Hotelling - T^2 Statistics

Hatteling - T^2 Statistics

Assumption for apply t-test :-

(1) Obs^n.s are ind. and normally distributed.
(2) Pop^n variance is unknown. i.e. σ is unknown.
(we use its estimate 's')
(3) Sample size should be less than 30 and greater than 15. If sample size is >30. Then we use z-test.

multivariate analogs of student-t^2 statistics is a Hatteling - T^2 statistics. If $x_1, x_2, \ldots x_n$ be a r.s. of size 'n' from normal dist^n with mean μ and var σ^2. i.e. $N(\mu, \sigma^2)$ where both μ and σ^2 are unknown and we wish to test $H_0: \mu = \mu_0$ v/s $H_1: \mu \neq \mu_0$ at the level of significance 'α'. Then our test statistics will be

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \sim t_{n-1} \text{ df}$$

where, $s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$ is an unbiased est^r of σ^2. If we square of this statistics, we get

$$t^2 = \frac{n(\bar{x} - \mu_0)^2}{s^2} = n(\bar{x} - \mu_0)' s^{-2} (\bar{x} - \mu_0) \quad \text{--- (1)}$$

The multivariate analogs of t^2 is given by,

$$T^2 = N(\underset{\sim}{\bar{X}} - \underset{\sim}{\mu_0})' \underset{\sim}{S}^{-1} (\underset{\sim}{\bar{X}} - \underset{\sim}{\mu_0}) \quad \text{--- (2)}$$

where, $\underset{\sim}{S} = \frac{1}{N-1} \sum_{\alpha=1}^{N} (\underset{\sim}{X}_\alpha - \underset{\sim}{\bar{X}})(\underset{\sim}{X}_\alpha - \underset{\sim}{\bar{X}})'$

and $\underset{\sim}{\bar{X}} = \frac{1}{N} \sum_{\alpha=1}^{N} \underset{\sim}{X}_\alpha$

② and μ_0 is specified value of $\mu(\mu_1, \mu_2, \ldots \mu_p)'$

Remark :- If $p=1$, then T^2 reduces to t^2.

Def : Let $X_\alpha \sim N_p(\mu, \Sigma)$; $\alpha = 1, 2, \ldots N$; $N > p$ be a r.s. where Σ be semi+ve definate matrix and unknown. Then for testing $H_0: \mu = \mu_0$ v/s $H_1: \mu \neq \mu_0$. The test statistics is given by

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

which is called the Hatteling- T^2 statistics.

where, $S = \frac{A}{N-1}$.

** Derivation of Hatteling- T^2 Statistics through Likelihood Ratio Approach :- Let $X_\alpha \sim N_p(\mu, \Sigma)$; $\alpha = 1, 2, \ldots N$; $N > p$. be a r.s. where Σ is semi +ve definate, and unknown (matrix)

Then for testing $H_0: \mu = \mu_0$ v/s $H_1: \mu \neq \mu_0$, we reject H_0 if the likelihood ratio

$$\lambda = \frac{\sup\limits_{\mu, \Sigma \in \Theta_0} L(X, \mu, \Sigma)}{\sup\limits_{\mu, \Sigma \in \Theta} L(X, \mu, \Sigma)} \le \lambda_0 \quad \text{---(1)}$$

where, λ_0 is a constant to be choosen such that $P_{H_0}[\lambda \le \lambda_0] = \alpha$ —— (2)

Now, the likelihood of the given sample is given by, ③

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

under $H_0: \mu = \mu_0$ and $\hat{\Sigma} = \frac{S}{N}$

under $H_1: \mu \neq \mu_0$ and $\hat{\Sigma} = \frac{A}{N}$,

where, $A = \sum_{\alpha=1}^{N} (X_\alpha - \bar{X})(X_\alpha - \bar{X})'$ and

$$S = \sum_{\alpha=1}^{N} (X_\alpha - \mu_0)(X_\alpha - \mu_0)'$$

$$= A + A_1$$

where, $A_1 = N(\bar{X} - \mu_0)(\bar{X} - \mu_0)'$

Then from eq^n (1), we have the likelihood ratio as

$$\lambda = \frac{\frac{1}{(2\pi)^{Np/2}} \cdot \frac{1}{|S/N|^{N/2}} \exp\left[-\frac{1}{2} tr \sum_{\alpha=1}^{N} (X_\alpha - \mu_0)' \left(\frac{S}{N}\right)^{-1} (X_\alpha - \mu_0)\right]}{\frac{1}{(2\pi)^{Np/2}} \cdot \frac{1}{|A/N|^{N/2}} \exp\left[-\frac{1}{2} tr \sum_{\alpha=1}^{N} (X_\alpha - \bar{X})' \left(\frac{A}{N}\right)^{-1} (X_\alpha - \bar{X})\right]}$$

$$= \frac{|A|^{N/2}}{|S|^{N/2}} \exp\left[-\frac{1}{2} tr(N S^{-1} S - N A^{-1} A)\right]$$

$$\lambda = \left(\frac{A}{S}\right)^{N/2}$$

$$\Rightarrow \lambda^{2/N} = \left|\frac{A}{S}\right| = \frac{|A|}{|A + A_1|}$$

$$= \frac{|A|}{|A| + N(\bar{X} - \mu_0)(\bar{X} - \mu_0)'}$$

$$= \frac{|A|}{\begin{vmatrix} 1 & \sqrt{N}(\bar{X}-\mu_0)' \\ -\sqrt{N}(\bar{X}-\mu_0) & A \end{vmatrix}}$$

$$= \frac{|A|}{|A|\,|1+N(\bar{X}-\mu_0)'A^{-1}(\bar{X}-\mu_0)|} \qquad \left\{ \begin{vmatrix} A & B \\ C & D \end{vmatrix} = |D|\,|A-BD^{-1}C| \right.$$

$$= \frac{1}{1+\frac{N}{N-1}(\bar{X}-\mu_0)'A^{-1}(N-1)(\bar{X}-\mu_0)}$$

$$= \frac{1}{1+\frac{N}{N-1}(\bar{X}-\mu_0)'S^{-1}(\bar{X}-\mu_0)} \qquad \left\{ \begin{array}{l} S = \frac{A}{N-1} \\ S^{-1} = A^{-1}(N-1) \end{array} \right.$$

$$\lambda^{2/N} = \frac{1}{1+\frac{T^2}{N-1}} \quad \text{under } H_0 \text{ ——— (3)}$$

$$\therefore \lambda^{2/N} \le \lambda_0^{2/N}$$

$$\Rightarrow \frac{1}{1+\frac{T^2}{N-1}} \le \lambda_0^{2/N}$$

from eq^n (1) and (2)

$$1+\frac{T^2}{N-1} \ge \lambda_0^{-2/N}$$

$$\Rightarrow \frac{T^2}{N-1} \ge \lambda_0^{-2/N} - 1$$

$$\Rightarrow T^2 \ge (N-1)\left(\lambda_0^{-2/N} - 1\right)$$

$$\Rightarrow \boxed{T^2 \ge T_0^2}$$

T_0^2 is defined as

$$P_{H_0}\left[T^2 \ge T_0^2\right] = \alpha$$

Thus, likelihood ratio test is the f^n of Hatteling-T^2 statistics.

Derivation of the dist^n of Hatteling-T^2 statistics

Thm:- Let $X_\alpha \sim N_p(\mu, \Sigma)$, $\alpha = 1, 2, \dots N$, $N > p$ be a r.s. and let $T^2 = N(\bar{X}-\mu_0)'S^{-1}(\bar{X}-\mu_0)$ where, $S = A/(N-1)$. Then, the dist^n of T^2

$\frac{T^2}{N-1} \times \frac{N-p}{p}$ is a non central F-dist^n with d.f. p and $(N-p)$. i.e. $\frac{T^2}{N-1} \times \frac{N-p}{p} \sim F_{NC}(p, N-p)$ with non central parameter

$$\lambda^2 = (\mu-\mu_0)'\Sigma^{-1}(\mu-\mu_0)$$

If $\mu = \mu_0$, Then F is central with d.f $(p, N-p)$

Proof:- we have $T^2 = N(\bar{X}-\mu_0)'S^{-1}(\bar{X}-\mu_0)$

and let $Y = \sqrt{N}(\bar{X}-\mu_0) \sim N_p(\nu, \Sigma)$ where ——(1) $\nu = \sqrt{N}(\mu-\mu_0)$

$$T^2 = Y'S^{-1}Y \quad \text{——— (2)}$$

Here, $A = nS \quad (n = N-1)$

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

Let C be a non singular square matrix of order p such that

$$C\Sigma C' = I$$

Then on making transformation such that

$y^* = Cy$ we get

$S^* = CSC'$

$\nu^* = C\nu C$

$z_\alpha^* = Cz_\alpha$

from eq^n (2), we have

$$T^{*2} = y^{*'} S^{*-1} y^* \quad \text{—— (3)}$$

where, $y^* = Cy \sim N_p(\nu^*, C\Sigma C')$

$\sim N_p(\nu^*, I)$

and $A^* = nS^* = \sum_{\alpha=1}^{n} z_\alpha^* z_\alpha^{*'}$

where $z_\alpha^* \sim N_p(0, I)$

Now, consider the orthogonal transformation

$$U = Qy^* \quad \text{—— (4)}$$

where, Q be a orthogonal matrix of order $p\times p$ with its first row as

$$q_{1i} = \frac{y_i^*}{\sqrt{y^{*'}y^*}} \quad \forall \; i = 1, 2, \dots p$$

(thus $\sum q_{1i} = 1$)

$\therefore$ we have

$$\begin{pmatrix} U_1 \\ U_2 \\ \vdots \\ U_p \end{pmatrix} = \begin{pmatrix} \frac{y_1^*}{\sqrt{y^{*'}y^*}} & \frac{y_2^*}{\sqrt{y^{*'}y^*}} & \cdots & \frac{y_p^*}{\sqrt{y^{*'}y^*}} \\ q_{21} & q_{22} & \cdots & q_{2p} \\ \vdots & & & \\ q_{p1} & q_{p2} & \cdots & q_{pp} \end{pmatrix} \begin{pmatrix} y_1^* \\ y_2^* \\ \vdots \\ y_p^* \end{pmatrix}$$

which provides us,

$$U_1 = \frac{y_1^* y_1^*}{\sqrt{y^{*'}y^*}} + \frac{y_2^* y_2^*}{\sqrt{y^{*'}y^*}} + \dots + \frac{y_p^* y_p^*}{\sqrt{y^{*'}y^*}}$$

$$= \frac{1}{\sqrt{y^{*'}y^*}} \left[y_1^{*2} + y_2^{*2} + \dots + y_p^{*2} \right]$$

$$= \frac{y^{*'}y^*}{\sqrt{y^{*'}y^*}}$$

$$U_1 = \sqrt{y^{*'}y^*} \quad \text{—— (a)}$$

And $U_j = \sum_{j=1}^{p} q_{ij} y_j^* \quad \forall \; i = 2, 3, \dots p$

$$= \sqrt{y^{*'}y^*} \; \frac{\sum_{j=1}^{p} q_{ij} y_j^*}{\sqrt{y^{*'}y^*}}$$

$$= \sqrt{y^{*'}y^*} \times 0$$

{$\because$ In orthogonal matrix mult. of two rows $= 0$}.

$$U_i = 0 \quad \forall \; i = 2, 3, \dots p$$

Hence,
$$U = \begin{bmatrix} \sqrt{y^{*'}y^*} \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{p\times 1} \quad \text{—— (5)}$$

Now, from eq^n (3), we have

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

(8)

$$= \underset{\sim}{U}' \underset{\sim}{Q}^{-1'} \underset{\sim}{S}^{*-1} \underset{\sim}{Q}^{-1} \underset{\sim}{U}$$

$$= \underset{\sim}{U}' (\underset{\sim}{Q} \underset{\sim}{S}^* \underset{\sim}{Q}')^{-1} \underset{\sim}{U}$$

$$= \underset{\sim}{U}' \left(\frac{\underset{\sim}{B}}{N-1} \right)^{-1} \underset{\sim}{U} \quad \left\{ \underset{\sim}{Q} \underset{\sim}{S}^* \underset{\sim}{Q}' = \frac{\underset{\sim}{B}}{N-1} \right\}$$

$$\Rightarrow \frac{T^{*2}}{N-1} = \underset{\sim}{U}' \underset{\sim}{B}^{-1} \underset{\sim}{U}$$

where, $\underset{\sim}{B} = \left[\begin{array}{c|ccc} b_{11} & b_{12} & \cdots & b_{1p} \\ \hline b_{21} & b_{22} & \cdots & b_{2p} \\ \vdots & & & \vdots \\ b_{p1} & b_{p2} & \cdots & b_{pp} \end{array}\right]$

$$\underset{\sim}{B} = \begin{bmatrix} b_{11} & b_{(1)} \\ b_{(1)}' & B_{22} \end{bmatrix}$$

Let $\underset{\sim}{B}^{-1} = \begin{bmatrix} b^{11} & b^{(1)} \\ b^{(1)'} & B^{22} \end{bmatrix}$

Then we have,

$$\frac{T^{*2}}{N-1} = (U_1 \; 0 \cdots 0) \begin{bmatrix} b^{11} & b^{(1)} \\ (b^{(1)'})^{-1} & B^{22} \end{bmatrix} \begin{bmatrix} U_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$= U_1^2 b^{11}$$

where, $(b^{11})^{-1} = b_{11} - b_{(1)} B_{22}^{-1} b_{(1)} = 1/b^{11}$

$= b_{11.2}$ (say)

$$\Rightarrow \frac{T^{*2}}{N-1} = \frac{U_1^2}{b_{11.2}} \quad \text{—— (6)}$$

(9)

As we have

$$\underset{\sim}{B} = \underset{\sim}{Q} (N-1) \underset{\sim}{S}^* \underset{\sim}{Q}'$$

$$= \underset{\sim}{Q} \sum_{\alpha=1}^{n} \underset{\sim}{Z}_\alpha^* \underset{\sim}{Z}_\alpha^{*'} \underset{\sim}{Q}'$$

$$= \sum_{\alpha=1}^{n} (\underset{\sim}{Q} \underset{\sim}{Z}_\alpha^*)(\underset{\sim}{Q} \underset{\sim}{Z}_\alpha^*)'$$

$$= \sum_{\alpha=1}^{n} W_\alpha W_\alpha' \text{ where, } W_\alpha = \underset{\sim}{Q} \underset{\sim}{Z}_\alpha^* \sim N_p(0, I)$$

Then, we have, $b_{11.2} = \sum_{\alpha=1}^{(N-1)-(p-1)} W_\alpha W_\alpha' = \sum_{\alpha=1}^{N-p} W_\alpha W_\alpha'$

$\Rightarrow b_{11.2} \sim \chi_{N-p}$

Then from eq^n (6), we have $\frac{T^2}{N-1} = \frac{y' y^*}{b_{11.2}} = \frac{\chi_p^2}{\chi_{N-p}^2}$

$U_1^2 = y^{*'} y^*$

$U_1^2 \sim \chi_p^2$ (Non Central)

from eq^n (6)

$$\therefore \frac{T^2}{N-1} \sim \frac{\chi_p^2}{\chi_{N-p}^2}$$

$$\Rightarrow \boxed{\frac{T^2/p}{N-1/N-p}} = \frac{\chi_p^2/p}{\chi_{N-p}^2/N-p} \sim F_{(p, N-p)}$$

with non central par λ^2. where, $\lambda^2 = N(\underset{\sim}{\mu} - \underset{\sim}{\mu}_0)' \Sigma^{-1} (\underset{\sim}{\mu} - \underset{\sim}{\mu}_0)$

Proved.

The pdf of Hatteling T^2 Statistics :- The pdf of non-central F dist^n with d.f p and $(N-p)$ is given by

Scanned with CamScanner

(10)

$$f(z) = \begin{cases} \frac{p}{N-p} e^{-\lambda^2/2} \left[\sum_{j=0}^{\infty} \frac{(\lambda^2/2)^j}{j!} \frac{1}{B\left(\frac{p}{2}+j, \frac{N-p}{2}\right)} \cdot \frac{\left(\frac{p}{N-p} z\right)^{\frac{p}{2}+j-1}}{\left(1+\frac{p}{N-p}z\right)^{\frac{N}{2}+j}} \right] \\ 0 \quad ; \quad ow \end{cases}$$

Let $z = \frac{T^2}{N-1} \cdot \frac{N-p}{p}$

$dz = \frac{dT^2}{N-1} \cdot \frac{N-p}{p}$

$$f(T^2) = \frac{1}{N-1} e^{-\lambda^2/2} \left[\sum_{j=0}^{\infty} \frac{(\lambda^2/2)^j}{j!} \frac{1}{B\left(\frac{p}{2}+j, \frac{N-p}{2}\right)} \frac{\left(\frac{T^2}{N-1}\right)^{\frac{p}{2}+j-1}}{\left(1+\frac{T^2}{N-1}\right)^{\frac{N}{2}+j}} \right]$$

where, $\lambda^2 = N(\underset{\sim}{\mu}-\underset{\sim}{\mu_0})' \underset{\sim}{\Sigma}^{-1} (\underset{\sim}{\mu}-\underset{\sim}{\mu_0})$

If $p=1$ and $T^2 = t^2$ with $\lambda^2 = 0$, Then

$$f(t^2) = \frac{1}{nB\left(\frac{1}{2}, \frac{n}{2}\right)} \frac{(t^2/n)^{\frac{1}{2}-1}}{\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}}$$

$$f(t) = \frac{1}{n^{1/2}B\left(\frac{1}{2}, \frac{n}{2}\right)} \cdot \frac{2}{\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}} \quad ; \quad t \geq 0$$

$$f(t) = \frac{1}{n^{1/2}B\left(\frac{1}{2}, \frac{n}{2}\right)} \cdot \frac{1}{\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}} \quad ; \quad -\infty \leq t < \infty$$

Applications of Hotteling T^2 statistics :- (11)

(1) Testing of $\underset{\sim}{\mu}$ when $\underset{\sim}{\Sigma}$ is unknown :-

Let $\underset{\sim}{X}_\alpha$ be iid $N_p(\underset{\sim}{\mu}, \underset{\sim}{\Sigma})$, $\alpha = 1, 2, \dots N$, $N>p$ and $\underset{\sim}{\Sigma}$ be semi +ve definite matrix and unknown. Then for testing $H_0 : \underset{\sim}{\mu} = \underset{\sim}{\mu_0}$ v/s $H_1 : \underset{\sim}{\mu} \neq \underset{\sim}{\mu_0}$. Then the statistics is given by

$$T^2 = N(\bar{\underset{\sim}{X}} - \underset{\sim}{\mu})' \underset{\sim}{S}^{-1} (\bar{\underset{\sim}{X}} - \underset{\sim}{\mu}), \quad \underset{\sim}{S} = \underset{\sim}{A}/N-1$$

The critical region is given by

Reject H_0 if $T^2 \geq T_0^2(\alpha)$ —(1)

where $T_0^2(\alpha)$ is such that

$$P[T^2 \geq T_0^2(\alpha) | H_0] = \alpha$$

eq^n (1) is equivalent to

$$\frac{T^2}{N-1} \cdot \frac{N-p}{p} \geq \frac{T_0^2(\alpha)}{N-1} \cdot \frac{N-p}{p}$$

$$F_{p, N-p} \geq F_{p, N-p}(\alpha)$$

where, $\frac{T_0^2(\alpha)}{N-1} \cdot \frac{N-p}{p} = F_{p, N-p}(\alpha)$ is the $\alpha \times 100\%$ upper point of non-central F-dist^n with p and $N-p$ d.f. The confidence region for $\underset{\sim}{\mu}$ when $\underset{\sim}{\Sigma}$ is unknown is given by

$$P[T^2 \leq T_0^2(\alpha)] = 1-\alpha$$

where, $\frac{T_0^2(\alpha)}{N-1} \cdot \frac{N-p}{p} = F_{p, N-p}(\alpha)$

where, $F_{p,N-p}(\alpha)$ is s.t.

$$P\left[F \le F_{p,N-p}(\alpha)\right] = 1-\alpha$$

Thus, $(1-\alpha)\times 100\%$ confidence region for $\underset{\sim}{\mu}$ is given by the set of values which satisfy the inequality $T^2 \le T_0^2(\alpha)$. where confidence region is ellipsoid in n-dimentional euclidean space with central at $\bar{\underset{\sim}{X}}$ with size and shape depending on $\underset{\sim}{S}$ and $T_0^2(\alpha)$.

Two Sample Problem :- Let

$$\underset{\sim}{X}_\alpha^{(1)} \sim N_p(\underset{\sim}{\mu}^{(1)}, \underset{\sim}{\Sigma});\ \alpha = 1,2,\dots N_1,\ N_1>p$$

$$\underset{\sim}{X}_\alpha^{(2)} \sim N_p(\underset{\sim}{\mu}^{(2)}, \underset{\sim}{\Sigma});\ \alpha = 1,2,\dots N_2,\ N_2>p$$

be two r.s. of size N_1 and N_2 and $\underset{\sim}{\Sigma}$ is +ve definate and unknown. The problem is to test $H_0: \underset{\sim}{\mu}^{(1)} = \underset{\sim}{\mu}^{(2)}$ v/s $H_1: \underset{\sim}{\mu}^{(1)} \ne \underset{\sim}{\mu}^{(2)}$. For testing the inequality of mean vector of two multivariate normal popn with same but unknown var-cov matrix $\underset{\sim}{\Sigma}$, we use Hatteling-T^2 statistics as follows

Here, $\bar{\underset{\sim}{X}}^{(i)} = \frac{1}{N_i}\sum_{\alpha=1}^{N_i} \underset{\sim}{X}_{(\alpha)}^{(i)}$; $i = 1,2$

Here $\underset{\sim}{S}$ is a palled estimate of var-cov matrix i.e. $\underset{\sim}{S} = \frac{1}{N_1+N_2-2}\sum_{i=1}^{2}\sum_{\alpha=1}^{N_i}(\underset{\sim}{X}_\alpha^{(i)} - \bar{\underset{\sim}{X}}_\alpha^{(i)})(\underset{\sim}{X}_{i\alpha}^{(i)} - \bar{\underset{\sim}{X}}_{i\alpha}^{(i)})'$

$$\underset{\sim}{S} = \frac{1}{n}\sum_{\alpha=1}^{n} \underset{\sim}{Z}_\alpha \underset{\sim}{Z}_\alpha' \text{ where, } \underset{\sim}{Z}_\alpha \sim N_p(\underset{\sim}{0}, \underset{\sim}{\Sigma});$$

$n = N_1 + N_2 - 2$

Here, $\bar{\underset{\sim}{X}}^{(1)} - \bar{\underset{\sim}{X}}^{(2)} \sim N_p\left(\underset{\sim}{0}, \frac{N_1+N_2}{N_1N_2}\underset{\sim}{\Sigma}\right)$ under H_0 ;

$$\therefore \sqrt{\frac{N_1N_2}{N_1+N_2}}(\bar{\underset{\sim}{X}}^{(1)} - \bar{\underset{\sim}{X}}^{(2)}) \sim N_p(\underset{\sim}{0}, \underset{\sim}{\Sigma})$$

Also, $(\bar{\underset{\sim}{X}}^{(1)} - \bar{\underset{\sim}{X}}^{(2)})$ and $\underset{\sim}{S}$ are independant. Therefore, by def of T^2 statistics

$$T^2 = \frac{N_1N_2}{N_1+N_2}(\bar{\underset{\sim}{X}}^{(1)} - \bar{\underset{\sim}{X}}^{(2)})'\underset{\sim}{S}^{-1}(\bar{\underset{\sim}{X}}^{(1)} - \bar{\underset{\sim}{X}}^{(2)})$$

$$\Rightarrow \frac{T^2}{N_1+N_2-2}\cdot\frac{N_1+N_2-p-1}{p} \sim F_{p,\underbrace{N_1+N_2-p-1}_{n-p+1}}$$

$\therefore$ The critical region is given by

$F_{p,N_1+N_2-p-1} \geq F_{p,N_1+N_2-p-1}(\alpha)$; Reject H_0

where, $F_{p,N_1+N_2-p-1}(\alpha)$ is the $\alpha\times 100\%$ upper points of F-distn with df p and (N_1+N_2-p-1).

Now, confidence region of

$$\underset{\sim}{\mu}^{(1)} - \underset{\sim}{\mu}^{(2)} = \underset{\sim}{\nu}$$

A $(1-\alpha)\times 100\%$ confidence region of $\underset{\sim}{\nu}$ consist of set of values of $\underset{\sim}{\nu}$ satisfying

$$\frac{N_1N_2}{N_1+N_2}(\bar{\underset{\sim}{X}}^{(1)} - \bar{\underset{\sim}{X}}^{(2)} - \underset{\sim}{\nu})'\underset{\sim}{S}^{-1}(\bar{\underset{\sim}{X}}^{(1)} - \bar{\underset{\sim}{X}}^{(2)} - \underset{\sim}{\nu}) \le T_0^2(\alpha)$$

$$\Rightarrow P\left[\frac{N_1N_2}{N_1+N_2}(\bar{\underset{\sim}{X}}^{(1)} - \bar{\underset{\sim}{X}}^{(2)} - \underset{\sim}{\nu})\underset{\sim}{S}^{-1}(\bar{\underset{\sim}{X}}^{(1)} - \bar{\underset{\sim}{X}}^{(2)} - \underset{\sim}{\nu}) \le T_0^2(\alpha)\right] = 1-\alpha.$$

q Sample Problem :- Let $\underset{\sim}{X}_{(\alpha)}^{(i)} \sim N_p(\underset{\sim}{\mu}^{(i)}, \underset{\sim}{\Sigma})$; $i = 1,2,\dots q$, $\alpha = 1,2,\dots N_q > p$.

where $\underset{\sim}{\Sigma}$ is the semi +ve definate and unknown

The problem is to test $H_0: \sum_{i=1}^{q}\beta_i\underset{\sim}{\mu}^{(i)} = \underset{\sim}{\mu}_0$ v/s

(14)

$H_1: \sum_{i=1}^{q} \beta_i \mu^{(i)} \neq \mu_0$

Here, $\bar{X}^{(i)} = \frac{1}{N_i} \sum_{\alpha=1}^{N_i} X_\alpha^{(i)}$,

$$S = \frac{1}{\sum_{i=1}^{q} N_i - q} \sum_{i=1}^{q} \sum_{\alpha=1}^{N_i} (X_\alpha^{(i)} - \bar{X}^{(i)})(X_\alpha^{(i)} - \bar{X}^{(i)})'$$

Suppose, $Z = \sum_{i=1}^{q} \beta_i \bar{X}^{(i)} \sim N_p(\mu_0, \Sigma/c)$ under H_0

where, $\frac{1}{c} = \frac{\sum \beta_i^2}{N_i}$

$\Rightarrow \sqrt{c}(Z - \mu_0) \sim N_p(0, \Sigma)$

Also, $nS = \sum_{\alpha=1}^{n} Z_\alpha Z_\alpha'$, where $n = \sum_{i=1}^{q} N_i - q$,

$; Z_\alpha \sim N_p(0, \Sigma)$

Here, Z and S are independent, Then by def of T^2 statistics, we have

$$\frac{T^2}{n} \cdot \frac{n-p+1}{p} = F_{p, n-p+1}$$

and the critical region is given by

$$F_{p, n-p+1} \geq F_{p, n-p+1}^{(\alpha)} \text{ ; Reject } H_0$$

where, $F_{p, n-p+1}(\alpha)$ is the $\alpha \times 100\%$ upper point of F- dist^n with d.f p and $(n-p+1)$ and

$$P_{H_0}\left[F_{p, n-p+1} \geq F_{p, n-p+1}^{(\alpha)}\right] = \alpha$$

Mahalanobis D^2- Statistics :- Prof. Mahalanobies

has introduce a name D^2- statistics using the concept of distance b/w two variate normal pop^n. Such as $N_p(\mu^{(1)}, \Sigma)$ and $N_p(\mu^{(2)}, \Sigma)$, Then the distance b/w them is defined as

$$D^2 = (\mu^{(1)} - \mu^{(2)})' \Sigma^{-1} (\mu^{(1)} - \mu^{(2)})$$

If the par are unknown, Then we use unbiased estimate of the par and then D^2 statistics is defined as

$$D^2 = (\bar{X}^{(1)} - \bar{X}^{(2)})' S^{-1} (\bar{X}^{(1)} - \bar{X}^{(2)}) \quad \text{--- (1)}$$

Here, $S = \frac{A_1 + A_2}{N_1 + N_2 - 2}$

This D^2 statistics is called mahalanobies D^2 statistics.

As we know that in case of two sample problem $T^2 = \frac{N_1 N_2}{N_1 + N_2} (\bar{X}^{(1)} - \bar{X}^{(2)})' S^{-1} (\bar{X}^{(1)} - \bar{X}^{(2)})$

** $\boxed{T^2 = \frac{N_1 N_2}{N_1 + N_2} D^2}$

$\therefore$ The uses of D^2 and two samples T^2 statistics can be put as identical.

In particular for testing $H_0: \mu^{(1)} = \mu^{(2)}$

$$\frac{T^2}{N_1 + N_2 - 2} \cdot \frac{N_1 + N_2 - p - 1}{p} \sim F_{p, N_1 + N_2 - p - 1}$$

(15)

Scanned with CamScanner

(16)

$$\Rightarrow \frac{N_1 N_2}{N_1+N_2} D^2 \cdot \frac{1}{N_1+N_2-2} \cdot \frac{N_1+N_2-p-1}{p} \sim F_{p,\, N_1+N_2-p-1}$$

14th Application

Problem of Symmetry:- Let $\underset{\sim}{X}_\alpha \sim N_p(\underset{\sim}{\mu}, \underset{\sim}{\Sigma})$, $\alpha = 1, 2, \ldots, N$, $N > p$. where, $\underset{\sim}{\Sigma}$ is +ve definate and unknown. and the problem is to test $H_0 : \mu_1 = \mu_2 = \cdots = \mu_p = \mu$ (say) v/s H_1 : The inequality holds atleast one place.

Let $\underset{\sim}{C}_{(p-1)\times p}$ be the constant matrix of order $(p-1)\times p$, such that $\underset{\sim}{C}\underset{\sim}{\epsilon} = \underset{\sim}{0}$ and rank $(p-1)$ where, $\underset{\sim}{\epsilon} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}_{p\times 1}$ be a column vector containing each element as 1

$$\underset{\sim}{y}_\alpha = \underset{\sim}{C}\underset{\sim}{X}_\alpha \sim N_{p-1}(\underset{\sim}{C}\underset{\sim}{\mu}, \underset{\sim}{C}\underset{\sim}{\Sigma}\underset{\sim}{C}') \text{ under } H_0$$

$$\underset{\sim}{C}\mu\underset{\sim}{\epsilon} = \mu\underset{\sim}{C}\underset{\sim}{\epsilon} = \mu\cdot\underset{\sim}{0} = \underset{\sim}{0}$$

Now, $H_0 : \underset{\sim}{C}\underset{\sim}{\mu} = \cancel{\underset{\sim}{C}\mu\underset{\sim}{\Sigma} = \mu\underset{\sim}{C}\underset{\sim}{\Sigma} = \mu\cdot\underset{\sim}{0}} = \underset{\sim}{0}$

$$\therefore \underset{\sim}{y}_\alpha \sim N_{p-1}(\underset{\sim}{0}, \underset{\sim}{C}\underset{\sim}{\Sigma}\underset{\sim}{C}') \text{ under } H_0$$

$$\bar{\underset{\sim}{y}}_\alpha \sim N_{p-1}(\underset{\sim}{0}, \underset{\sim}{C}\underset{\sim}{\Sigma}\underset{\sim}{C}'/N)$$

$$\text{and } \sqrt{N}\,\bar{\underset{\sim}{y}} \sim N_{p-1}(\underset{\sim}{0}, \underset{\sim}{C}\underset{\sim}{\Sigma}\underset{\sim}{C}')$$

Also
$$\underset{\sim}{S}_y = \frac{1}{N-1}\sum_{\alpha=1}^{N}(\underset{\sim}{y}_\alpha - \bar{\underset{\sim}{y}}_\alpha)(\underset{\sim}{y}_\alpha - \bar{\underset{\sim}{y}}_\alpha)'$$

$$= \frac{1}{N-1}\underset{\sim}{C}\sum_{\alpha=1}^{N}(\underset{\sim}{X}_\alpha - \bar{\underset{\sim}{X}})(\underset{\sim}{X}_\alpha - \bar{\underset{\sim}{X}})'\underset{\sim}{C}'$$

$$\underset{\sim}{S}_y = \underset{\sim}{C}\underset{\sim}{S}_x\underset{\sim}{C}'$$

$$\Rightarrow (N-1)\underset{\sim}{S}_y = \sum_{\alpha=1}^{n}\underset{\sim}{Z}_\alpha^{*}\underset{\sim}{Z}_\alpha^{*\prime} \text{ where,}$$

$$\underset{\sim}{Z}_\alpha^{*} = \underset{\sim}{C}\underset{\sim}{Z}_\alpha \sim N_p(\underset{\sim}{0}, \underset{\sim}{C}\underset{\sim}{\Sigma}\underset{\sim}{C}')$$

$$\text{and } \underset{\sim}{Z}_\alpha \sim N_p(\underset{\sim}{0}, \underset{\sim}{\Sigma})$$

Then by def of T^2 statistics (17)

$$T^2 = N\bar{\underset{\sim}{y}}'\underset{\sim}{S}_y^{-1}\bar{\underset{\sim}{y}}$$

and the corresponding F-statistics is given by

$$F_{p-1,\, n-(p-1)+1}$$

$\therefore$ The critical region is given by ~~reject~~ reject H_0 if

$$F_{p-1,\, n-(p-1)+1} \geq F_{p-1,\, n-(p-1)+1}(\alpha)$$

where, $F_{p-1,\, n-(p-1)+1}(\alpha)$ is such that $\alpha \times 100\%$ upper pt of F-distn with df $p-1,\, n-(p-1)+1$

i.e. $P_{H_0}\left[F_{p-1,\, n-(p-1)+1} \geq F_{p-1,\, n-(p-1)+1}(\alpha)\right] = \alpha$

$$\Rightarrow H_0 : \underset{\sim}{C}\underset{\sim}{\mu} = \underset{\sim}{C}\mu\underset{\sim}{\epsilon} = \mu\underset{\sim}{C}\underset{\sim}{\epsilon} = \mu\times\underset{\sim}{0} = \underset{\sim}{0}$$

$$\because \underset{\sim}{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \end{pmatrix}_{p\times 1} = \begin{pmatrix} \mu \\ \mu \\ \vdots \\ \mu \end{pmatrix} = \mu\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \mu\underset{\sim}{\epsilon}$$

Anderson
Page No - 182