10 Further Multivariate Hypothesis Testing: More One and Two-Sample Procedures

10.1 Introduction

In the last chapter we introduced two new distributions based on the multivariate normal distribution. These were the Wishart distribution (the multivariate generalization of the χ^2 distribution) and the Hotelling T^2 distribution (generalizing the t and F distributions).

Consider a random sample, $\mathbf{X}_1, \ldots, \mathbf{X}_n$, from $\mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are unknown. We have now established the following results.

(i) The maximum likelihood estimates of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are $\overline{\mathbf{X}}$ and $\mathbf{S}$, and these are independently distributed.

(ii) The corresponding sampling distributions are given by

$$\overline{\mathbf{X}} \sim N_p(\boldsymbol{\mu}, \frac{1}{n}\boldsymbol{\Sigma})$$

and

$$n\,\mathbf{S} \;=\; (n-1)\mathbf{S}_U \;\sim\; \mathrm{W}_p(n-1, \boldsymbol{\Sigma}) \quad \text{(Proposition 9.2)}.$$

(iii) From Corollary 9.5 we have

$$(n-1)\,(\overline{\mathbf{X}} - \boldsymbol{\mu})^T \mathbf{S}^{-1} (\overline{\mathbf{X}} - \boldsymbol{\mu}) \;=\; n\,(\overline{\mathbf{X}} - \boldsymbol{\mu})^T \mathbf{S}_U^{-1} (\overline{\mathbf{X}} - \boldsymbol{\mu}) \;\sim\; \mathrm{T}_p^2(n-1).$$

(iv) We also recall the more general result of Theorem 9.4, which tells us that if $\mathbf{Y}$ and $\mathbf{A}$ are independently distributed as $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\mathrm{W}_p(m, \boldsymbol{\Sigma})$ respectively, then

$$m(\mathbf{Y} - \boldsymbol{\mu})^T \mathbf{A}^{-1} (\mathbf{Y} - \boldsymbol{\mu}) \sim \mathrm{T}_p^2(m).$$

(v) And Theorem 9.6 tells us that, for $m > (p-1)$,

$$\mathrm{T}_p^2(m) \;=\; \frac{mp}{m-p+1}\mathrm{F}(p, m-p+1).$$

We considered two general approaches to the development of suitable multivariate test statistics, the union intersection test (UIT), and the generalized likelihood ratio test (LRT).

Considering the one-sample test

$$H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0 \text{ vs } H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$$

both the UIT and LRT approaches to test construction gave the test statistic

$$T^2 \;=\; (n-1)(\overline{\mathbf{x}}-\boldsymbol{\mu}_0)^T\mathbf{S}^{-1}(\overline{\mathbf{x}}-\boldsymbol{\mu}_0) \;=\; n(\overline{\mathbf{x}}-\boldsymbol{\mu}_0)^T\mathbf{S}_U^{-1}(\overline{\mathbf{x}}-\boldsymbol{\mu}_0).$$

Under H_0, T^2 is an observation drawn from the distribution $\mathrm{T}_p^2(n-1)$ and large values for T^2 are evidence against H_0. At significance level α, the rejection (critical) region for the test is given by

$$\left\{T^2 : T^2 > \frac{(n-1)p}{n-p}\mathrm{F}_\alpha(p, n-p)\right\}.$$

Note again that, because T^2 is the UIT statistic, $T^2 = \max_{\mathbf{a}} t^2(\mathbf{a})$, where the maximizing $\mathbf{a}$ gives the linear compound with the largest $|t|$ statistic in a test of

$$H_0 : \mathbf{a}^T\boldsymbol{\mu} = \mathbf{a}^T\boldsymbol{\mu}_0 \text{ vs } H_1 : \mathbf{a}^T\boldsymbol{\mu} \neq \mathbf{a}^T\boldsymbol{\mu}_0.$$

i.e. we have the combination of variables that deviates most from its expected value under H_0. Recall that a maximizing $\mathbf{a}$, is given by $\mathbf{a}^* = \mathbf{S}^{-1}(\overline{\mathbf{x}}-\boldsymbol{\mu}_0)$ or $\mathbf{a}^* = \mathbf{S}_U^{-1}(\overline{\mathbf{x}}-\boldsymbol{\mu}_0)$.

10.2 Confidence Regions for the Mean Vector $\boldsymbol{\mu}$

The T^2 test can be inverted to give a confidence region for the elements of the population mean vector $\boldsymbol{\mu}$ given a sample of size n with sample mean vector $\bar{\mathbf{x}}$ and sample covariance matrix $\mathbf{S}$.

The $100(1-\alpha)\%$ confidence region consists of vectors $\boldsymbol{\mu}$ satisfying

$$(n-1)(\bar{\mathbf{x}}-\boldsymbol{\mu})^T\mathbf{S}^{-1}(\bar{\mathbf{x}}-\boldsymbol{\mu}) \leq \mathrm{T}^2_{p;\alpha}(n-1) = \frac{(n-1)p}{n-p}\mathrm{F}_\alpha(p, n-1)$$

or,

$$(n-1)(\boldsymbol{\mu}-\bar{\mathbf{x}})^T\mathbf{S}^{-1}(\boldsymbol{\mu}-\bar{\mathbf{x}}) \leq \frac{(n-1)p}{n-p}\mathrm{F}_\alpha(p, n-p)$$

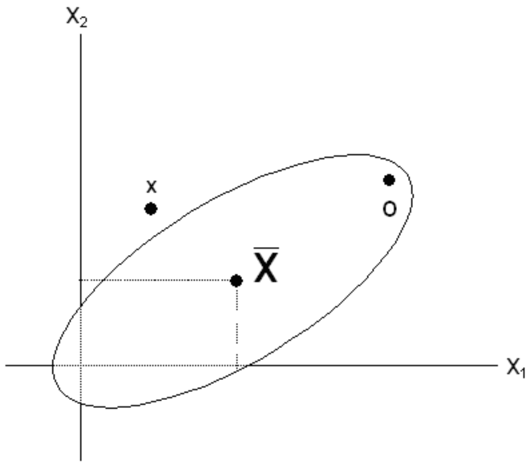
The boundary of this region is a hyperellipsoid centred at $\boldsymbol{\mu} = \bar{\mathbf{x}}$. The interior points of the region represent those points $\boldsymbol{\mu}_0$ which would *not* be rejected in a single application of the T^2 test using this sample. Thus

$$\begin{cases} \text{if } \boldsymbol{\mu}_0 \text{ is within confidence region} & \text{do not reject } H_0 \\ \text{if } \boldsymbol{\mu}_0 \text{ is outside confidence region} & \text{reject } H_0 \end{cases}$$

In the bivariate case it can be helpful to plot $100(1-\alpha)\%$ confidence region. For example, it can tell us more about why an hypothesis has been rejected or not.

If we look at the points $\bar{\mathbf{x}}$, x and O on the following plot, the multivariate result appears surprising in terms of the Euclidean distances seen by the eye - but these distances take no account of correlations. Points x and O represent two extremes. Univariate $t-$tests on x_1 and x_2 may give a different result from the multivariate T^2 test.

95% Confidence Region for $\boldsymbol{\mu}$



We would like to have a measure of distance making all the points on the boundary equidistant from $\bar{\mathbf{x}}$. This distance is the *Mahalanobis* distance (see autumn term lectures).

The sample Mahalanobis distance between any two points $\mathbf{x}$ and $\mathbf{y}$ is given by

$$D^2(\mathbf{x}, \mathbf{y}) = (\mathbf{x} - \mathbf{y})^T \mathbf{S}^{-1} (\mathbf{x} - \mathbf{y})$$

We can now see that the one-sample T^2 test statistic is $(n-1)D^2(\bar{\mathbf{x}}, \boldsymbol{\mu}_0)$.

Note: If we make the transformation $\mathbf{z} = \mathbf{S}^{-\frac{1}{2}}(\mathbf{x} - \bar{\mathbf{x}})$, so that $\bar{\mathbf{z}} = \mathbf{0}$ and sample $\text{Cov}(\mathbf{z}, \mathbf{z}) = \mathbf{S}^{-\frac{1}{2}} \mathbf{S} \mathbf{S}^{-\frac{1}{2}} = \mathbf{I}$. Then, since $\mathbf{x} = \mathbf{S}^{\frac{1}{2}} \mathbf{z} + \bar{\mathbf{x}}$

$$\begin{aligned} D^2(\mathbf{x}_1, \mathbf{x}_2) &= (\mathbf{x}_1 - \mathbf{x}_2)^T \mathbf{S}^{-1} (\mathbf{x}_1 - \mathbf{x}_2) \\ &= (\mathbf{z}_1 - \mathbf{z}_2)^T \mathbf{S}^{\frac{1}{2}} \mathbf{S}^{-1} \mathbf{S}^{\frac{1}{2}} (\mathbf{z}_1 - \mathbf{z}_2) \\ &= (\mathbf{z}_1 - \mathbf{z}_2)^T (\mathbf{z}_1 - \mathbf{z}_2) \end{aligned}$$

which is the (squared) Euclidean distance between $\mathbf{z}_1$ and $\mathbf{z}_2$. ($\mathbf{z} = \mathbf{S}^{-\frac{1}{2}}(\mathbf{x} - \bar{\mathbf{x}})$ is called the *Mahalanobis transformation*).

Note also that the sample Mahalanobis distance is often taken to be

$$D^2(\mathbf{x}, \mathbf{y}) = (\mathbf{x} - \mathbf{y})^T \mathbf{S}_U^{-1} (\mathbf{x} - \mathbf{y})$$

in which case $T^2 = nD^2(\bar{\mathbf{x}}, \boldsymbol{\mu}_0)$.

10.3 Simultaneous Confidence Intervals for all Linear Compounds of the Mean Vector

The significance of T^2 does not show which of the components of $\mathbf{x}$ has led to the rejection of H_0.

To test the components separately, or to test particular linear combinations of the components, we use the union-intersection derivation of the T^2 test. This allows us to control the $\alpha-$probability for *any number* of tests on linear compounds of $\bar{\mathbf{x}}$.

Thus, with $\mathbf{x}_1, \dots, \mathbf{x}_n$ independently drawn from $\mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, and $\mathbf{a}$ any vector of constants

$$t^2(\mathbf{a}) \;=\; \frac{(n-1)[\mathbf{a}^T(\bar{\mathbf{x}}-\boldsymbol{\mu})]^2}{\mathbf{a}^T\mathbf{S}\mathbf{a}} \;\leq\; (n-1)(\bar{\mathbf{x}}-\boldsymbol{\mu})^T\mathbf{S}^{-1}(\bar{\mathbf{x}}-\boldsymbol{\mu}) \;\sim\; \mathrm{T}^2_p(n-1).$$

Hence,

$$\mathbb{P}\left\{t^2(\mathbf{a}) \leq \mathrm{T}^2_{p;\alpha}(n-1), \quad \text{for all } \mathbf{a}\right\} = 1-\alpha,$$

where

$$\mathrm{T}^2_{p;\alpha}(n-1) = \frac{(n-1)p}{n-p}\mathrm{F}_\alpha(p, n-p).$$

So, if we set $s^2_{\mathbf{a}} = \mathbf{a}^T\mathbf{S}\mathbf{a}$ (the sample variance of $\mathbf{a}^T\mathbf{x}$), we have

$$\mathbb{P}\left\{\mathbf{a}^T\bar{\mathbf{x}} - \frac{s_{\mathbf{a}}}{\sqrt{n-1}}\mathrm{T}_{p;\alpha}(n-1) \;\leq\; \mathbf{a}^T\boldsymbol{\mu} \leq \mathbf{a}^T\bar{\mathbf{x}} + \frac{s_{\mathbf{a}}}{\sqrt{n-1}}\mathrm{T}_{p;\alpha}(n-1), \quad \text{for all } \mathbf{a}\right\} \;=\; 1-\alpha$$

i.e.

$$\mathbb{P}\left\{\mathbf{a}^T\boldsymbol{\mu} \in \mathbf{a}^T\bar{\mathbf{x}} \pm \frac{s_{\mathbf{a}}}{\sqrt{n-1}}\mathrm{T}_{p;\alpha}(n-1), \quad \text{for all } \mathbf{a}\right\} = 1-\alpha$$

For a given sample of size n with sample mean vector $\bar{\mathbf{x}}$ and sample covariance matrix $\mathbf{S}$, the above are *simultaneous* confidence intervals for all linear compounds of the population mean vector $\boldsymbol{\mu}$.

e.g. If $\mathbf{a}^T = (1, 0, \dots, 0)$, then we have $\mu_1 \in \bar{x}_1 \pm \frac{s_1}{\sqrt{n-1}}\mathrm{T}_{p;\alpha}(n-1)$ with $100(1-\alpha)\%$ confidence.

$\Big($Compare $\mu_1 \in \bar{x}_1 \pm \frac{s_1}{\sqrt{n-1}}\mathrm{t}_{\alpha/2}(n-1)$, the <u>univariate</u> $100(1-\alpha)\%$ confidence interval.$\Big)$

If H_0 has been rejected at significance level α, then we know there is at least one $\mathbf{a}$ such that the interval $\mathbf{a}^T\bar{\mathbf{x}} \pm \mathrm{T}_{p;\alpha}(n-1)\frac{s_{\mathbf{a}}}{\sqrt{n-1}}$ does *not* contain $\mathbf{a}^T\boldsymbol{\mu}_0$.

In particular we know that $\mathbf{a} = \mathbf{S}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0) = \mathbf{a}^*$ maximizes $t^2(\mathbf{a})$ and therefore

$$\mathbf{a}^{*T}\boldsymbol{\mu}_0 \notin \mathbf{a}^{*T}\bar{\mathbf{x}} \pm \mathrm{T}_{p;\alpha}(n-1)\frac{s_{\mathbf{a}}}{\sqrt{n-1}}$$

It is always worth looking at $\mathbf{a}^*$ to see exactly which linear compound of $\mathbf{x}$ most differs from its expected value under H_0.

10.4 Testing for Structural Relations within Components of Mean

Here, we test whether the data are consistent with a particular hypothesis regarding the relationships between the components of $\boldsymbol{\mu}$.

As usual, it will be assumed that the data arise from the $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution, where both $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are unknown.

Example 10.1 (Equality of the components of the mean vector)
Determine whether or not the components of the (population) mean vector are equal to each other.

Solution: Formally, we seek a test for

$$H_0 : \mu_1 = \mu_2 = \ \ldots \ \mu_p \quad \text{versus} \quad H_1 : \ \mu_k \neq \mu_\ell \text{ for some } k \neq \ell.$$

i.e.

$$H_0 : \mu_1 - \mu_k = 0,\ k{=}2,\ldots,p, \ \text{ vs. } \ H_1 : \ \mu_k \neq \mu_1 \text{ for some } k \neq \ell.$$

This can be expressed in vector form as

$$H_0 : \begin{bmatrix} \mu_1 - \mu_2 \\ \mu_1 - \mu_3 \\ \vdots \\ \mu_1 - \mu_p \end{bmatrix} = \mathbf{0}$$

using a $(p-1) \times 1$ column vector.

Again, this can be rewritten as

$$\begin{bmatrix} 1 & -1 & 0 & 0 & \ldots & 0 \\ 1 & 0 & -1 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & 0 & \ldots & -1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} = \mathbf{0}$$

or

$$\mathbf{C}\boldsymbol{\mu} = \mathbf{0}$$

where

$$\mathbf{C} = \begin{bmatrix} 1 & -1 & 0 & 0 & \dots & 0 \\ 1 & 0 & -1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & 0 & \dots & -1 \end{bmatrix}$$

which is a $(p-1) \times p$ matrix of constants.

The test to be carried out is therefore

$$H_0 : \mathbf{C}\boldsymbol{\mu} = \mathbf{0} \text{ versus } H_1 : \mathbf{C}\boldsymbol{\mu} \neq \mathbf{0}.$$

Now

$$\mathbf{X}_i \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}),\ i=1,\dots,n,$$

which means that

$$\overline{\mathbf{X}} \sim N_p(\boldsymbol{\mu}, \frac{1}{n}\boldsymbol{\Sigma})$$

and so

$$\mathbf{C}\overline{\mathbf{X}} \sim N_{p-1}(\mathbf{C}\boldsymbol{\mu}, \frac{1}{n}\mathbf{C}\boldsymbol{\Sigma}\mathbf{C}^{\mathbf{T}}).$$

Also

$$n\,\mathbf{CSC}^T \sim W_{p-1}(n-1, \mathbf{C}\boldsymbol{\Sigma}\mathbf{C}^{\mathbf{T}})$$

independently (using the fact that $n\,\mathbf{S} \sim W_p(n-1, \boldsymbol{\Sigma})$ and Remarks 9.1 (iv)).

Hence, under H_0 and using Theorem 9.4,

$$\begin{aligned} T^2 &= (n-1)\,(\mathbf{C}\overline{\mathbf{X}} - \mathbf{C}\boldsymbol{\mu})^T(\mathbf{CSC}^T)^{-1}(\mathbf{C}\overline{\mathbf{X}} - \mathbf{C}\boldsymbol{\mu}) \\ &= (n-1)\,\overline{\mathbf{X}}^T\mathbf{C}^T(\mathbf{CSC}^T)^{-1}\mathbf{C}\overline{\mathbf{X}} \\ &\sim T^2_{p-1}(n-1) \end{aligned}$$

for $n > p-1$. (Check this result as an exercise. See the proof of Theorem 9.4.)

Hence, we reject H_0 at the $100\alpha\%$ level of significance if

$$\frac{n-p+1}{(n-1)(p-1)}T^2_{obs} > F_{p-1,n-p+1}(\alpha)$$

and fail to reject otherwise. It is easily shown that this is both the UIT and LRT test of H_0 versus H_1.

Generalization

The procedure described above can be generalized to any linear hypothesis involving the components of the mean. Assume that $\mathbf{X}_i \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $i = 1, \ldots, n$, mutually independent, where both $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are unknown.

If we seek a test for

$$H_0 : \mathbf{C}\boldsymbol{\mu} = \boldsymbol{\phi} \text{ vs. } H_1 : \mathbf{C}\boldsymbol{\mu} \neq \boldsymbol{\phi}$$

where $\mathbf{C}$ is a $m \times p$ matrix of constants, $\text{rank}(\mathbf{C}) = m \leq p < n$, and $\boldsymbol{\phi}$ is a $m \times 1$ vector of constants,

then we use

$$T^2 = (n-1)\,(\mathbf{C}\overline{\mathbf{X}} - \boldsymbol{\phi})^T (\mathbf{C}\mathbf{S}\mathbf{C}^T)^{-1} (\mathbf{C}\overline{\mathbf{X}} - \boldsymbol{\phi}).$$

A simple extended derivation using the UIT and LRT principles shows that this T^2 is the UIT and LRT statistic, and so if H_0 is true, T^2 should be 'small'; while if H_0 is false, then T^2 should be 'large'.

Under H_0, an easy application of Theorem 9.4 tells us that $T^2 \sim T^2_m(n-1)$.

Thus, we reject H_0 if

$$\frac{n-m}{(n-1)m} T^2_{obs} > F_{m,n-m}(\alpha)$$

$n > m$, and accept H_0 otherwise.

10.5 Two-Sample Hotelling T^2 Test

Suppose that we draw two samples of sizes n_1 and n_2 from populations 1 and 2 respectively.

Observations from population i are assumed to be mutually independent with distribution $N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$, where both $\boldsymbol{\mu}_i$, and $\boldsymbol{\Sigma}$, are unknown.

We wish to test whether the two populations have the same mean or not, i.e.

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \text{ versus } H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2.$$

Notice that we have assumed that the two populations share the same covariance matrix.

Let $\overline{\mathbf{X}}_1$ and $\mathbf{S}_1$ be the sample mean vector and maximum likelihood sample covariance matrix for the random sample drawn from population 1.

Similarly, let $\overline{\mathbf{X}}_2$ and $\mathbf{S}_2$ be the sample mean vector and the maximum likelihood sample covariance matrix for the random sample from population 2, drawn independently of the first.

We have

$$\overline{\mathbf{X}}_i \sim N_p(\boldsymbol{\mu}_i, \frac{1}{n_i}\boldsymbol{\Sigma}) \text{ independently for } i = 1, 2,$$

and it follows easily that

$$\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2 \sim N_p\left(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2, \left(\frac{1}{n_1} + \frac{1}{n_2}\right)\boldsymbol{\Sigma}\right).$$

Under H_0

$$\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2 \sim N_p\left(\mathbf{0}, \left(\frac{1}{n_1} + \frac{1}{n_2}\right)\boldsymbol{\Sigma}\right). \qquad (A)$$

We now define

$$\mathbf{S} = \frac{n_1\mathbf{S}_1 + n_2\mathbf{S}_2}{n_1 + n_2},$$

the pooled estimate of $\boldsymbol{\Sigma}$ for which

$$(n_1 + n_2)\mathbf{S} \sim W_p(n_1 + n_2 - 2, \boldsymbol{\Sigma}). \qquad (B)$$

It is easily shown that $\mathbf{S}$ is the *maximum likelihood* estimate of $\boldsymbol{\Sigma}$, and that $(\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2)$ and $\mathbf{S}$ are independently distributed.

Now, from (A),

$$\sqrt{\frac{n_1 n_2}{n_1 + n_2}}(\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2) \sim \mathrm{N}_p(\mathbf{0}, \boldsymbol{\Sigma}) \qquad (C)$$

so that, from Theorem 9.4 (result (iv) in Section 10.1), (B) and (C)

$$\begin{aligned} T^2 &= (n_1 + n_2 - 2)\left[\sqrt{\frac{n_1 n_2}{n_1 + n_2}}(\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2)^T (n_1 + n_2)^{-1}\mathbf{S}^{-1}\sqrt{\frac{n_1 n_2}{n_1 + n_2}}(\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2)\right] \\ &= \frac{n_1 + n_2 - 2}{n_1 + n_2}\,\frac{n_1 n_2}{n_1 + n_2}(\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2)^T\mathbf{S}^{-1}(\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2) \sim \mathrm{T}^2_p(n_1 + n_2 - 2) \end{aligned}$$

Equivalently,

$$T^2 = \frac{n_1 n_2}{n_1 + n_2}(\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2)^T\mathbf{S}_U^{-1}(\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2) \sim \mathrm{T}^2_p(n_1 + n_2 - 2)$$

where $\mathbf{S}_U = \dfrac{(n_1 - 1)\mathbf{S}_{1U} + (n_2 - 1)\mathbf{S}_{2U}}{n_1 + n_2 - 2}$ is the *unbiased* estimate of $\boldsymbol{\Sigma}$.

i.e.

$$\frac{n_1 n_2}{n_1+n_2}(\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2)^T \mathbf{S}_U^{-1}(\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2) \;\sim\; \frac{(n_1+n_2-2)\,p}{n_1+n_2-p-1}\,\mathrm{F}(p, n_1+n_2-p-1)$$

or

$$\frac{n_1+n_2-p-1}{(n_1+n_2-2)\,p}\,T^2 \;\sim\; \mathrm{F}(p, n_1+n_2-p-1).$$

It would therefore seem sensible to reject H_0 at the significance level α if

$$T^2 \;>\; \frac{(n_1+n_2-2)\,p}{n_1+n_2-p-1}\,\mathrm{F}_\alpha(p, n_1+n_2-p-1).$$

Once again this is the likelihood ratio test. It can also be derived using the Union-Intersection Principle. That is, for observed sample mean vectors $\overline{\mathbf{x}}_1$ and $\overline{\mathbf{x}}_2$

Theorem 10.2

$$\begin{aligned}\max_{\mathbf{a}} t^2(\mathbf{a}) &= \max_{\mathbf{a}} \frac{n_1 n_2}{n_1+n_2}\frac{[\mathbf{a}^T(\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2)]^2}{\mathbf{a}^T \mathbf{S}_U \mathbf{a}} \\ &= \frac{n_1 n_2}{n_1+n_2}(\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2)^T \mathbf{S}_U^{-1}(\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2)\end{aligned}$$

and the maximum is achieved for

$$\mathbf{a} \;=\; \mathbf{S}_U^{-1}(\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2) \;=\; \mathbf{a}^*,$$

(or, equivalently, for $\mathbf{a} \;=\; \mathbf{S}^{-1}(\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2)$).

Proof: Do as an Exercise.

Note 1: The compound $\mathbf{y} = \mathbf{a}^{*T}\mathbf{x}$ for a maximizing $\mathbf{a}^*$ is called the linear discriminant function and was first proposed by R. A. Fisher for the purpose of classifying a new individual into one of two populations. It is the combination of original variables that distinguishes most clearly between the populations.

Note 2: As before, T^2 is related to a quantity called the sample Mahalanobis distance between two samples. i.e.

$$D^2(\overline{\mathbf{x}}_1, \overline{\mathbf{x}}_2) \;=\; (\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2)^T \mathbf{S}_U^{-1}(\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2)$$

and, under H_0,

$$T^2 \;=\; \frac{n_1 n_2}{n_1+n_2} D^2(\overline{\mathbf{x}}_1, \overline{\mathbf{x}}_2) \;\sim\; \mathrm{T}^2_p(n_1+n_2-2).$$

Confidence Regions for $\Delta = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$

If $\Delta = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$ then, for given sample mean vectors $\overline{\mathbf{x}}_1$ and $\overline{\mathbf{x}}_2$, and pooled, unbiased sample covariance matrix $\mathbf{S}_U$, the $100(1-\alpha)\%$ confidence region for Δ is given by

$$\frac{n_1 n_2}{n_1+n_2}(\Delta - (\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2)^T \mathbf{S}_U^{-1}(\Delta - (\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2)) \;\leq\; \mathrm{T}^2_{p;\alpha}(n_1+n_2-2),$$

where

$$\mathrm{T}_{p;\alpha}(n_1+n_2-2) \;=\; \frac{p\,(n_1+n_2-2)}{n_1+n_2-p-1}\,\mathrm{F}_\alpha(p, n_1+n_2-p-1).$$

The boundary of this region is a hyperellipsoid centred at the point $\Delta = (\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2)$. Note that if $\mathbf{0}$ is not contained within this region then H_0 would be rejected by the T^2 test.

Simultaneous Confidence Intervals

As before, the Union-Intersection approach can give us further information if H_0 has been rejected. The $100(1-\alpha)\%$ simultaneous confidence intervals for all linear compounds of $\mathbf{a}^T\Delta$ of the mean difference vector are given by

$$\mathbf{a}^T\Delta \;\in\; \mathbf{a}^T(\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2) \;\pm\; s_a\sqrt{\frac{n_1+n_2}{n_1 n_2}\,\mathrm{T}_{p;\alpha}(n_1+n_2-2)}$$

where $s_a^2 = \mathbf{a}^T\mathbf{S}_U\mathbf{a}$ and

$$\mathrm{T}^2_{p;\alpha}(n_1+n_2-2) \;=\; \frac{(n_1+n_2-2)\,p}{n_1+n_2-p-1}\,\mathrm{F}_\alpha(p, n_1+n_2-p-1).$$

Once again, if H_0 is rejected, there is at least one $\mathbf{a}$ such that the corresponding confidence interval does not include 0. e.g.

$$\mathbf{a} \;=\; \mathbf{a}^* \;=\; \mathbf{S}_U^{-1}(\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2)$$

(or, equivalently, $\mathbf{a} \;=\; \mathbf{S}^{-1}(\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2)$).

Example 10.3 (Whale Data)
An ocean research laboratory conducted a survey on two species of whale: the Blue whale and the Bowhead whale. Three measurements were taken on each one of the whales in the survey: body length (in metres), weight (in tonnes), and right flipper length (in metres). Data on two independent random samples from the two species are presented below.

Blue Whale(A):
(sample size n_1=4)

Body length	Weight	Flipper length
24.30	109.74	2.46
24.96	108.95	1.95
25.36	109.12	1.75
25.74	109.44	2.35

Bowhead Whale(B):
(sample size n_2=5)

Body length	Weight	Flipper length
22.39	83.07	2.53
22.45	81.84	2.62
22.75	82.81	3.39
20.92	81.90	2.94
21.64	82.65	2.19

We would like to test the hypothesis that the population mean values for body length, weight and flipper length are the same for Blue whales and Bowhead whales.

```
> blue.l <- c(24.3, 24.96, 25.36, 25.74)
> blue.w <- c(109.74, 108.95, 109.12, 109.44)
> blue.f <- c(2.46, 1.95, 1.75, 2.35)
> bow.l <- c(22.39, 22.45, 22.75, 20.92, 21.64)
> bow.w <- c(83.07, 81.84, 82.81, 81.9, 82.65)
> bow.f <- c(2.53, 2.62, 3.39, 2.94, 2.19)

> blue.whale <- data.frame(blue.l, blue.w, blue.f)
> blue.whale
  blue.l blue.w blue.f
1  24.30 109.74   2.46
2  24.96 108.95   1.95
3  25.36 109.12   1.75
4  25.74 109.44   2.35

> bowhead.whale <- data.frame(bow.l, bow.w, bow.f)
> bowhead.whale
  bow.l bow.w bow.f
1 22.39 83.07  2.53
2 22.45 81.84  2.62
3 22.75 82.81  3.39
4 20.92 81.90  2.94
5 21.64 82.65  2.19
```

```
> sm.blue <- apply(blue.whale, 2, mean)
> sm.bow <- apply(bowhead.whale, 2, mean)
> cov.blue <- var(blue.whale)
> cov.bow <- var(bowhead.whale)

> sm.blue
  blue.l   blue.w   blue.f
 25.0900 109.3125   2.1275

> sm.bow
 bow.l  bow.w  bow.f
22.030 82.454  2.734

> sm.blue - sm.bow
 blue.l  blue.w  blue.f
 3.0600 26.8585 -0.6065

> p <- 3
> n1 <- 4
> n2 <- 5

> s.pooled <- ((n1 - 1) * cov.blue + (n2 - 1) * cov.bow)/(n1 + n2 - 2)
> s.pooled
> s.pooled
           blue.l     blue.w     blue.f
blue.l 0.47785714 0.07128571 0.01965714
blue.w 0.07128571 0.22799929 0.03780643
blue.f 0.01965714 0.03780643 0.16534214

> T2 <- ((n1 * n2)/(n1 + n2)) * t(sm.blue - sm.bow) %*% solve(s.pooled) %*% (sm.blue - sm.bow)

> df <- n1 + n2 - 2
> F.stat <- drop(((df - p + 1)/(df * p)) * T2)
> F.stat
[1] 1789.309

> p.value <- pf(F.stat, p, df - p + 1, lower.tail=F)
> p.value
[1] 5.380012e-08
```

The small p-value provides strong evidence that the two species are indeed dissimilar with respect to their average measurements. As an exercise, investigate this difference further.

10.6 Further Two Sample Tests

The test procedure of the previous section can be generalized still further.

Suppose that independent random samples of sizes n_1 and n_2 are drawn from the $N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$ and $N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$ distributions, respectively, and that we wish to test

$$H_0 : \mathbf{C}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = \boldsymbol{\phi}$$

versus

$$H_1 : \mathbf{C}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \neq \boldsymbol{\phi},$$

where $\mathbf{C}$ is an $m \times p$ matrix of constants, of rank m, and $\boldsymbol{\phi}$ is a $m \times 1$ vector of constants.

It is easy to see that

$$\mathbf{C}(\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2) \sim N_m\left(\mathbf{C}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2), \left(\frac{1}{n_1} + \frac{1}{n_2}\right)\mathbf{C}\boldsymbol{\Sigma}\mathbf{C}^T\right).$$

Also

$$(n_1 + n_2 - 2)\mathbf{C}\mathbf{S}_U\mathbf{C}^T \sim W_m(n_1 + n_2 - 2, \mathbf{C}\boldsymbol{\Sigma}\mathbf{C}^T),$$

independently, where $\mathbf{S}_U$ is the pooled *unbiased* sample covariance matrix.

Now, under H_0,

$$\mathbf{C}(\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2) \sim N_m\left(\boldsymbol{\phi}, \frac{n_1 + n_2}{n_1 n_2}\mathbf{C}\boldsymbol{\Sigma}\mathbf{C}^T\right).$$

Thus, it is not difficult to show, using Theorem 9.4(result (iv) in Section 10.1), that

$$T^2 = \frac{n_1 n_2}{n_1 + n_2}\left[\mathbf{C}(\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2) - \boldsymbol{\phi}\right]^T (\mathbf{C}\mathbf{S}_U\mathbf{C})^{-1}\left[\mathbf{C}(\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2) - \boldsymbol{\phi}\right] \sim T_m^2(n_1 + n_2 - 2)$$

and

$$\left[\frac{n_1 + n_2 - m - 1}{(n_1 + n_2 - 2)m}\right] T^2 \sim F_{m, n_1 + n_2 - m - 1}.$$

So we reject $H_0 : \mathbf{C}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = \boldsymbol{\phi}$ at the $100\alpha\%$ level of significance if

$$\left[\frac{n_1 + n_2 - m - 1}{(n_1 + n_2 - 2)\,m}\right] T_{obs}^2 > F_{m, n_1 + n_2 - m - 1}(\alpha).$$