alpha)$ with $(n-1)$ d.f. $\hookrightarrow$ (1)

where $\bar{X} = \frac{1}{n} \sum x_i$ and $S^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2$

The decision rule is

accept H_0 if $|t| \leq t(\alpha)$ with $(n-1)$ d.f.

reject H_0 if $|t| > t(\alpha)$ with $(n-1)$ d.f.

The multivariate analog of the square of 't' given (1) is

$$T^2 = N (\bar{x} - \mu_0)' S^{-1} (\bar{x} - \mu_0)$$

where, $\bar{x}$ is a mean vector of the sample of size N and S is the sample co-variance matrix.

Hotelling proposed the T^2 statistic for two samples and derived the distribution when μ is the population mean.

Derivation of the generalised Hotelling T^2 Statistic

The likelihood ratio test of the hypothesis $H_0 : \mu = \mu_0$ on the basis of the sample from $N(\mu, \Sigma)$ is based on the T^2 – Statistic,

$$T^2 = N(\bar{x} - \mu)' S^{-1} (\bar{x} - \mu)$$

Suppose we have N observations $x_1, x_2 \cdots x_N$ $(N > p)$, the likelihood function is

$$L(\mu, \Sigma) = (2\pi)^{\frac{-Np}{2}} |\Sigma|^{-N/2} \exp\left\{-\frac{1}{2} \sum_{\alpha=1}^{N} (x_\alpha - \mu)' \Sigma^{-1} (x_\alpha - \mu)\right\}$$

The likelihood ratio criterion for testing the hypothesis $H_0 : \mu = \mu_0$ is

$$\lambda = \frac{\max_{\Sigma} L(\mu_0, \Sigma)}{\max_{\mu, \Sigma} L(\mu, \Sigma)} \longrightarrow (2)$$

when the parameters are unrestricted, the maximum occurs when μ, Σ are defined by the maximum likelihood estimators of μ and Σ.

ie $\hat{\mu}_\Omega = \bar{x}$ and $\hat{\Sigma}_\Omega = \frac{1}{N} \sum_{\alpha=1}^{N} (x_\alpha - \bar{x})^{T} (x_\alpha - \bar{x})'$

when $\mu = \mu_0$ the likelihood function is maximized at

$$\hat{\Sigma}_\omega = \frac{1}{N} \sum_{\alpha=1}^{N} (x_\alpha - \mu_0)(x_\alpha - \mu_0)' \longrightarrow (3)$$

under ω, the likelihood function reduced to

$$\max_{\Sigma} L(\mu_0, \Sigma) = (2\pi)^{-Np/2} |\Sigma_\omega|^{-N/2} \exp\left\{-\frac{1}{2} \sum_{\alpha=1}^{N} (x_\alpha - \mu_0) \hat{\Sigma}_\omega^{-1} (x_\alpha - \mu_0)'\right\}$$

$$= (2\pi)^{-Np/2} |\Sigma_\omega|^{-N/2} \exp\left\{-\frac{1}{2} \operatorname{tr}\left[\sum_{\alpha=1}^{N} (x_\alpha - \mu_0)' \hat{\Sigma}_\omega^{-1} (x_0 - \mu_0)\right]\right\}$$

$$= \text{"} \quad \text{"} \quad \exp\left\{-\frac{1}{2} \operatorname{tr} \sum_{\alpha=1}^{N} \left[(x_\alpha - \mu_0)' \hat{\Sigma}_\omega^{-1} (x_0 - \mu_0)\right]\right\}$$

$$= \text{"} \quad \text{"} \quad \exp\left\{-\frac{1}{2} \sum_{\alpha=1}^{N} \operatorname{tr}\left[\hat{\Sigma}_\omega^{-1} (x_\alpha - \mu_0)' (x_0 - \mu_0)\right]\right\}$$

$$= \;,, \exp\left\{-\frac{1}{2}\operatorname{tr}\hat{\Sigma}_{\omega}^{-1}\left[(x_\alpha-\mu_0)'(x_0-\mu_0)\right]\right\},$$

$$= \;,, \exp\left\{-\frac{1}{2}\operatorname{tr}\hat{\Sigma}_{\omega}^{-1}\left[N\,\hat{\Sigma}_{\omega}\right]\right\}$$

$$= \;,, \exp\left\{-\frac{1}{2}\operatorname{tr}\, NI\right\}$$

$$= \;,, \exp\left\{-\frac{1}{2}NP\right\}$$

$$\therefore \max_{\Sigma} L(\mu_0,\Sigma) = (2\pi)^{-NP/2}\left|\hat{\Sigma}_{\omega}\right|^{-N/2}\exp\left(-\frac{NP}{2}\right) \longrightarrow (4)$$

Similarly,

$$\max_{\mu,\Sigma} L(\mu,\Sigma) = (2\pi)^{-\frac{NP}{2}}\left|\hat{\Sigma}_{\Omega}\right|^{-N/2}\exp\left(-\frac{NP}{2}\right) \longrightarrow (5)$$

Thus the likelihood ratio criterion is

$$\lambda = \frac{\max_{\Sigma} L(\mu_0,\Sigma)}{\max_{\mu,\Sigma} L(\mu,\Sigma)}$$

$$= \frac{\left|\hat{\Sigma}_{\Omega}\right|^{N/2}}{\left|\hat{\Sigma}_{\omega}\right|^{N/2}}$$

$$= \frac{\left|\frac{1}{N}\sum_{\alpha=1}^{N}(x_\alpha-\bar{x})(x_\alpha-\bar{x})'\right|^{N/2}}{\left|\frac{1}{N}\sum_{\alpha=1}^{N}(x_\alpha-\mu_0)(x_\alpha-\mu_0)'\right|^{N/2}}$$

$$\therefore \lambda = \frac{|A|^{N/2}}{\left|A+N(\bar{x}-\mu_0)(x_0-\mu_0)'\right|^{N/2}}$$

$$\lambda^{2/N} = \frac{|A|}{\left|A+N(\bar{x}-\mu_0)(x_0-\mu_0)'\right|}$$

$$= \frac{|(N-1)S|}{|(N-1)S + N(\bar{x}-\mu_0)(\bar{x}-\mu_0)'|} \quad \text{where } S = \frac{1}{N-1}(A)$$

$$= \frac{|(N-1)S|}{|(N-1)S|\left|1 + \frac{N(\bar{x}-\mu_0)(\bar{x}-\mu_0)'}{(N-1)S}\right|}$$

$$= \frac{|(N-1)S|}{|(N-1)S|\left|1 + \frac{N(\bar{x}-\mu_0)S^{-1}(\bar{x}-\mu_0)'}{N-1}\right|}$$

$$\lambda^{2/N} = \frac{1}{1 + \frac{T^2}{N-1}} \longrightarrow (6)$$

where $T^2 = N(\bar{x}-\mu_0)S^{-1}(\bar{x}-\mu_0)'$

The likelihood ratio test is defined by the critical region $\lambda \le \lambda_0 \longrightarrow (7)$

where λ_0 is choosen so that the Prob. of (7) when the null hypothesis H_0 is true.

$\therefore$ (5) $\Rightarrow \lambda^{2/N} \le \lambda_0^{2/N}$

$\Rightarrow \lambda^{-2/N} \ge \lambda_0^{-2/N}$

$\therefore \lambda^{-2/N} - 1 \ge \lambda_0^{-2/N} - 1$

$(N-1)(\lambda^{-2/N} - 1) \ge (N-1)(\lambda_0^{-2/N} - 1)$

$T^2 \ge T_0^2$

$\therefore T_0^2 = (N-1)(\lambda_0^{-2/N} - 1)$

The distribution of Hotelling's T^2

W.K.T $T^2 = N(\bar{x}-\mu_0)' S^{-1}(\bar{x}-\mu_0) \longrightarrow (1)$

let $T^2 = Y' S^{-1} Y \longrightarrow (2)$

where $Y \sim N(\nu, \Sigma)$ and nS is distributed independently as $\sum_{\alpha=1}^{n} Z_\alpha Z_\alpha'$ with $z_1, z_2, z_3 \ldots z_n$ are independent, each with distribution $N(0, \Sigma)$. The T^2 defined with $Y = \sqrt{N}(\bar{x}-\mu_0)$ and $\nu = \sqrt{N}(\mu-\mu_0)$ and $n = N-1$.

$$T^2 = N(\bar{x}-\mu_0)' S^{-1}(\bar{x}-\mu_0)$$

$$T^2 = \sqrt{N}(\bar{x}-\mu_0)' S^{-1} \sqrt{N}(\bar{x}-\mu_0) \longrightarrow (3)$$

From (2) and (3)

$$Y = \sqrt{N}(\bar{x}-\mu_0)$$

$$E(Y) = E\{\sqrt{N}(\bar{x}-\mu_0)\}$$

$$E(Y) = \sqrt{N}\{E(\bar{x}-\mu_0)\}$$

$$\nu = \sqrt{N}(\mu-\mu_0) \quad \text{and} \quad n = N-1$$

Let D be a non singular matrix such that $D'\Sigma D = I$ and define

$$\left.\begin{aligned} Y^* &= DY \\ S^* &= DSD' \\ \nu^* &= D\nu \end{aligned}\right\} \longrightarrow (4)$$

Now $Y^* = DY$

$\Rightarrow DY = Y^*$

$Y = D^{-1} Y^*$

$Y' = (D^{-1} Y^*)'$

$$\boxed{Y' = Y^{*'} (D^{-1})'}$$

$S^* = DSD'$

$DSD' = S^*$

$S = D^{-1} S^* (D')^{-1}$

$= (D^{-1})[(S^*)^{-1}]^{-1} (D')^{-1}$

$$\boxed{S = (D'(S^*)^{-1} D)^{-1}}$$

$$\boxed{S^{-1} = D'(S^*)^{-1} D}$$

Then from (2)

$$T^2 = Y' S^{-1} Y$$

$$= \left((Y^*)' (D^{-1})'\right)\left(D' (S^*)^{-1} D\right)(D^{-1} Y^*)$$

$$= (Y^*)' (D^{-1})' (D')^{-1} (S^*)^{-1} Y^*$$

$$T^2 = (Y^*)' (S^*)^{-1} Y^*$$

where $Y^* \sim N(\nu^*, I)$. Since $Y \sim N(\nu, \Sigma)$

$Y^* = DY \sim N(D\nu, D\Sigma D')$

$\therefore Y^* \sim N(\nu^*, I)$

$$nS = \sum_{\alpha=1}^{n} Z_\alpha Z_\alpha' \quad \text{where } Z_\alpha \sim N(0, \Sigma)$$

$$nS^* = n(DSD') \quad \{\because \text{from (4)}$$

$$= D\, nS\, D'$$

$$= D\left(\sum_{\alpha=1}^{n} Z_\alpha Z_\alpha'\right) D'$$

$$= \sum_{\alpha=1}^{n} (DZ_\alpha)(Z_\alpha' D')$$

$$= \sum_{\alpha=1}^{n} (DZ_\alpha)(DZ_\alpha)'$$

$$= \sum_{\alpha=1}^{n} Z_\alpha^* (Z_\alpha^*)' \quad \text{where } DZ_\alpha = Z_\alpha^*$$

where Z_α^* are independent each with distribution $N(0, D\Sigma D')$

$\because Z_\alpha \sim N(0, \Sigma)$

$Z_\alpha^* = DZ_\alpha \sim N(0, D\Sigma D')$

$\sim N(0, I)$

Let the first row of a $p \times p$ orthogonal matrix Q be defined by

$$q_{1i} = \frac{\gamma_i^*}{\sqrt{\gamma^{*\prime} \gamma^*}} \quad ; \quad i = 1, 2, \ldots p \longrightarrow (5)$$

This is permissible since $\sum_i q_{1i}^2 = 1$

now let $U = Q \gamma^*$ and $B = Q n S^* Q^{-1}$

Q was defined by $U_1 = \sum q_{1i} \gamma_i^* = \frac{\gamma_i^{*\prime} \gamma_i^*}{\sqrt{\gamma^{*\prime} \gamma^*}}$

$$= \sqrt{\gamma^{*\prime} \gamma^*}$$

and $U_j = \sum q_{ji} \gamma_i^*$

$$= \sum q_{ji} q_{1i} \sqrt{\gamma^{*\prime} \gamma^*} \quad \text{using (5)}$$

$$= \sqrt{\gamma^{*\prime} \gamma^*} \sum q_{ji} q_{1i} \quad (6)$$

$U_j = 0$ for $j \neq 1$

since $q_{ji} q_{1i} = 0$

Then $\frac{T^2}{n} = \frac{\gamma^{*\prime} S^{*-1} \gamma^*}{n}$

Substitute $U = Q \gamma^*$

$\Rightarrow \gamma^* = Q^{-1} U$

$\therefore \frac{T^2}{n} = \frac{(Q^{-1}U)' (S^*)^{-1} (Q^{-1}U)}{n}$

$$= \frac{U' (Q^{-1})' (S^*)^{-1} (Q^{-1} U)}{n}$$

$$= \frac{U' (Q')' (S^*)^{-1} Q' U)}{n} \qquad (\because Q^{-1} = Q' \text{ if } Q \text{ is orthogonal})$$

$$\frac{T^2}{n} = \frac{U' Q (S^*)^{-1} Q' U}{n} \longrightarrow (b)$$

and $B = Q n S^* Q'$

$$B^{-1} = (Q n S^* Q')^{-1}$$

$$= (Q')^{-1} (S^*)^{-1} n^{-1} Q^{-1}$$

$$B^{-1} = \frac{(Q')^{-1} (S^*)^{-1} Q^{-1}}{n} \qquad \{\because Q' = Q^{-1}$$

$$\boxed{B^{-1} = \frac{Q (S^*)^{-1} Q'}{n}}$$

$\therefore$ (b) $\Rightarrow \frac{T^2}{n} = U' B^{-1} U$

Since $U_1 = \sqrt{Y^{*'} Y^*}$ and $U_j = 0$ for $j \neq i$

The vector $U = \begin{bmatrix} U_1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$

$$\therefore \frac{T^2}{n} = U' B^{-1} U$$

$$= (U_1 \; 0 \; 0 \ldots 0) \begin{bmatrix} b^{11} & b^{12} & \ldots & b^{1p} \\ b^{21} & b^{22} & \ldots & b^{2p} \\ \vdots & \vdots & & \vdots \\ b^{p1} & b^{p2} & \ldots & b^{pp} \end{bmatrix} \begin{bmatrix} U_1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$\boxed{\frac{T^2}{n} = U_1^2 \, b^{11}}$$

where $b^{ij} = B^{-1}$

Applications of T^2 in tests on mean vector

(i) Tests the hypothesis that the mean vector is a given vector

T^2 test statistic is used to testing the hypothesis that the mean vector $\mu = \mu_0$ when the dispersion matrix Σ is known.

The likelihood ratio test of the hypothesis $H_0 : \mu = \mu_0$ on the basis of the sample N from $N(\mu, \Sigma)$ is equivalent to $T^2 \geq T_0^2$ with the level of significance α. The hypothesis is rejected if

$$T^2 \geq \frac{(N-1)p}{N-p} F_{(p, N-p) \text{ d.f}} (\alpha)$$

(ii) Finding the confidence region for the mean vector

Let $X_\alpha = (X_{\alpha 1}, X_{\alpha 2} \cdots X_{\alpha N})'$ $\alpha = 1, 2 \cdots N$ be a sample of size N from p-variate normal distn. with unknown mean μ and unknown positive definite Covariance matrix Σ.

WKT $\bar{x} = \frac{1}{N} \sum_{\alpha=1}^{N} X_\alpha$

and $S = \frac{1}{N-1} \sum_{\alpha=1}^{N} (X_\alpha - \bar{x})(X_\alpha - \bar{x})'$

$N(\bar{x} - \mu)' S^{-1} (\bar{x} - \mu)$ is distributed as Hotelling T^2 distribution with $(N-1)$ d.f.

Let $T_0^2(\alpha)$; $0 < \alpha < 1$ Such that $P_r(T^2 \geq T_0^2) = \alpha$, then the probabilities of drawing a sample of X_α with mean vector $\bar{x}$ and sample covariance matrix S Such that

$$N(\bar{x} - \mu_0)' S^{-1} (\bar{x} - \mu_0) \leq T_0^2(\alpha) \longrightarrow (1)$$

Thus if we compute ① for a particular sample, we have confidence $(1-\alpha)$ the equation ① is true statement concerning μ.

(iii) Test for the equality of two mean vectors

The T^2 statistic is used to test the null hypothesis that the mean vector of one normal population is equal to the mean vector of the other normal population, where the covariance matrices are assumed to be equal but unknown.

$H_0 : \mu^{(1)} = \mu^{(2)}$

Suppose $Y_1^{(1)}, Y_2^{(1)}, \ldots Y_N^{(1)}$ be a sample from $N_p(\mu^{(1)}, \Sigma)$

The sample mean $\bar{Y}^{(1)} = \frac{1}{N_1} \sum_{\alpha=1}^{N_1} Y_\alpha^{(1)}$

Then $\bar{Y}^{(1)} \sim N\left(\mu^{(1)}, \frac{\Sigma}{N_1}\right)$

Similarly $\bar{Y}^{(2)} \sim N\left(\mu^{(2)}, \frac{\Sigma}{N_2}\right)$

$\therefore (\bar{Y}^{(1)} - \bar{Y}^{(2)}) \sim N\left(\mu^{(1)} - \mu^{(2)}, \frac{\Sigma}{N_1} + \frac{\Sigma}{N_2}\right)$

$\sim N\left(\underset{\sim}{0}, \left(\frac{N_1+N_2}{N_1 N_2}\right)\Sigma\right)$ {$\because$ by H_0

$$\sqrt{\frac{N_1 N_2}{N_1+N_2}} (\bar{Y}^{(1)} - \bar{Y}^{(2)}) \sim N(\underset{\sim}{0}, \Sigma)$$

If we let

$$S = \frac{1}{N_1+N_2-2} \left\{ \sum_{\alpha=1}^{N_1} (Y_\alpha^{(1)} - \bar{Y}^{(1)})(Y_\alpha^{(1)} - \bar{Y}^{(1)})' + \sum_{\alpha=1}^{N_2} (Y_\alpha^{(2)} - \bar{Y}^{(2)})(Y_\alpha^{(2)} - \bar{Y}^{(2)})' \right\}$$

Then $(N_1+N_2)S$ is distributed as $\sum_{\alpha=1}^{N_1+N_2-2} Z_\alpha Z_\alpha'$ where $Z_\alpha \sim N(\underset{\sim}{0}, \Sigma)$

Thus $T^2 = \frac{N_1 N_2}{N_1+N_2} (\bar{y}^{(1)} - \bar{y}^{(2)})' S^{-1} (\bar{y}^{(1)} - \bar{y}^{(2)})$ is distributed as T^2 with (N_1+N_2-2) d.f.

The critical region is

$$T^2 > \frac{(N_1+N_2-2)\,p}{N_1+N_2-p-1} F^{(\alpha)}_{(p,\ N_1+N_2-p-1)} \text{ d.f.}$$

(iv) Two sample problem with unequal covariance matrices

Let $\{x_\alpha^{(i)}\}$ $\alpha = 1, 2 \ldots N_i$ be samples from $N(\mu^{(i)}, \Sigma_i)$ $i = 1, 2$

we wish to test the hypothesis that

$H_0 : \mu^{(1)} = \mu^{(2)}$

The mean of the first sample is normally distributed with expected value $E(\bar{x}^{(1)}) = \mu^{(1)}$ and covariance matrix

$$\Sigma_1 = \frac{1}{N_1} \sum_{\alpha=1}^{N_1} (x_\alpha^{(1)} - \bar{x}^{(1)})(x_\alpha^{(1)} - \bar{x}^{(1)})'$$

Similarly $E(\bar{x}^{(2)}) = \mu^{(2)}$

and $\Sigma_2 = \frac{1}{N_2} \sum_{\alpha=1}^{N_2} (x_\alpha^{(2)} - \bar{x}^{(2)})(x_\alpha^{(2)} - \bar{x}^{(2)})'$

Thus $(\bar{x}^{(1)} - \bar{x}^{(2)})$ has mean $(\mu^{(1)} - \mu^{(2)})$ and covariance matrix $\frac{\Sigma_1}{N_1} + \frac{\Sigma_2}{N_2}$

If $N_1 = N_2 = N$ (say)

and let $y_\alpha = x_\alpha^{(1)} - x_\alpha^{(2)}$

$$\bar{y} = \frac{1}{N} \sum y_\alpha$$

$$= \bar{x}^{(1)} - \bar{x}^{(2)}$$

$$\therefore \bar{y} \sim N\left((\mu^{(1)} - \mu^{(2)}), \frac{1}{N}(\Sigma_1 + \Sigma_2)\right)$$

$$S = \frac{1}{N-1} \sum_{\alpha=1}^{N} (Y_\alpha - \bar{Y})(Y_\alpha - \bar{Y})'$$

$\therefore$ our test statistic $T^2 = N \bar{Y}' S^{-1} \bar{Y}$ is distributed T^2 distribution with $(N-1)$ d.f.