Chapter 5
Generalized T^2 Statistics

The test of hypothesis concerning the population mean μ is conducted using the t-statistic given by:

$$t = \frac{\sqrt{N}(\bar{x} - \mu)}{s}$$

where $\bar{x}$ and s^2 are the sample mean and sample variance based on a sample of size N from a univariate normal distribution $N(\mu, \sigma^2)$ where σ^2 is unknown. Under the null hypothesis $H_0 : \mu = \mu_0$ (known μ_0), the t-statistic follows the t-distribution with $N - 1$ degrees of freedom.

In the multivariate setup, the null hypothesis $H_0 : \underset{\sim}{\mu} = \underset{\sim}{\mu}_0$ is tested using a multivariate analogue of t-statistics, which is essentially a multivariate analogue of the square of t-statistics, based on a sample from a multivariate normal population.

The multivariate analogue of the square of $t = \frac{\sqrt{N}(\bar{x} - \mu)}{s}$ is

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

where $\underset{\sim}{\bar{x}}$ is the sample mean vector and S is the sample covariance matrix based on a sample of size N, and $S = \frac{A}{N-1}$. This is called the **Generalized T^2-statistic** or **Hotelling's T^2-statistic**.

This statistic is used for testing the hypothesis and for obtaining the confidence regions for unknown $\underset{\sim}{\mu}$ from $N_p(\underset{\sim}{\mu}, \Sigma)$.

The geometric method, likelihood function approach, and Union-Intersection principle (given by Professor S.N. Roy) are used to obtain the T^2-statistic.

Derivation of T^2-Statistic as a Function of the Likelihood Ratio Criterion

We consider the likelihood ratio test of hypothesis $H_0 : \underset{\sim}{\mu} = \underset{\sim}{\mu}_0$ on the basis of a sample from $N(\underset{\sim}{\mu}, \Sigma)$.

Let $\underset{\sim}{X} \sim N_p(\underset{\sim}{\mu}, \Sigma)$. Suppose a sample of N observations $\underset{\sim}{x}_1, \underset{\sim}{x}_2, \ldots, \underset{\sim}{x}_N$ $(N > p)$ is available. Suppose the null hypothesis to be tested is $H_0 : \underset{\sim}{\mu} = \underset{\sim}{\mu}_0$, where $\underset{\sim}{\mu}_0$ is known and Σ is unknown.

The likelihood function is

$$L \equiv L(\underset{\sim}{\mu}, \Sigma) = \frac{1}{(2\pi)^{\frac{Np}{2}} |\Sigma|^{\frac{N}{2}}} \exp\left[-\frac{1}{2} \sum_{\alpha=1}^{N} (\underset{\sim}{x}_\alpha - \underset{\sim}{\mu})' \Sigma^{-1} (\underset{\sim}{x}_\alpha - \underset{\sim}{\mu}) \right].$$

The likelihood ratio test statistic is

$$\lambda = \frac{\max\limits_{H_0} L}{\max\limits_{\Omega} L} = \frac{\max\limits_{\Sigma} L(\underset{\sim}{\mu}_0, \Sigma)}{\max\limits_{\underset{\sim}{\mu}, \Sigma} L(\underset{\sim}{\mu}, \Sigma)}$$

The parametric space under H_0 is $\hat{\Omega}_{H_0} = [\underset{\sim}{\mu}_0, \hat{\Sigma}_\omega]$ and the whole parametric space Ω is $\hat{\Omega}_\Omega = [\hat{\underset{\sim}{\mu}}_\Omega, \hat{\Sigma}_\Omega]$.

When the parameters are unrestricted, the maximum occurs when $\underset{\sim}{\mu}$ and Σ are defined by the maximum likelihood estimates as

$$\hat{\underset{\sim}{\mu}}_\Omega = \bar{\underset{\sim}{x}}$$

$$\hat{\Sigma}_\Omega = \frac{1}{N} \sum_{\alpha=1}^{N} (\underset{\sim}{x}_\alpha - \bar{\underset{\sim}{x}})(\underset{\sim}{x}_\alpha - \bar{\underset{\sim}{x}})' = \frac{A}{N}.$$

Note that the maximum likelihood estimators are obtained by maximizing the likelihood function, so when these estimators are substituted into the likelihood function, the likelihood function will automatically be maximized.

Under $H_0 : \underset{\sim}{\mu} = \underset{\sim}{\mu}_0$, the likelihood function is maximized at

$$\hat{\Sigma}_\omega = \frac{1}{N} \sum_{\alpha=1}^{N} (\underset{\sim}{x}_\alpha - \underset{\sim}{\mu}_0)(\underset{\sim}{x}_\alpha - \underset{\sim}{\mu}_0)'$$

and $\underset{\sim}{\mu} = \underset{\sim}{\mu_0}$ is known.

Now we substitute these estimators into the likelihood function and obtain the likelihood function under H_0 and under Ω, to be used in the likelihood ratio test statistic λ.

Calculating $\underset{H_0}{\max} L$:

Consider maximizing the likelihood under H_0.

$$
\begin{aligned}
\max_{H_0} L &= \max_{H_0} L(\mu_0, \hat{\Sigma}_\omega) \\
&= \frac{1}{(2\pi)^{\frac{Np}{2}} |\hat{\Sigma}_\omega|^{\frac{N}{2}}} \exp\left[-\frac{1}{2} \left\{ \sum_{\alpha=1}^{N} (\underset{\sim}{x}_\alpha - \underset{\sim}{\mu_0})' \hat{\Sigma}_\omega^{-1} (\underset{\sim}{x}_\alpha - \underset{\sim}{\mu_0}) \right\} \right] \\
&= \frac{1}{(2\pi)^{\frac{Np}{2}} |\hat{\Sigma}_\omega|^{\frac{N}{2}}} \exp\left[-\frac{1}{2} tr\left(\hat{\Sigma}_\omega^{-1} \sum_{\alpha=1}^{N} (\underset{\sim}{x}_\alpha - \underset{\sim}{\mu_0})(\underset{\sim}{x}_\alpha - \underset{\sim}{\mu_0})' \right) \right] \\
&= \frac{1}{(2\pi)^{\frac{Np}{2}} |\hat{\Sigma}_\omega|^{\frac{N}{2}}} \exp\left[-\frac{1}{2} tr\left(\hat{\Sigma}_\omega^{-1} \cdot N\hat{\Sigma}_\omega \right) \right] \qquad \text{(Substituting the definition of } \hat{\Sigma}_\omega) \\
&= \frac{1}{(2\pi)^{\frac{Np}{2}} |\hat{\Sigma}_\omega|^{\frac{N}{2}}} \exp\left[-\frac{N}{2} tr(\hat{\Sigma}_\omega^{-1} \hat{\Sigma}_\omega) \right] \\
&= \frac{1}{(2\pi)^{\frac{Np}{2}} |\hat{\Sigma}_\omega|^{\frac{N}{2}}} \exp\left[-\frac{N}{2} tr(I_p) \right] \\
&= \frac{1}{(2\pi)^{\frac{Np}{2}} |\hat{\Sigma}_\omega|^{\frac{N}{2}}} \exp\left[-\frac{Np}{2} \right].
\end{aligned}
$$

Consider

$$
\begin{aligned}
A_0 &= \sum_{\alpha=1}^{N} (\underset{\sim}{x}_\alpha - \underset{\sim}{\mu})(\underset{\sim}{x}_\alpha - \underset{\sim}{\mu})' \\
&= \sum_{\alpha=1}^{N} [(\underset{\sim}{x}_\alpha - \bar{\underset{\sim}{x}}) + (\bar{\underset{\sim}{x}} - \underset{\sim}{\mu})][(\underset{\sim}{x}_\alpha - \bar{\underset{\sim}{x}}) + (\bar{\underset{\sim}{x}} - \underset{\sim}{\mu})]' \\
&= \sum_{\alpha=1}^{N} (\underset{\sim}{x}_\alpha - \bar{\underset{\sim}{x}})(\underset{\sim}{x}_\alpha - \bar{\underset{\sim}{x}})' + N(\bar{\underset{\sim}{x}} - \underset{\sim}{\mu})(\bar{\underset{\sim}{x}} - \underset{\sim}{\mu})' \\
&= A + N(\bar{\underset{\sim}{x}} - \underset{\sim}{\mu})(\bar{\underset{\sim}{x}} - \underset{\sim}{\mu})'
\end{aligned}
$$

where $A = \sum_{\alpha=1}^{N} (\underset{\sim}{x}_\alpha - \bar{\underset{\sim}{x}})(\underset{\sim}{x}_\alpha - \bar{\underset{\sim}{x}})'$.

Calculating $\max\limits_{\Omega} L$:

Consider maximizing the likelihood under Ω.

$$\max_{\Omega} L = \max_{\Omega} L(\hat{\mu}_{\Omega}, \hat{\Sigma}_{\Omega})$$

$$= \frac{1}{(2\pi)^{\frac{Np}{2}} |\hat{\Sigma}_{\Omega}|^{\frac{N}{2}}} \exp\left[-\frac{1}{2}\sum_{\alpha=1}^{N}(x_{\alpha} - \bar{x})'\hat{\Sigma}_{\Omega}^{-1}(x_{\alpha} - \bar{x})\right]$$

$$= \frac{1}{(2\pi)^{\frac{Np}{2}} |\hat{\Sigma}_{\Omega}|^{\frac{N}{2}}} \exp\left[-\frac{1}{2}tr\left(\hat{\Sigma}_{\Omega}^{-1}\sum_{\alpha=1}^{N}(x_{\alpha} - \bar{x})(x_{\alpha} - \bar{x})'\right)\right]$$

$$= \frac{1}{(2\pi)^{\frac{Np}{2}} |\hat{\Sigma}_{\Omega}|^{\frac{N}{2}}} \exp\left[-\frac{1}{2}tr(\hat{\Sigma}_{\Omega}^{-1}A)\right]$$

$$= \frac{1}{(2\pi)^{\frac{Np}{2}} |\hat{\Sigma}_{\Omega}|^{\frac{N}{2}}} \exp\left[-\frac{N}{2}tr(A^{-1}A)\right]$$

$$= \frac{1}{(2\pi)^{\frac{Np}{2}} |\hat{\Sigma}_{\Omega}|^{\frac{N}{2}}} \exp\left(-\frac{Np}{2}\right).$$

Thus, the likelihood ratio test statistic is

$$\lambda = \frac{\left(\frac{1}{(2\pi)^{\frac{Np}{2}} |\hat{\Sigma}_{\omega}|^{\frac{N}{2}}} \exp\left(-\frac{Np}{2}\right)\right)}{\left(\frac{1}{(2\pi)^{\frac{Np}{2}} |\hat{\Sigma}_{\Omega}|^{\frac{N}{2}}} \exp\left(-\frac{Np}{2}\right)\right)}$$

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

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

Now we simplify $\lambda^{2/N}$ as

$$
\lambda^{2/N} = \frac{|\sum_{\alpha=1}^{N}(\underset{\sim}{x}_\alpha - \underset{\sim}{\bar{x}})(\underset{\sim}{x}_\alpha - \underset{\sim}{\bar{x}})'|}{|\sum_{\alpha=1}^{N}(\underset{\sim}{\bar{x}} - \underset{\sim}{\mu}_0)(\underset{\sim}{\bar{x}} - \underset{\sim}{\mu}_0)'|}
$$

$$
= \frac{|A|}{\left| \sum_{\alpha=1}^{N}\left\{(\underset{\sim}{x}_\alpha - \underset{\sim}{\bar{x}}) + (\underset{\sim}{\bar{x}} - \underset{\sim}{\mu}_0)\right\}\left\{(\underset{\sim}{x}_\alpha - \underset{\sim}{\bar{x}}) + (\underset{\sim}{\bar{x}} - \underset{\sim}{\mu}_0)\right\}' \right|}
$$

$$
= \frac{|A|}{|A + N(\underset{\sim}{\bar{x}} - \underset{\sim}{\mu}_0)(\underset{\sim}{\bar{x}} - \underset{\sim}{\mu}_0)'|}
$$

$$
= \frac{|A|}{|A + N(\underset{\sim}{\bar{x}} - \underset{\sim}{\mu}_0)(\underset{\sim}{\bar{x}} - \underset{\sim}{\mu}_0)'|}
$$

$$
= \frac{|A|}{\left| \begin{matrix} 1 & \sqrt{N}(\underset{\sim}{\bar{x}} - \underset{\sim}{\mu}_0)' \\ -\sqrt{N}(\underset{\sim}{\bar{x}} - \underset{\sim}{\mu}_0) & A \end{matrix} \right|}
$$

$$
= \frac{1}{1 + N(\underset{\sim}{\bar{x}} - \underset{\sim}{\mu}_0)'A^{-1}(\underset{\sim}{\bar{x}} - \underset{\sim}{\mu}_0)}
$$

$$
= \frac{1}{1 + \frac{T^2}{N-1}}
$$

where

$$
T^2 = N(\underset{\sim}{\bar{x}} - \underset{\sim}{\mu}_0)'S^{-1}(\underset{\sim}{\bar{x}} - \underset{\sim}{\mu}_0)
$$

$$
= N(N-1)(\underset{\sim}{\bar{x}} - \underset{\sim}{\mu}_0)'A^{-1}(\underset{\sim}{\bar{x}} - \underset{\sim}{\mu}_0).
$$

The following result is used in deriving the expression for $\lambda^{2/N}$.

Result 1: For $B \neq 0$, and using the result:

$$
\left| \begin{matrix} B & C \\ D & E \end{matrix} \right| = \left| \begin{matrix} B & C \\ D & E \end{matrix} \right| \cdot \left| \begin{matrix} I & -B^{-1}C \\ 0 & I \end{matrix} \right|
$$

$$
= \left| \begin{matrix} B & 0 \\ D & E - DB^{-1}C \end{matrix} \right|
$$

$$
= |B| \cdot |E - DB^{-1}C|.
$$

Now the decision rule for H_0 based on likelihood ratio test statistic λ is to:

Reject H_0 if $\lambda \leq \lambda_0$

where λ_0 is chosen such that $P(\lambda \leq \lambda_0)$, when H_0 is true, is equal to the level of significance α, i.e. $P(\lambda \leq \lambda_0 | H_0) = \alpha$.

Now simplify λ as follows:

$$\lambda \leq \lambda_0$$

$$\text{or} \quad \lambda^{2/N} \leq \lambda_0^{2/N}$$

$$\text{or} \quad \lambda^{-2/N} \geq \lambda_0^{-2/N}$$

$$\text{or} \quad (\lambda^{-2/N} - 1) \geq (\lambda_0^{-2/N} - 1)$$

$$\text{or} \quad (N-1)(\lambda^{-2/N} - 1) \geq (N-1)(\lambda_0^{-2/N} - 1).$$

Substituting $\lambda^{-2/N} = 1 + \frac{T^2}{N-1}$:

$$(N-1)\left(1 + \frac{T^2}{N-1} - 1\right) \geq (N-1)\left(1 + \frac{T_0^2}{N-1} - 1\right)$$

$$\text{or} \qquad T^2 \geq T_0^2(\alpha)$$

where $T_0^2(\alpha)$ is a constant to be determined by the size condition and

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

with $P(\lambda \leq \lambda_0 | H_0) = \alpha$.

The next step is to determine the distribution of T^2 to facilitate the test of the hypothesis.

Distribution of T^2

The distribution of T^2 is derived under general conditions, including the case when H_0 is not true.

Let $T^2 = \underset{\sim}{Y}' S^{-1} \underset{\sim}{Y}$ where $\underset{\sim}{Y} \sim N_p(\underset{\sim}{\gamma}, \Sigma)$ and nS is distributed independently as $\sum_{\alpha=1}^{n} \underset{\sim}{Z}_\alpha \underset{\sim}{Z}_\alpha'$ with $\underset{\sim}{Z}_1, \underset{\sim}{Z}_2, \ldots, \underset{\sim}{Z}_n$ independent, each $\underset{\sim}{Z}_\alpha \sim N(\underset{\sim}{0}, \Sigma)$, $\alpha = 1, 2, \ldots, n$.

The derived T^2 statistic is defined as

$$T^2 = N(\bar{\underset{\sim}{x}} - \underset{\sim}{\mu}_0)' S^{-1}(\bar{\underset{\sim}{x}} - \underset{\sim}{\mu}_0)$$

and is a special case of $T^2 = \underset{\sim}{Y}'S^{-1}\underset{\sim}{Y}$ with

$$\underset{\sim}{Y} = \sqrt{N}(\bar{\underset{\sim}{x}} - \underset{\sim}{\mu_0}), \quad \underset{\sim}{\gamma} = \sqrt{N}(\underset{\sim}{\mu} - \underset{\sim}{\mu_0})$$

and $n = N - 1$.

Let D be a non-singular matrix such that $D\Sigma D' = I$. Next, define

$$\underset{\sim}{Y}^* = D\underset{\sim}{Y}$$

$$\underset{\sim}{S}^* = DSD'$$

$$\underset{\sim}{\gamma}^* = D\underset{\sim}{\gamma}.$$

The following Lemma is used further:

Lemma 2: For any $p \times p$ non-singular matrix C and H and any vector $\underset{\sim}{k}$,

$$\underset{\sim}{k}'H^{-1}\underset{\sim}{k} = (C\underset{\sim}{k})'(CHC')^{-1}(C\underset{\sim}{k}).$$

Proof: The proof of the lemma follows by considering the right-hand side as

$$(C\underset{\sim}{k})'(CHC')^{-1}(C\underset{\sim}{k}) = \underset{\sim}{k}'C'(C')^{-1}H^{-1}C^{-1}C\underset{\sim}{k} = \underset{\sim}{k}'H^{-1}\underset{\sim}{k}.$$

Consider

$$T^2 = \underset{\sim}{Y}'S^{-1}\underset{\sim}{Y}$$
$$= \underset{\sim}{Y}'D'D'^{-1}S^{-1}D^{-1}D\underset{\sim}{Y}$$
$$= (D\underset{\sim}{Y})'(DSD')^{-1}(D\underset{\sim}{Y})$$
$$= \underset{\sim}{Y}^{*'}S^{*-1}\underset{\sim}{Y}^*.$$

Then $T^2 = \underset{\sim}{Y}^{*'}S^{*-1}\underset{\sim}{Y}^*$, where $\underset{\sim}{Y}^* \sim N_p(\gamma^*, I)$, and nS^* is distributed independently as $\sum_{\alpha=1}^{n} \underset{\sim}{Z}_\alpha^* \underset{\sim}{Z}_\alpha^{*'} = \sum_{\alpha=1}^{n}(D\underset{\sim}{Z}_\alpha)(D\underset{\sim}{Z}_\alpha)'$, where $\underset{\sim}{Z}_\alpha^* = D\underset{\sim}{Z}_\alpha \sim N_p(0, I), \quad \alpha = 1, 2, \ldots, n$.

Note that $\underset{\sim}{\gamma}^{*'}\underset{\sim}{\gamma}^* = \underset{\sim}{\gamma}'\Sigma^{-1}\underset{\sim}{\gamma}$.

Orthogonal Transformation:

Define an orthogonal matrix Q of order $p \times p$ such that the first row of Q is

$$q_{1i} = \frac{Y_i^*}{\sqrt{\underset{\sim}{Y}^{*'}\underset{\sim}{Y}^*}}, \quad i = 1, 2, \ldots, p.$$

This is permissible since
$$\sum_{i=1}^{p} q_{1i}^2 = 1$$
due to the property of an orthogonal matrix. Other $(p-1)$ rows can be defined arbitrarily.

This will define an orthogonal matrix due to the following Lemma:

Lemma 3: Let A be an $n \times m (n > m)$ such that $A'A = I_m$, then there exists an $n \times (n-m)$ matrix B such that the matrix $(A \quad B)$ is an orthogonal matrix.

Proof: Since $rank(A) = m$, then there exists a matrix C such that the matrix $(A \quad C)$ is nonsingular. Take a matrix D as $(C - AA'C)$. Then $D'A = (C - AA'C)A = CA - AA'CA = 0$. Let E be a matrix of order $(n-m) \times (n-m)$ such that $E'D'DE = I$. Then B can be taken as DE.

Note that Q depends on $\underset{\sim}{Y}^*$, so it is a random matrix.

Let
$$\underset{\sim}{U} = Q\underset{\sim}{Y}^*, \qquad B = Q\left(nS^*\right)Q'.$$

Then
$$U_1 = \sum_i q_{1i} Y_i^* = \frac{\underset{\sim}{Y}^{*'}\underset{\sim}{Y}^*}{\sqrt{\underset{\sim}{Y}^{*'}\underset{\sim}{Y}^*}} = \sqrt{\underset{\sim}{Y}^{*'}\underset{\sim}{Y}^*},$$
$$U_j = \sum_i q_{ji} Y_i^* = \sqrt{\underset{\sim}{Y}^{*'}\underset{\sim}{Y}^*} \sum_i q_{ji} q_{1i} = 0, \quad j \neq 1 \quad \left(\text{using } Y_i^* = q_{1i}\sqrt{\underset{\sim}{Y}^{*'}\underset{\sim}{Y}^*}\right).$$

Then, we have
$$\frac{T^2}{n} = \frac{1}{n}\underset{\sim}{Y}^{*'}Q'Q^{-1'}S^{-1}Q^{-1}Q\underset{\sim}{Y}^*$$
$$= (\underset{\sim}{Y}^*Q)'(Q\ nS\ Q')^{-1}Q\underset{\sim}{Y}^*$$
$$= \underset{\sim}{U}'B^{-1}\underset{\sim}{U}$$
$$= (U_1, 0, \ldots, 0)\begin{pmatrix} b^{11} & b^{12} & \cdots & b^{1p} \\ b^{21} & b^{22} & \cdots & b^{2p} \\ \vdots & \vdots & \ddots & \vdots \\ b^{p1} & b^{p2} & \cdots & b^{pp} \end{pmatrix}\begin{pmatrix} U_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$
$$= U_1^2\, b^{11},$$

where $B^{-1} = (b^{ij})$ and b^{ij} denotes the $(i,j)^{th}$ element of B^{-1}.

If B is partitioned as

$$B = \begin{pmatrix} b_{11} & \underset{\sim}{b}'_{(1)} \\ \underset{\sim}{b}_{(1)} & B_{22} \end{pmatrix},$$

then

$$\frac{1}{b^{11}} = b_{11} - \underset{\sim}{b}'_{(1)} B_{22}^{-1} \underset{\sim}{b}_{(1)} = b_{11.23...p}.$$

This result is obtained using the following theorem:

Theorem 4: Let

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

be nonsingular with A_{22} be a square and nonsingular matrix. Then

$$A^{-1} = \begin{pmatrix} A_{11.2}^{-1} & -A_{11.2}^{-1} A_{12} A_{22}^{-1} \\ -A_{22}^{-1} A_{21} A_{11.2}^{-1} & A_{22}^{-1} + A_{22}^{-1} A_{21} A_{11.2}^{-1} A_{12} A_{22}^{-1} \end{pmatrix},$$

where

$$A_{11.2} = A_{11} - A_{12} A_{22}^{-1} A_{21}.$$

Also, we use the following theorem to derive the distribution of $b_{11.2...p}$. The distribution of $A_{11.2}$ is $W_q(\Sigma_{11.2}, n - p + q)$ where $A = \sum_{\alpha=1}^{n} \underset{\sim}{Y}_\alpha \underset{\sim}{Y}'_\alpha$, each $\underset{\sim}{Y}_\alpha \sim N_p(\underset{\sim}{0}, \Sigma), \alpha = 1, 2, \ldots, n$ are independently distributed and partitioned as $\underset{\sim}{Y}_\alpha = (\underset{\sim}{Y}_\alpha^{(1)} \quad \underset{\sim}{Y}_\alpha^{(2)})'$ such that $\underset{\sim}{Y}_\alpha^{(1)}$ and $\underset{\sim}{Y}_\alpha^{(2)}$ are of oreders $q \times 1$ and $(p - q) \times 1$, respectively.

It is also clear that $b_{11.23...p}$ is based only on one variable, so using that Wishart distribution reduces to chi-square distribution for one variable case (i.e., $p = 1$), it is clear that $b_{11.23...p} \sim \chi^2_{n-(p-1)}$, i.e., $b_{11.23...p}$ follows a chi-square distribution with $n - (p - 1)$ degrees of freedom. Since the conditional distribution of $b_{11.23...p}$ does not depend on Q, it is unconditionally distributed as a chi-square distribution.

Based on $A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$, we have already proved that $A_{11.2}$ and A_{22} are independently distributed. So any function of $A_{11.2}$ and A_{22} will also be independently distributed.

Thus $\underset{\sim}{Y}^{*\prime}\underset{\sim}{Y}^{*}$ has a non-central χ^2 distribution with p degrees of freedom and non-centrality parameter given by

$$\underset{\sim}{\gamma}^{*\prime}\underset{\sim}{\gamma}^{*} = \underset{\sim}{\gamma}^{\prime}\Sigma^{-1}\underset{\sim}{\gamma}.$$

Thus

$$\frac{T^2}{n} = \frac{\underset{\sim}{Y}^{*\prime}\underset{\sim}{Y}^{*}}{b_{11\cdot23\ldots p}}$$

is distributed as the ratio of a non-central χ^2 and an independent χ^2. So using the distributional property of F-distribution, we can write:

$$\frac{T^2}{n} = \left(\frac{\underset{\sim}{Y}^{*\prime}\underset{\sim}{Y}^{*}/p}{b_{11\cdot23\ldots p}/(n-p+1)}\right) \times \left(\frac{p}{n-p+1}\right)$$

where the first term follows a non-central F distribution with p and $(n-p+1)$ degrees of freedom and non-centrality parameter $\underset{\sim}{\gamma}^{\prime}\Sigma^{-1}\underset{\sim}{\gamma}$.

Thus

$$\frac{n-p+1}{p} \cdot \frac{T^2}{n} \sim \text{Noncentral } F(p, n-p+1, \delta^2)$$

Substituting $n = N-1$, we get

$$\frac{N-p}{p} \cdot \frac{T^2}{N-1} \sim \text{Noncentral } F(p, N-p, \underset{\sim}{\gamma}^{\prime}\Sigma^{-1}\underset{\sim}{\gamma}).$$

We call this the T^2 distribution with n degrees of freedom.

If $\underset{\sim}{\mu} = \underset{\sim}{\mu_0}$, then the distribution is central F distribution.

The exact form of the probability density of T^2 can be written using the probability density function of the non-central F-distribution.

Limiting Behavior of T^2 Statistic:

The following theorem states the limiting behaviour of T^2 statistics.

Theorem 5: Let $\underset{\sim}{X}_1, \underset{\sim}{X}_2, \ldots$ be a sequence of identically and independently distributed random vectors with mean vector $\underset{\sim}{\mu}$ and covariance matrix Σ. Let $\bar{\underset{\sim}{X}}_N = \frac{1}{N}\sum_{\alpha=1}^{N} \underset{\sim}{X}_\alpha$ and $S_N = \frac{1}{N-1}\sum_{\alpha=1}^{N}(\underset{\sim}{X}_\alpha - \bar{\underset{\sim}{X}}_N)(\underset{\sim}{X}_\alpha - \bar{\underset{\sim}{X}}_N)'$, and

$$T_N^2 = N(\bar{\underset{\sim}{X}}_N - \underset{\sim}{\mu_0})'S_N^{-1}(\bar{\underset{\sim}{X}}_N - \underset{\sim}{\mu_0})$$

Then the limiting distribution of T_N^2 as $N \to \infty$ is the χ^2 distribution with p degrees of freedom when $\underset{\sim}{\mu} = \underset{\sim}{\mu}_0$.

Uses of T^2 Statistic

1. One Sample Mean Problem

Let $\underset{\sim}{X}_1, \underset{\sim}{X}_2, \ldots, \underset{\sim}{X}_N$ be independently distributed and $\underset{\sim}{X}_\alpha \sim N_p(\underset{\sim}{\mu}, \Sigma)$, $\alpha = 1, 2, \ldots, N$. Suppose $H_0 : \underset{\sim}{\mu} = \underset{\sim}{\mu}_0$ is to be tested.

The likelihood ratio test statistic has already been derived, and the critical region is obtained as:

Reject H_0 if $\lambda \leq \lambda_0$

where

$$\lambda^{2/N} = \frac{1}{1 + \frac{T^2}{N-1}}.$$

So the rejection region $\lambda \leq \lambda_0$ is equivalent to

$$\lambda^{2/N} \leq \lambda_0^{2/N}$$

$$\text{or} \quad \lambda^{N/2} \leq \lambda_0^{N/2}$$

$$\text{or} \quad \lambda^{-\frac{2}{N}} - 1 \geq \lambda_0^{-\frac{2}{N}} - 1$$

$$\text{or} \quad (N-1)(\lambda^{-\frac{2}{N}} - 1) \geq (N-1)(\lambda_0^{-\frac{2}{N}} - 1)$$

$$\text{or} \quad (N-1)\left(1 + \frac{T^2}{N-1} - 1\right) \geq (N-1)\left(1 + \frac{T_0^2}{N-1} - 1\right)$$

$$\text{or} \quad T^2 \geq T_0^2(\alpha),$$

where $T_0^2 = (N-1)(\lambda_0^{-2/N} - 1)$ is a constant to be determined by the size condition

$$P(\lambda \leq \lambda_0 | H_0) = \alpha.$$

This is equivalent to saying that if the level of significance is α, then the $100\alpha\%$ point of the F-distribution is taken, i.e.,

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

or

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

$$= T_{p,N-1}^2(\alpha)$$

$$= T_{p,n}^2(\alpha).$$

An important point to be noticed here is that in order to find the value of T_0^2 at α level of significance, first find the value of $F(p, N-p)$ at α level of significance and then find $T_{p,N-1}^2(\alpha)$.

Observe that while computing the T^2 statistic using $\bar{X}$ and A, the vector $A^{-1}(\bar{X}-\mu_0)$ is the solution b of $Ab = \bar{X} - \mu_0$. Then

$$\frac{T^2}{N-1} = N(\bar{X} - \mu_0)'A^{-1}(\bar{X} - \mu_0) = N(\bar{X} - \mu_0)'b.$$

Note that $\frac{T^2}{N-1}$ is the non-zero characteristic root of the equation:

$$|N(\bar{X} - \mu_0)(\bar{X} - \mu_0)' - \lambda A| = 0.$$

This follows from the following lemma.

Lemma 6: If B is a $p \times p$ non-singular matrix and v is a $p \times 1$ vector, then $v'B^{-1}v$ is the non-zero root of $|vv' - \lambda B| = 0$.

Proof: Let λ_1 be a nonzero root of

$$|vv' - \lambda B| = 0$$

with a characteristic vector β satisfying

$$vv'\beta = \lambda B\beta.$$

Since $\lambda_1 \neq 0$, $v'\beta \neq 0$. Hence, multiplying both sides by $v'B^{-1}$, we get

$$(v'B^{-1}v)(v'\beta) = \lambda(v'\beta).$$

In the case to show that λ is the nonzero root of

$$\left| N(x - \mu_0)(x - \mu_0)' - \lambda A \right| = 0,$$

choose

$$\underset{\sim}{v} = \sqrt{N}(\underset{\sim}{x} - \mu_0) \quad \text{and} \quad B = A$$

in the lemma.

Confidence Region for the Mean Vector

If $\underset{\sim}{\mu}$ is the mean vector of $N_p(\mu, \Sigma)$, the probability is $(1 - \alpha)$ of drawing a sample of size N with mean $\bar{x}$ and covariance matrix S such that

$$N(\bar{\underset{\sim}{x}} - \underset{\sim}{\mu})' S^{-1} (\bar{\underset{\sim}{x}} - \underset{\sim}{\mu}) \le T^2_{p, N-1}(\alpha).$$

We can also obtain the confidence intervals for linear functions $\underset{\sim}{\delta}' \underset{\sim}{\mu}$ that hold simultaneously with a given confidence coefficient $(1 - \alpha)$ using the following generalized Cauchy–Schwarz inequality.

Lemma 7: For a positive definite matrix S,

$$(\underset{\sim}{\delta}' \underset{\sim}{y})^2 \le (\underset{\sim}{\delta}' S \underset{\sim}{\delta})\,(\underset{\sim}{y}' S^{-1} \underset{\sim}{y}).$$

Proof: Let

$$d = \frac{\underset{\sim}{\delta}' \underset{\sim}{y}}{\underset{\sim}{\delta}' S \underset{\sim}{\delta}}.$$

Then,

$$0 \le (\underset{\sim}{y} - dS\underset{\sim}{\delta})' S^{-1} (\underset{\sim}{y} - dS\underset{\sim}{\delta}).$$
$$= \underset{\sim}{y}' S^{-1} \underset{\sim}{y} - d\,\underset{\sim}{\delta}' S S^{-1} \underset{\sim}{y} - \underset{\sim}{y}' S^{-1} S \underset{\sim}{\delta} d + d^2 \underset{\sim}{\delta}' S S^{-1} S \underset{\sim}{\delta}$$
$$= \underset{\sim}{y}' S^{-1} \underset{\sim}{y} - \frac{(\underset{\sim}{\delta}' \underset{\sim}{y})^2}{\underset{\sim}{\delta}' S \underset{\sim}{\delta}},$$

which gives

$$(\underset{\sim}{\delta}' \underset{\sim}{y})^2 \le (\underset{\sim}{\delta}' S \underset{\sim}{\delta})\,(\underset{\sim}{y}' S^{-1} \underset{\sim}{y}).$$

When $\underset{\sim}{y} = \underset{\sim}{x} - \underset{\sim}{\mu}$, the generalized Cauchy–Schwarz inequality implies

$$|\underset{\sim}{\delta}'(\underset{\sim}{x} - \mu)| \le \sqrt{\underset{\sim}{\delta}' S \underset{\sim}{\delta}\,(\underset{\sim}{x} - \mu)' S^{-1}(\underset{\sim}{x} - \mu)}$$
$$\le \sqrt{\underset{\sim}{\gamma}' S \underset{\sim}{\gamma}}\,\sqrt{\frac{T^2_{p, N-1}(\alpha)}{N}}.$$

This holds for all $\underset{\sim}{\delta}$ with probability $(1 - \alpha)$.

Thus, we can state with confidence coefficient $(1 - \alpha)$ that the unknown parameter vector satisfies simultaneously for all $\underset{\sim}{\delta}$ the inequalities

$$\left| \underset{\sim}{\delta}' \underset{\sim}{x} - \underset{\sim}{\delta}' \underset{\sim}{\mu} \right| \leq \sqrt{\underset{\sim}{\gamma}' S \underset{\sim}{\gamma}} \ \sqrt{\frac{T^2_{p,\, N-1}(\alpha)}{N}}.$$

2. Two-Sample Problem

In a univariate setup, the two-sample problem is considered as the testing of the null hypothesis

$$H_0 : \mu_1 = \mu_2$$

based on two samples drawn independently from two normal populations with different means but the same variance.

Now, we want to consider the same problem in a multivariate setup and test the hypothesis of equality of two mean vectors.

Suppose the two samples of sizes N_1 and N_2 are drawn from $N_p(\underset{\sim}{\mu}^{(1)}, \Sigma)$ and $N_p(\underset{\sim}{\mu}^{(2)}, \Sigma)$, as

$$\underset{\sim}{x}^{(1)}_1, \underset{\sim}{x}^{(1)}_2, \ldots, \underset{\sim}{x}^{(1)}_{N_1} \sim N_p(\underset{\sim}{\mu}^{(1)}, \Sigma)$$

and

$$\underset{\sim}{x}^{(2)}_1, \underset{\sim}{x}^{(2)}_2, \ldots, \underset{\sim}{x}^{(2)}_{N_2} \sim N_p(\underset{\sim}{\mu}^{(2)}, \Sigma).$$

Note that the covariance matrix Σ is the same in both populations. This can be different as Σ_1 and Σ_2. We will consider this separately later as the Fisher–Behrens problem.

At present, we assume

$$\Sigma_1 = \Sigma_2 = \Sigma.$$

Next, find the sample mean vectors $\bar{\underset{\sim}{x}}^{(1)}$ and $\bar{\underset{\sim}{x}}^{(2)}$ from the two samples and find the sample covariance matrices S_1 and S_2 from the two samples. Then, find the pooled

covariance matrix as

$$S = \frac{(N_1 - 1)S_1 + (N_2 - 1)S_2}{N_1 + N_2 - 2} = \frac{n_1 S_1 + n_2 S_2}{n_1 + n_2}.$$

where

$$\bar{x}^{(1)} = \frac{1}{N_1} \sum_{\alpha=1}^{N_1} x_\alpha^{(1)}, \qquad \bar{x}^{(2)} = \frac{1}{N_2} \sum_{\alpha=1}^{N_2} x_\alpha^{(2)},$$

$$S_1 = \frac{1}{N_1 - 1} \sum_{\alpha=1}^{N_1} \left(x_\alpha^{(1)} - \bar{x}^{(1)} \right) \left(x_\alpha^{(1)} - \bar{x}^{(1)} \right)',$$

$$S_2 = \frac{1}{N_2 - 1} \sum_{\alpha=1}^{N_2} \left(x_\alpha^{(2)} - \bar{x}^{(2)} \right) \left(x_\alpha^{(2)} - \bar{x}^{(2)} \right)'.$$

Recalling the definition of the T^2 statistic, we construct the T^2 statistic for

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

Such a null hypothesis can be written as

$$H_0 : \mu^{(1)} - \mu^{(2)} = 0, \quad \text{or} \quad H_0 : \mu^* = 0.$$

The null hypothesis $H_0 : \mu^* = 0$ where $\mu^* = \mu^{(1)} - \mu^{(2)}$ is equivalent to test the null hypothesis in a one sample test case. So we need to find a statistic for testing $H_0 : \mu^* = 0$ based on T^2 - distribution. To find such a statistic, we require a random variable which follows $N_p(\mu^*, \Sigma)$. Since $\bar{X}^{(1)} \sim N_p(\mu^{(1)}, \frac{\Sigma}{N_1})$ and $\bar{X}^{(2)} \sim N_p(\mu^{(2)}, \frac{\Sigma}{N_2})$, so intuitively $\bar{X}^{(1)} - \bar{X}^{(2)}$ can be considered as an initial estimator of $\mu^{(1)} - \mu^{(2)}$ and can be modified to have covariance matrix as Σ. Note that both the samples are independently drawn, so $Cov(X_{(\alpha)}^{(1)}, X_{(\beta)}^{(2)}) = 0$ for $\alpha \neq \beta = 1, 2, \ldots, N$.

We observe that

$$E(\bar{X}^{(1)} - \bar{X}^{(2)}) = \mu^{(1)} - \mu^{(2)}$$

so $\bar{X}^{(1)} - \bar{X}^{(2)}$ is an unbiased estimator of $\mu^{(1)} - \mu^{(2)}$.

Further,

$$Var(\bar{X}^{(1)} - \bar{X}^{(2)}) = \frac{\Sigma}{N_1} + \frac{\Sigma}{N_2}$$
$$= \frac{N_1 + N_2}{N_1 N_2}\Sigma$$

or

$$Var\left[\sqrt{\frac{N_1 N_2}{N_1 + N_2}}(\bar{X}^{(1)} - \bar{X}^{(2)})\right] = \Sigma.$$

Thus, the construction of a test statistic for $H_0 : \mu^{(1)} = \mu^{(2)}$ can be restated as follows:

Let $x_\alpha^{(i)}$ be a random sample from $N_p(\mu^{(i)}, \Sigma)$, $\alpha = 1, 2, \ldots, N_i$, $i = 1, 2$. Then

$$\sqrt{\frac{N_1 N_2}{N_1 + N_2}}(\bar{X}^{(1)} - \bar{X}^{(2)}) \sim N_p(0, \Sigma) \text{ under } H_0.$$

If we let

$$S = \frac{1}{N_1 + N_2 - 2}\left[\sum_{\alpha=1}^{N_1}(X_\alpha^{(1)} - \bar{X}^{(1)})(X_\alpha^{(1)} - \bar{X}^{(1)})' + \sum_{\alpha=1}^{N_2}(X_\alpha^{(2)} - \bar{X}^{(2)})(X_\alpha^{(2)} - \bar{X}^{(2)})'\right]$$

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

Thus

$$T^2 = \frac{N_1 N_2}{N_1 + N_2}(\bar{x}^{(1)} - \bar{x}^{(2)})'S^{-1}(\bar{x}^{(1)} - \bar{x}^{(2)})$$

is distributed as T^2 with $(N_1 + N_2 - 2)$ degrees of freedom under H_0. Equivalently, under H_0,

$$\frac{N_1 + N_2 - p - 1}{p} \cdot \frac{T^2}{N_1 + N_2 - 2} \sim F(p, N_1 + N_2 - p - 1; \alpha).$$

The critical region is given as

$$T^2 > \frac{(N_1 + N_2 - 2)p}{N_1 + N_2 - p - 1}F(p, N_1 + N_2 - p - 1; \alpha)$$

with significance level α.

The confidence region for $\mu^{(1)} - \mu^{(2)}$ with confidence level $(1 - \alpha)$ is the set of vectors μ^* satisfying

$$(\bar{x}^{(1)} - \bar{x}^{(2)} - \mu^*)'S^{-1}(\bar{x}^{(1)} - \bar{x}^{(2)} - \mu^*) \leq \frac{N_1 + N_2}{N_1 N_2}T^2_{p,N_1+N_2-2}(\alpha)$$

or

$$(\bar{\underset{\sim}{x}}^{(1)} - \bar{\underset{\sim}{x}}^{(2)} - \underset{\sim}{\mu}^*)' S^{-1} (\bar{\underset{\sim}{x}}^{(1)} - \bar{\underset{\sim}{x}}^{(2)} - \underset{\sim}{\mu}^*) \leq \frac{N_1 + N_2}{N_1 N_2} \cdot \frac{(N_1 + N_2 - 2)p}{(N_1 + N_2 - p - 1)} F(p, N_1 + N_2 - p - 1; \alpha).$$

The simultaneous confidence Interval for linear functions $\underset{\sim}{\delta}'(\mu^{(1)} - \mu^{(2)})$, on the lines of setup discussed in the case of a one-sample problem, are given by

$$|\underset{\sim}{\delta}'(\bar{\underset{\sim}{x}}^{(1)} - \bar{\underset{\sim}{x}}^{(2)}) - \underset{\sim}{\delta}'\underset{\sim}{\mu}^*| \leq \sqrt{\underset{\sim}{\delta}' S \underset{\sim}{\delta}} \sqrt{\frac{N_1 + N_2}{N_1 N_2} T^2_{p, N_1 + N_2 - 2}(\alpha)} \ .$$

3. Problem of Symmetry and Invariance Property

Consider the test of the hypothesis $H_0 : \underset{\sim}{\mu_1} = \underset{\sim}{\mu_2} = \cdots = \underset{\sim}{\mu_p}$ on the basis of a random sample $\underset{\sim}{x}_1, \underset{\sim}{x}_2, \ldots, \underset{\sim}{x}_N$ of size N from $N_p(\underset{\sim}{\mu}, \Sigma)$ where $\underset{\sim}{\mu} = (\mu_1, \mu_2, \ldots, \mu_p)'$.

The null hypothesis $H_0 : \mu_1 = \mu_2 = \cdots = \mu_p$ has p parameters. It can be expressed as $H_0 : \mu_1 - \mu_2 = 0, \underset{\sim}{\mu_2} - \mu_3 = 0, \ldots, \mu_{p-1} - \mu_p = 0$ which has $(p - 1)$ simple linear contrasts and can be represented by considering a $(p-1) \times p$ matrix C with rank $(p-1)$ such that

$$C\underset{\sim}{\varepsilon} = \underset{\sim}{0}$$

where $\underset{\sim}{\varepsilon}' = (1, 1, \ldots, 1)$. Then we can define $\underset{\sim}{y}_\alpha = C\underset{\sim}{x}_\alpha$ and $\underset{\sim}{y}_\alpha \sim N_{p-1}(C\underset{\sim}{\mu}, C\Sigma C')$.

Now the null hypothesis can be re-expressed as

$$H_0 : C\underset{\sim}{\mu} = \underset{\sim}{0}.$$

The statistic to be used for $H_0 : C\underset{\sim}{\mu} = \underset{\sim}{0}$ is

$$T_y^2 = N\bar{\underset{\sim}{y}}' S^{-1} \bar{\underset{\sim}{y}}$$

where

$$\bar{\underset{\sim}{y}} = \frac{1}{N} \sum_{\alpha=1}^{N} \underset{\sim}{y}_\alpha = C\bar{\underset{\sim}{x}},$$

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

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

The statistic T_y^2 has a distribution with $(N-1)$ degrees of freedom for a $(p-1)$ dimensional distribution.

This T^2 statistic is invariant under any linear transformation in the $(p-1)$ dimensions orthogonal to $\underset{\sim}{\varepsilon}$. This can be seen as follows:

$$
\begin{aligned}
T_y^2 &= N\bar{\underset{\sim}{y}}' S^{-1} \bar{\underset{\sim}{y}} \\
&= N(C\bar{\underset{\sim}{x}})'(CSC')^{-1}(C\bar{\underset{\sim}{x}}) \\
&= N\bar{\underset{\sim}{x}}' C' C'^{-1} S^{-1} C^{-1} C\bar{\underset{\sim}{x}} \\
&= N\bar{\underset{\sim}{x}}' S^{-1} \bar{\underset{\sim}{x}} \\
&= T_x^2.
\end{aligned}
$$

So T_y^2 is independent of C.

Such hypothesis $H_0 : C\underset{\sim}{\mu} = \underset{\sim}{0}$ can be used in various situations. For example

$$
H_0 : \underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_2 + \underset{\sim}{\mu}_3 = \underset{\sim}{\mu}_4
$$

can be expressed as $H_0 : C\underset{\sim}{\mu} = \underset{\sim}{0}$ where $C = (1, \; -1, \; 1, \; -1)$ and $\underset{\sim}{\mu} = (\mu_1, \mu_2, \mu_3, \mu_4)'$.

Mahalanobis Distance

Define

$$
\Delta^2 = (\underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_2)' \Sigma^{-1} (\underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_2)
$$

which gives the distance between two populations $N_p(\underset{\sim}{\mu}_1, \Sigma)$ and $N_p(\underset{\sim}{\mu}_2, \Sigma)$. This is known as the Mahalanobis distance.

The statistical distance or Mahalanobis distance between two points $\underset{\sim}{x} = (x_1, x_2, \ldots, x_p)'$ and $\underset{\sim}{y} = (y_1, y_2, \ldots, y_p)'$ in the p-dimensional space $\mathbb{R}^p$ is defined as

$$
d_S(\underset{\sim}{x}, \underset{\sim}{y}) = \sqrt{(\underset{\sim}{x} - \underset{\sim}{y})' S^{-1} (\underset{\sim}{x} - \underset{\sim}{y})}
$$

which satisfies the following properties:

(i) $d_S(\underset{\sim}{x}, \underset{\sim}{y}) = d_S(\underset{\sim}{y}, \underset{\sim}{x})$

(ii) $d_S(\underset{\sim}{x}, \underset{\sim}{y}) > 0$ if $\underset{\sim}{x} \neq \underset{\sim}{y}$

(iii) $d_S(\underset{\sim}{x}, \underset{\sim}{y}) = 0$ if $\underset{\sim}{x} = \underset{\sim}{y}$

(iv) $d_S(\underset{\sim}{x}, \underset{\sim}{y}) \leq d_S(\underset{\sim}{x}, \underset{\sim}{z}) + d_S(\underset{\sim}{z}, \underset{\sim}{y})$.

The condition (iv) is known as the Triangle inequality.

The Mahalanobis distance is the squared distance between two points in multivariate space taking into account the correlation structure through the covariance matrix.

4. Testing the distance between Two Populations

The null hypothesis $H_0 : \Delta^2 = 0$ tests that the distance between the two multivariate populations is zero, meaning that both populations are the same.

Since $\Delta^2 = (\mu_1 - \mu_2)'\Sigma^{-1}(\mu_1 - \mu_2)$, so $\Delta^2 = 0$ when $\underset{\sim}{\mu_1} = \underset{\sim}{\mu_2}$. Thus $\underset{\sim}{\mu_1} = \underset{\sim}{\mu_2}$ and $\Delta^2 = 0$ are equivalent. The same test statistic and test procedure are used to test $H_0 : \underset{\sim}{\mu_1} = \underset{\sim}{\mu_2}$, also used to test $H_0 : \Delta^2 = 0$.

Next, we want to discuss Rao's U-statistic. Before discussing this statistic, we discuss the following result, which is used in the discussion of Rao's U-statistics.

Result: Consider a $p \times 1$ vector $\underset{\sim}{Y}$ and partition $\underset{\sim}{Y} = (\underset{\sim}{Y_1'} \quad \underset{\sim}{Y_2'})'$ where $\underset{\sim}{Y_1}$ is $q \times 1$ vector and $\underset{\sim}{Y_2}$ is $(p - q) \times 1$ vector. Similarly μ, Σ and A are partitioned as:

$$\underset{\sim}{\mu} = (\underset{\sim}{\mu_1'} \quad \underset{\sim}{\mu_2'})', \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \quad A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}.$$

Consider a quantity

$$\Delta_p^2 = \underset{\sim}{\mu}'\Sigma^{-1}\underset{\sim}{\mu}$$

which is analogous to Hotelling's T^2 in the population.

To find an expression of $\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}^{-1}$, we use the following result.

Suppose

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} P & Q \\ R & S \end{pmatrix}$$

then

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} P & Q \\ R & S \end{pmatrix} = \begin{pmatrix} I & O \\ O & I \end{pmatrix}$$

or

(i) $AP + BR = I$

(ii) $CQ + DS = I$

(iii) $AQ + BS = O$

(iv) $CP + DR = O$

Solving (i)–(iv) with the result $(I + LM)^{-1} = I - L(I + ML)^{-1}M$ gives the following solution:

$$P = (A - BD^{-1}C)^{-1} = A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1}$$
$$Q = -A^{-1}B(D - CA^{-1}B)^{-1}$$
$$R = -D^{-1}C(A - BD^{-1}C)^{-1}$$
$$S = (D - CA^{-1}B)^{-1}.$$

Thus, we can write:

$$\Sigma^{-1} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}^{-1} = \begin{pmatrix} \Sigma_{11}^{-1} + B'\Sigma_{22\cdot1}^{-1}B & -B'\Sigma_{22\cdot1}^{-1} \\ -\Sigma_{22\cdot1}^{-1}B & \Sigma_{22\cdot1}^{-1} \end{pmatrix}$$

where $B = \Sigma_{21}\Sigma_{11}^{-1}$.

Similarly,

$$A^{-1} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}^{-1} = \begin{pmatrix} A_{11}^{-1} + \hat{B}'A_{22\cdot1}^{-1}\hat{B} & -\hat{B}'A_{22\cdot1}^{-1} \\ -A_{22\cdot1}^{-1}\hat{B} & A_{22\cdot1}^{-1} \end{pmatrix}$$

where $\hat{B} = A_{21}A_{11}^{-1}$.

Consider $\Delta_p^2 = \underset{\sim}{\mu}'\Sigma^{-1}\underset{\sim}{\mu}$ and substitute the partitioned vector and matrix as follows:

$$\Delta_p^2 = \underset{\sim}{\mu}'\Sigma^{-1}\underset{\sim}{\mu}$$

$$= (\mu_1' \quad \mu_2') \begin{pmatrix} \Sigma_{11}^{-1} + B'\Sigma_{22\cdot1}^{-1}B & -B'\Sigma_{22\cdot1}^{-1} \\ -\Sigma_{22\cdot1}^{-1}B & \Sigma_{22\cdot1}^{-1} \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$$

$$= \underset{\sim}{\mu_1'}\Sigma_{11}^{-1}\underset{\sim}{\mu_1} + \underset{\sim}{\mu_1'}B'\Sigma_{22\cdot1}^{-1}B\underset{\sim}{\mu_1} - \underset{\sim}{\mu_1'}B'\Sigma_{22\cdot1}^{-1}\underset{\sim}{\mu_2} - \underset{\sim}{\mu_2'}\Sigma_{22\cdot1}^{-1}B\underset{\sim}{\mu_1} + \underset{\sim}{\mu_2'}\Sigma_{22\cdot1}^{-1}\underset{\sim}{\mu_2}$$

$$= \underset{\sim}{\mu_1'}\Sigma_{11}^{-1}\underset{\sim}{\mu_1} + (\underset{\sim}{\mu_2} - B\underset{\sim}{\mu_1})'\Sigma_{22\cdot1}^{-1}(\underset{\sim}{\mu_2} - B\underset{\sim}{\mu_1})$$

$$\Delta_p^2 = \Delta_q^2 + (\underset{\sim}{\mu_2} - B\underset{\sim}{\mu_1})'\Sigma_{22\cdot1}^{-1}(\underset{\sim}{\mu_2} - B\underset{\sim}{\mu_1}).$$

Here Δ_p^2 and Δ_q^2 are the Mahalanobis distances based on populations with p and q random variables, respectively.

Rao's U-Statistic

Let $\underset{\sim}{Y} \sim N_p(\underset{\sim}{\mu}, \Sigma)$ and $A \sim W_p(\Sigma, n)$. The distributions of $\underset{\sim}{Y}$ and A are independent.

Now partition $\underset{\sim}{Y} = (\underset{\sim}{Y_1'} \quad \underset{\sim}{Y_2'})'$ where $\underset{\sim}{Y_1}$ and $\underset{\sim}{Y_2}$ are $(q \times 1)$ and $(p - q) \times 1$ vectors. The vectors and matrices $\underset{\sim}{\mu}$, Σ and A are accordingly partitioned as

$$\underset{\sim}{\mu} = (\mu_1 \quad \mu_2), \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \quad A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}.$$

Consider $T_p^2 = \underset{\sim}{y}'S^{-1}\underset{\sim}{y}$ where $\underset{\sim}{y} = \sqrt{N}(\bar{\underset{\sim}{x}} - \underset{\sim}{\mu_0})$ will provide the usual T^2-statistics.

$$T_p^2 = (\underset{\sim}{y_1'} \quad \underset{\sim}{y_2'}) \begin{pmatrix} S_{11} & S_{12} \\ S_{22} & S_{22} \end{pmatrix}^{-1} \begin{pmatrix} y_1 \\ y_2. \end{pmatrix}$$

Using the earlier steps followed in the result, we can express

$$T_p^2 = T_q^2 + (N - 1)(\underset{\sim}{y_2} - \hat{B}\underset{\sim}{y_1})'A_{22\cdot1}^{-1}(\underset{\sim}{y_2} - \hat{B}\underset{\sim}{y_1})$$

or

$$\frac{T_p^2}{N - 1} = \frac{T_q^2}{N - 1} + \underset{\sim}{z}'A_{22\cdot1}^{-1}\underset{\sim}{z}$$

or

$$\frac{T_p^2}{n} = \frac{T_q^2}{n} + \underset{\sim}{z}'A_{22\cdot1}^{-1}\underset{\sim}{z}$$

where $n = N - 1$, $A_{22 \cdot 1} = A_{22} - A_{21}A_{11}^{-1}A_{12}$, $\hat{B} = A_{21}A_{11}^{-1}$, and $\underset{\sim}{z} = \underset{\sim}{y}_2 - \hat{B}\underset{\sim}{y}_1$.

Rao defined

$$U = \frac{1 + \frac{T_q^2}{n}}{1 + \frac{T_p^2}{n}}$$

which is known as Rao's U-statistic.

Rao has further shown that

$$\frac{n - p - 1}{p - q}\left(\frac{1}{U} - 1\right) \sim F(p - q, n - p + 1)$$

if $\Delta_p^2 = \Delta_q^2$.

Note that Δ_p^2 and Δ_q^2 are the Mahalanobis distances between two populations, considering all the p variables and the first q variables, respectively.

Suppose $H_0 : \Delta_p^2 = \Delta_q^2$, then accepting H_0 implies that the distance based on p variables is the same as that distance based on the first q variables,

i.e., $(p - q)$ variables are useless,

i.e., first q variables are equally informative.

Such $(p - q)$ variables are called redundant.

Further, $\Delta_p^2 = \Delta_q^2$ whenever any of the following conditions is satisfied:

(i) $\underset{\sim}{\mu}_2 = B\underset{\sim}{\mu}_1$

(ii) If $\underset{\sim}{\mu}_2 = \underset{\sim}{0}$ and $\underset{\sim}{X}_1$ and $\underset{\sim}{X}_2$ are independently distributed which implies $\Sigma_{12} = 0$, $\Sigma_{21} = 0$, $B = 0$.

(iii) If $\underset{\sim}{\mu}_2 = \underset{\sim}{0}$ given $\underset{\sim}{\mu}_1 = \underset{\sim}{0}$.

(iv) If we have a linear function of $\underset{\sim}{x}$, say $\underset{\sim}{a}'\underset{\sim}{x}$, uncorrelated with $\underset{\sim}{x}_1$. This condition can be seen as follows:

$$\underset{\sim}{a}'\underset{\sim}{x} = \underset{\sim}{a}_1'\underset{\sim}{x}_1 + \underset{\sim}{a}_2'\underset{\sim}{x}_2 \quad \text{if } \underset{\sim}{a} = \begin{pmatrix} \underset{\sim}{a}_1 \\ \underset{\sim}{a}_2 \end{pmatrix}.$$

Since $\underset{\sim}{a}'\underset{\sim}{x}$ is uncorrelated with $\underset{\sim}{x}_1$,

$$Cov(\underset{\sim}{a}'\underset{\sim}{x}, \underset{\sim}{x}_1) = \underset{\sim}{0}$$

or $\quad Cov(\underset{\sim}{a}_1'\underset{\sim}{x}_1 + \underset{\sim}{a}_2'\underset{\sim}{x}_2, \underset{\sim}{x}_1) = \underset{\sim}{0}$

or $\quad \underset{\sim}{a}_1'\Sigma_{11} + \underset{\sim}{a}_2'\Sigma_{21} = \underset{\sim}{0}$

or $\quad \underset{\sim}{a}_1' = -\underset{\sim}{a}_2'\Sigma_{21}\Sigma_{11}^{-1}$

or $\quad \underset{\sim}{a}_1' = -\underset{\sim}{a}_2'B,$

and therefore

$$E(\underset{\sim}{a}'\underset{\sim}{x}) = \underset{\sim}{a}_1'\mu_1 + \underset{\sim}{a}_2'\mu_2$$
$$= \underset{\sim}{a}_2'(\mu_2 - B\mu_1)$$

Thus if $E(\underset{\sim}{a}'\underset{\sim}{x}) = 0$ for all $\underset{\sim}{a}$, this implies $\underset{\sim}{\mu}_2 - B\underset{\sim}{\mu}_1 = \underset{\sim}{0}$, or equivalently $\Delta_p^2 = \Delta_q^2$.

This means that if the mean of a linear combination is 0 and the linear combination is uncorrelated with the first partition, then $\Delta_p^2 = \Delta_q^2$.

The reason behind the conditions (i)–(iv) is as follows:

Note that:

$$\Delta_p^2 = \Delta_q^2 + (\mu_2 - B\mu_1)'\Sigma_{22.1}^{-1}(\mu_2 - B\mu_1)$$

Thus $\Delta_p^2 = \Delta_q^2$ when

$$(\underset{\sim}{\mu}_2 - B\underset{\sim}{\mu}_1)'\Sigma_{22.1}^{-1}(\underset{\sim}{\mu}_2 - B\underset{\sim}{\mu}_1) = 0$$

which is possible if

(i) $\underset{\sim}{\mu}_2 = B\underset{\sim}{\mu}_1$,

or

(ii) $\Sigma_{22.1} = 0$ (which implies deterministic relationship as this is possible when $\Sigma_{21} = 0$ or $B = 0$),

or

(iii) $\mu_2 = \underset{\sim}{0}$ knowing $\mu_1 = \underset{\sim}{0}$.

To use any of (i), (ii), (iii), and (iv), the following F-statistic is used:

$$\frac{n-p+1}{p-q}\left(\frac{1}{U}-1\right) \sim F(p-q, n-p+1)$$

The Multivariate Behrens-Fisher Problem

Now we consider the test of the hypothesis $H_0 : \mu_1 = \mu_2$ when the samples are originating from two independent multivariate normal distributions with **unequal** and **unknown** covariance matrices.

It is important to recall that in the earlier test of hypothesis for $H_0 : \mu_1 = \mu_2$ (where $\Sigma_1 = \Sigma_2$), we considered constructing a statistic which is an unbiased estimator of $(\mu_1 - \mu_2)$ and has the covariance matrix Σ. This facilitated the use of the Wishart distribution to construct a relevant T^2 statistic. In simple terms, the test statistic was constructed with the Wishart distribution in mind. Now we see more clearly the differences that arise when the covariance matrices are unequal.

Consider two samples of sizes N_1 and N_2 drawn from two independent multivariate normal populations as:

$$\underset{\sim}{x}_1^{(1)}, \underset{\sim}{x}_2^{(1)}, \ldots, \underset{\sim}{x}_{N_1}^{(1)} \sim N_p(\underset{\sim}{\mu}^{(1)}, \Sigma_1)$$

and

$$\underset{\sim}{x}_1^{(2)}, \underset{\sim}{x}_2^{(2)}, \ldots, \underset{\sim}{x}_{N_2}^{(2)} \sim N_p(\underset{\sim}{\mu}^{(2)}, \Sigma_2)$$

where Σ_1 and Σ_2 both are unknown and unequal.

The sample means from two respective populations are $\bar{\underset{\sim}{x}}^{(1)}$ and $\bar{\underset{\sim}{x}}^{(2)}$. The sample covariance matrices from the two respective populations are S_1 and S_2.

The technique of considering $(\bar{\underset{\sim}{x}}^{(1)} - \bar{\underset{\sim}{x}}^{(2)})$ as a statistic as in the case of $\Sigma_1 = \Sigma_2$ cannot be considered when $\Sigma_1 \neq \Sigma_2$.

$$(\bar{\underset{\sim}{X}}^{(1)} - \bar{\underset{\sim}{X}}^{(2)}) \sim N\left(\underset{\sim}{\mu}^{(1)} - \underset{\sim}{\mu}^{(2)}, \frac{\Sigma_1}{N_1} + \frac{\Sigma_2}{N_2}\right)$$

and the sample covariance matrix

$$\sum_{\alpha=1}^{N_1}(\underset{\sim}{X}_\alpha^{(1)} - \underset{\sim}{\bar{X}}^{(1)})(\underset{\sim}{X}_\alpha^{(1)} - \underset{\sim}{\bar{X}}^{(1)})' + \sum_{\alpha=1}^{N_2}(\underset{\sim}{X}_\alpha^{(2)} - \underset{\sim}{\bar{X}}^{(2)})(\underset{\sim}{X}_\alpha^{(2)} - \underset{\sim}{\bar{X}}^{(2)})'$$

does not have Wishart distribution with covariance matrix as a multiple of $\left(\frac{\Sigma_1}{N_1} + \frac{\Sigma_2}{N_2}\right)$.
Recall the result that if $A_i \sim W_p(\Sigma, n_i), i = 1, 2, \ldots, q$ are independent then $\sum_{i=1}^{q} A_i \sim W_p(\Sigma, \sum_{i=1}^{q} n_i)$ will no more follow if Σ_i's are different.

Now we consider two cases when $N_1 = N_2$ and $N_1 \neq N_2$. Then we aim to develop a statistic which can be framed in such a way that it is an unbiased estimator of $\underset{\sim}{\mu}^{(1)} - \underset{\sim}{\mu}^{(2)}$ and can follow a Wishart distribution.

Case I: when $N_1 = N_2 = N$

The null hypothesis: $H_0 : \underset{\sim}{\mu}^{(1)} = \underset{\sim}{\mu}^{(2)}$ and $N_1 = N_2 = N$.

The statistic

$$\underset{\sim}{\bar{X}}^{(1)} - \underset{\sim}{\bar{X}}^{(2)} \sim N_p\left(\underset{\sim}{\mu}^{(1)} - \underset{\sim}{\mu}^{(2)}, \frac{\Sigma_1 + \Sigma_2}{N}\right).$$

In case, $\Sigma_1 \neq \Sigma_2$ and $N_1 \neq N_2$, then $(\underset{\sim}{\bar{x}}^{(1)} - \underset{\sim}{\bar{x}}^{(2)}) \sim N_p\left(\underset{\sim}{\mu}^{(1)} - \underset{\sim}{\mu}^{(2)}, \frac{\Sigma_1}{N_1} + \frac{\Sigma_2}{N_2}\right)$ and the sample covariance matrix

$$\sum_{\alpha=1}^{N_1}(\underset{\sim}{X}_\alpha^{(1)} - \underset{\sim}{\bar{X}}^{(1)})(\underset{\sim}{X}_\alpha^{(1)} - \underset{\sim}{\bar{X}}^{(1)})' + \sum_{\alpha=1}^{N_2}(\underset{\sim}{X}_\alpha^{(2)} - \underset{\sim}{\bar{X}}^{(2)})(\underset{\sim}{X}_\alpha^{(2)} - \underset{\sim}{\bar{X}}^{(2)})'$$

does not follow a Wishart distribution with covariance matrix $\left(\frac{\Sigma_1}{N_1} + \frac{\Sigma_2}{N_2}\right)$.

When $N_1 = N_2 = N$, define

$$\underset{\sim}{y}_\alpha = \underset{\sim}{x}_\alpha^{(1)} - \underset{\sim}{x}_\alpha^{(2)}, \quad \alpha = 1, 2, \ldots, N$$

assuming that the numbering of the observations in the two samples is independent of the observations themselves.

Then $\underset{\sim}{Y}_\alpha \sim N_p(\underset{\sim}{\mu}^{(1)} - \underset{\sim}{\mu}^{(2)}, \Sigma_1 + \Sigma_2), \alpha = 1, 2, \ldots, N$ and independently of $\underset{\sim}{Y}_\beta$ $(\beta \neq \alpha$.

Based on the sample, $\underset{\sim}{y}_1, \underset{\sim}{y}_2, \ldots, \underset{\sim}{y}_N$, let the sample mean vector is

$$\underset{\sim}{\bar{y}} = \frac{1}{N}\sum_{\alpha=1}^{N}\underset{\sim}{y}_\alpha = \underset{\sim}{\bar{x}}^{(1)} - \underset{\sim}{\bar{x}}^{(2)}$$

and the sample covariance matrix is

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

$$= \frac{1}{N-1} \sum_{\alpha=1}^{N} \left[(x_\alpha^{(1)} - \bar{x}^{(1)}) - (x_\alpha^{(2)} - \bar{x}^{(2)}) \right] \left[(x_\alpha^{(1)} - \bar{x}^{(1)}) - (x_\alpha^{(2)} - \bar{x}^{(2)}) \right]'.$$

Then

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

is suitable for testing $H_0 : \mu^{(1)} = \mu^{(2)}$ and T^2 follows the Hotelling T^2 distribution with $(N-1)$ degrees of freedom when H_0 is true.

It is important to note that if we had known $\Sigma_1 = \Sigma_2$, then T^2 will have $2(N-1)$ degrees of freedom. So we lose $(N-1)$ degrees of freedom in constructing a test which is independent of the two covariance matrices.

Case II: $N_1 \neq N_2$

Consider the case when $N_1 \neq N_2$. Suppose $N_1 < N_2$. Once again, we aim to construct such a variable whose expected value is $\mu^{(1)} - \mu^{(2)}$ and can be appropriately applicable for the application of Wishart distribution.

Define a new variable

$$y_\alpha = \left(x_\alpha^{(1)} - \sqrt{\frac{N_1}{N_2}} x_\alpha^{(2)} \right) + \frac{1}{\sqrt{N_1 N_2}} \left(\sum_{\beta=1}^{N_1} x_\beta^{(2)} - \sqrt{\frac{N_1}{N_2}} \sum_{\gamma=1}^{N_2} x_\gamma^{(2)} \right)$$

$$= \left(x_\alpha^{(1)} - \sqrt{\frac{N_1}{N_2}} x_\alpha^{(2)} \right) + \left(\frac{1}{\sqrt{N_1 N_2}} \sum_{\beta=1}^{N_1} x_\beta^{(2)} - \frac{1}{N_2} \sum_{\gamma=1}^{N_2} x_\gamma^{(2)} \right)$$

for $\alpha = 1, 2, \ldots, N_1$. Then

$$E(y_\alpha) = \mu^{(1)} - \sqrt{\frac{N_1}{N_2}} \mu^{(2)} + \frac{N_1}{\sqrt{N_1 N_2}} \mu^{(2)} - \frac{N_2}{N_2} \mu^{(2)}$$

$$= \mu^{(1)} - \mu^{(2)}.$$

The covariance matrix of $\underset{\sim}{y}_\alpha$ and $\underset{\sim}{y}_\beta$ is

$$E\left[(\underset{\sim}{y}_\alpha - E(\underset{\sim}{y}_\alpha))(\underset{\sim}{y}_\beta - E(\underset{\sim}{y}_\beta))'\right]$$

$$= E\left[(\underset{\sim}{x}_\alpha^{(1)} - \underset{\sim}{\mu}^{(1)}) - \sqrt{\frac{N_1}{N_2}}(\underset{\sim}{x}_\alpha^{(2)} - \underset{\sim}{\mu}^{(2)}) + \frac{1}{\sqrt{N_1 N_2}}\sum_{\gamma=1}^{N_1}(\underset{\sim}{x}_\gamma^{(2)} - \underset{\sim}{\mu}^{(2)}) - \frac{1}{N_2}\sum_{\gamma=1}^{N_2}(\underset{\sim}{x}_\gamma^{(2)} - \underset{\sim}{\mu}^{(2)})\right]$$

$$\times \left[(\underset{\sim}{x}_\beta^{(1)} - \underset{\sim}{\mu}^{(1)}) - \sqrt{\frac{N_1}{N_2}}(\underset{\sim}{x}_\beta^{(2)} - \underset{\sim}{\mu}^{(2)}) + \frac{1}{\sqrt{N_1 N_2}}\sum_{\gamma=1}^{N_1}(\underset{\sim}{x}_\gamma^{(2)} - \underset{\sim}{\mu}^{(2)}) - \frac{1}{N_2}\sum_{\gamma=1}^{N_2}(\underset{\sim}{x}_\gamma^{(2)} - \underset{\sim}{\mu}^{(2)})\right]'$$

$$= \delta_{\alpha\beta}\left[\left(\Sigma_1 + \frac{N_1}{N_2}\Sigma_2\right) + \Sigma_2\left(-\frac{2}{N_2} + \frac{2}{N_2}\sqrt{\frac{N_1}{N_2}} + \frac{N_1}{N_1 N_2} - \frac{2N_1}{N_2\sqrt{N_1 N_2}} + \frac{N_2}{N_2^2}\right)\right]$$

$$= \left(\Sigma_1 + \frac{N_1}{N_2}\Sigma_2\right)\delta_{\alpha\beta}$$

where $\delta_{\alpha\beta} = \begin{cases} 1 & \text{if } \alpha = \beta \\ 0 & \text{otherwise} \end{cases}$ is the Kronecker delta.

Thus $H_0 : \underset{\sim}{\mu}^{(1)} = \underset{\sim}{\mu}^{(2)}$ now becomes $E(\underset{\sim}{y}_\alpha) = 0$. A suitable statistic for testing H_0 is

$$T^2 = N_1 \bar{y}' S^{-1} \bar{y}$$

which is distributed as Hotelling T^2 distribution with $(N_1 - 1)$ degrees of freedom where

$$\underset{\sim}{\bar{y}} = \frac{1}{N_1}\sum_{\alpha=1}^{N_1}\underset{\sim}{y}_\alpha = \bar{\underset{\sim}{x}}^{(1)} - \bar{\underset{\sim}{x}}^{(2)}$$

and

$$(N_1 - 1)S = \sum_{\alpha=1}^{N_1}(\underset{\sim}{y}_\alpha - \underset{\sim}{\bar{y}})(\underset{\sim}{y}_\alpha - \underset{\sim}{\bar{y}})'.$$

If we define

$$\underset{\sim}{u}_\alpha = \underset{\sim}{x}_\alpha^{(1)} - \sqrt{\frac{N_1}{N_2}}\underset{\sim}{x}_\alpha^{(2)} \ , \alpha = 1, 2, \ldots, N_1$$

then

$$(N_1 - 1)S = \sum_{\alpha=1}^{N_1}(\underset{\sim}{u}_\alpha - \underset{\sim}{\bar{u}})(\underset{\sim}{u}_\alpha - \underset{\sim}{\bar{u}})'$$

where $\underset{\sim}{\bar{u}} = \frac{1}{N_1}\sum_{\alpha=1}^{N_1}\underset{\sim}{u}_\alpha$.

This procedure was given by Scheffé (1943) in the univariate case, and it was shown that, in the univariate case, this technique yields the shortest confidence intervals among

those obtained by the t-test. One advantage is that the statistic $\bar{x}^{(1)} - \bar{x}^{(2)}$ is the most relevant statistic to $\mu^{(1)} - \mu^{(2)}$.

The sacrifice of observations in estimating the covariance matrix is not so important given the advantages. The case considers $N_1 < N_2$ which is just for convenience. The case $N_2 < N_1$ can be dealt similarly.

Another approach:

Let $x_\alpha^{(i)}$, $\alpha = 1, 2, \ldots, N_i$; $i = 1, 2, \ldots, k$ be samples from $N(\mu_i, \Sigma_i)$ respectively. Consider the hypothesis

$$H_0 : \sum_{i=1}^{k} \beta_i \mu_i = \mu$$

where $\beta_1, \beta_2, \ldots, \beta_k$ are given scalars and μ is given vector. If N_i's are unequal, take N_1 to be the smallest.

Let

$$y_\alpha = \beta_1 x_\alpha^{(1)} + \sum_{i=2}^{k} \beta_i \sqrt{\frac{N_1}{N_i}} \left(x_\alpha^{(i)} - \frac{1}{N_1} \sum_{\beta=1}^{N_1} x_\beta^{(i)} + \frac{1}{\sqrt{N_1 N_i}} \sum_{\gamma=1}^{N_i} x_\gamma^{(i)} \right).$$

Then

$$E(y_\alpha) = \sum_{i=1}^{k} \beta_i \mu_i$$

and

$$E\left[(y_\alpha - E(y_\alpha))(y_\beta - E(y_\beta))' \right] = \delta_{\alpha\beta} \left(\sum_{i=1}^{k} \frac{\beta_i^2 N_1}{N_i} \Sigma_i \right).$$

Let $\bar{y}$ and S be defined by $\bar{y} = \frac{1}{N_1} \sum_{\alpha=1}^{N_1} y_\alpha = \sum_{i=1}^{q} \beta_i \bar{x}_i$ where $\bar{x}_i = \frac{1}{N_i} \sum_{\beta=1}^{N_i} x_\beta^{(i)}$ and $(N_1 - 1)S = \sum_{\alpha=1}^{N_1} (y_\alpha - \bar{y})(y_\alpha - \bar{y})'$.

Then

$$T^2 = N_1 (\bar{y} - \mu)' S^{-1} (\bar{y} - \mu)$$

is suitable for testing H_0. When H_0 is true, T^2 follows T^2 distribution with $(N_1 - 1)$ degrees of freedom.

If we define

$$\underset{\sim}{u}_\alpha = \sum_{i=1}^{q} \beta_i \sqrt{\frac{N_1}{N_i}} x_\alpha^{(i)} \ , \alpha = 1, 2, \ldots, N_1$$

then S can be defined as

$$(N_1 - 1)S = \sum_{\alpha=1}^{N_1} (\underset{\sim}{u}_\alpha - \bar{\underset{\sim}{u}})(\underset{\sim}{u}_\alpha - \bar{\underset{\sim}{u}})'$$

where $\bar{\underset{\sim}{u}} = \frac{1}{N_1} \sum_{\alpha=1}^{N_1} \underset{\sim}{u}_\alpha$.

Another problem that is amenable to this kind of treatment is the test of a hypothesis about two subvectors of the same dimensions having equal means.

Let

$$\underset{\sim}{x} = \begin{pmatrix} \underset{\sim}{x}^{(1)} \\ \underset{\sim}{x}^{(2)} \end{pmatrix} \sim N_{2p} \left[\begin{pmatrix} \underset{\sim}{\mu}^{(1)} \\ \underset{\sim}{\mu}^{(2)} \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right]$$

where $\underset{\sim}{x}^{(1)}$ and $\underset{\sim}{x}^{(2)}$ are of orders $q \times 1$. Then

$$\underset{\sim}{x}^{(1)} - \underset{\sim}{x}^{(2)} \sim N_q(\underset{\sim}{\mu}^{(1)} - \underset{\sim}{\mu}^{(2)}, \Sigma^*)$$

where

$$\Sigma^* = E \left[(\underset{\sim}{x}^{(1)} - \underset{\sim}{\mu}^{(1)}) - (\underset{\sim}{x}^{(2)} - \underset{\sim}{\mu}^{(2)}) \right] \left[(\underset{\sim}{x}^{(1)} - \underset{\sim}{\mu}^{(1)}) - (\underset{\sim}{x}^{(2)} - \underset{\sim}{\mu}^{(2)}) \right]'$$

$$= \Sigma_{11} - \Sigma_{12} - \Sigma_{21} + \Sigma_{22}.$$

To test $H_0 : \underset{\sim}{\mu}^{(1)} = \underset{\sim}{\mu}^{(2)}$, the T^2 statistic is used as

$$T^2 = N(\bar{\underset{\sim}{x}}^{(1)} - \bar{\underset{\sim}{x}}^{(2)})'(S_{11} - S_{12} - S_{21} + S_{22})^{-1}(\bar{\underset{\sim}{x}}^{(1)} - \bar{\underset{\sim}{x}}^{(2)})$$

with $\bar{\underset{\sim}{x}} = \begin{pmatrix} \bar{\underset{\sim}{x}}^{(1)} \\ \bar{\underset{\sim}{x}}^{(2)} \end{pmatrix}$, $S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}$.

Application of the Union-Intersection Principle to obtain Hotelling's T^2-Statistics

We use the Union-Intersection principle to derive Hotelling's T^2 statistic.

Union-Intersection Principle:

Consider $\underset{\sim}{X}_\alpha \sim N_p(\underset{\sim}{\mu}, \Sigma)$, $\alpha = 1, 2, \ldots, N$ where Σ is unknown and independent of $\underset{\sim}{x}_\alpha$. Consider the hypothesis $H_0 : \underset{\sim}{\mu} = 0$ versus $H_1 : \underset{\sim}{\mu} \neq 0$. Let $\underset{\sim}{a}$ be an arbitrary nonnull vector of p constant elements of order $p \times 1$.

When H_0 is true, then $\underset{\sim}{a}'\underset{\sim}{\mu} = 0$ for every $\underset{\sim}{a}$. Conversely, if $\underset{\sim}{a}'\underset{\sim}{\mu} = 0$ for every $\underset{\sim}{a}$, then H_0 will be true.

Thus H_0 will be rejected if at least one of the hypothesis

$$H_{\underset{\sim}{a}} : \underset{\sim}{a}'\underset{\sim}{\mu} = 0, \; \underset{\sim}{a} \in A$$

is rejected where A is the set of all p-component non-null vectors of constants. Thus

$$H_0 = \bigcap_{\underset{\sim}{a}} H_{\underset{\sim}{a}}.$$

Let $W_{\underset{\sim}{a}}$ be the critical region for $H_{\underset{\sim}{a}}$. If sample observations are such that the "sample point" falls in the critical region for at least one of the hypotheses $H_{\underset{\sim}{a}}$ ($\underset{\sim}{a} \in A$) and hence H_0 will be rejected.

So the critical region for H_0 will be formed by the Union of these critical regions for the separate $H_{\underset{\sim}{a}}$. Thus $W = \bigcup_{\underset{\sim}{a}} W_{\underset{\sim}{a}}$. This is the Union-Intersection principle.

The sizes of $W_{\underset{\sim}{a}}$ should be such that the final size of W turns out to be α, which is preassigned.

The merit is that $H_{\underset{\sim}{a}}$ is about the scalar $\underset{\sim}{a}'\underset{\sim}{\mu}$ (which is $E(\underset{\sim}{a}'\underset{\sim}{x})$) and we have optimum test for univariate hypothesis, we may be able to obtain a good test (in some sense) for the multivariate H_0.

Derivation of T^2 statistics:

Let $\underset{\sim}{X} \sim N_p(\underset{\sim}{\mu}, \Sigma)$. Consider the hypothesis

$$H_0 : \underset{\sim}{a}'\underset{\sim}{\mu} = \underset{\sim}{a}'\underset{\sim}{\mu}_0,$$

$H_1 : \underset{\sim}{a}'\mu \neq \underset{\sim}{a}'\mu_0.$

Then $\underset{\sim}{a}'\underset{\sim}{x} \sim N_p(\underset{\sim}{a}'\mu, \underset{\sim}{a}'\Sigma\underset{\sim}{a})$. Also $\frac{\underset{\sim}{a}'S\underset{\sim}{a}}{\underset{\sim}{a}'\Sigma\underset{\sim}{a}} \sim \chi_n^2$ and is independent of $\underset{\sim}{a}'\underset{\sim}{x}$. The hypothesis $H_{\underset{\sim}{a}}$ can be tested by Student's t-test statistic

$$t_{\underset{\sim}{a}} = \frac{\underset{\sim}{a}'\underset{\sim}{\bar{x}} - \underset{\sim}{a}'\underset{\sim}{\mu_0}}{\sqrt{\underset{\sim}{a}'S\underset{\sim}{a}/N}}$$

and the critical region $W_{\underset{\sim}{a}}$ for $H_{\underset{\sim}{a}}$ is thus $|t_{\underset{\sim}{a}}| > t_*$, which is a suitable constant depending on N and size of $W_{\underset{\sim}{a}}$.

Alternatively, it is equivalent to $t_{\underset{\sim}{a}}^2 > t_*^2$.

If each $t_{\underset{\sim}{a}}^2 < t_*^2$, none of $H_{\underset{\sim}{a}}$ will be rejected. If any $t_{\underset{\sim}{a}}^2 \geq t_*^2$, the corresponding $H_{\underset{\sim}{a}}$ and hence H_0 will be rejected. Thus H_0 will not be rejected if each $t_{\underset{\sim}{a}}^2 \leq t_*^2$. This will be so if $\max_{\underset{\sim}{a}} t_{\underset{\sim}{a}}^2 < t_*^2$ with respect to $a < t_*^2$. Hence, this procedure will be simpler if we find this maximum value instead of verifying whether $t_{\underset{\sim}{a}}^2 < t_*^2$ or not for each and every $\underset{\sim}{a}$ in A.

Now

$$t_{\underset{\sim}{a}}^2 = \frac{N(\underset{\sim}{a}'\underset{\sim}{\bar{x}} - \underset{\sim}{a}'\underset{\sim}{\mu_0})^2}{\underset{\sim}{a}'S\underset{\sim}{a}} = \frac{N\underset{\sim}{a}'(\underset{\sim}{\bar{x}} - \underset{\sim}{\mu_0})(\underset{\sim}{\bar{x}} - \underset{\sim}{\mu_0})'\underset{\sim}{a}}{\underset{\sim}{a}'S\underset{\sim}{a}}.$$

Using the lemma

$$\sup_{\underset{\sim}{\theta} \in \Theta} \frac{\underset{\sim}{\theta}'A\underset{\sim}{\theta}}{\underset{\sim}{\theta}'B\underset{\sim}{\theta}} = tr(AB^{-1})$$

with $A = (\underset{\sim}{\bar{x}} - \underset{\sim}{\mu_0})(\underset{\sim}{\bar{x}} - \underset{\sim}{\mu_0})'$ and $B = S$, then

$$tr(AB^{-1}) = tr\left[\sqrt{N}(\underset{\sim}{\bar{x}} - \underset{\sim}{\mu_0})\right]\left[\sqrt{N}(\underset{\sim}{\bar{x}} - \underset{\sim}{\mu_0})\right]'S^{-1}$$

$$= (\underset{\sim}{\bar{x}} - \underset{\sim}{\mu_0})'S^{-1}(\underset{\sim}{\bar{x}} - \underset{\sim}{\mu_0}).$$

Thus

$$\max_{\underset{\sim}{a}} t_{\underset{\sim}{a}}^2 = N(\underset{\sim}{\bar{x}} - \underset{\sim}{\mu_0})'S^{-1}(\underset{\sim}{\bar{x}} - \underset{\sim}{\mu_0})$$

which is Hotelling T^2 based on N degrees of freedom.

The critical region for H_0 is $T^2 > t_*^2$ where t_* is to be chosen such that the size of W is predetermined constant α.