9 Multivariate Hypothesis Testing and Further Distributional Theory

9.1 Introduction

We consider now how to approach the construction of multivariate hypothesis tests. There are two fundamental problems. First the large number of possible hypotheses that exist in the multivariate context, and secondly the difficulty in choosing between various plausible test statistics. We shall look at two general approaches to the development of suitable test statistics, the union intersection test (UIT), and the generalized likelihood ratio test (LRT). We shall consider a range of possible hypothesis tests, some of which are obvious generalizations of univariate hypothesis tests, and others that have no univariate equivalent. In some cases UIT and LRT lead to the same test statistic, but in others they do not.

All the multivariate hypothesis tests considered in this course are based on the assumption of multivariate normality. To carry out the corresponding hypothesis tests we need to introduce further distributions that are based on the multivariate normal distribution. These are the Wishart distribution (the multivariate generalization of the χ^2 distribution) and the Hotelling T^2 distribution (generalizing the t and F distributions).

9.2 One-sample Test for the Mean Vector

Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be a random sample from $\mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are unknown. That is we have the $n \times p$ data matrix

$$X = \begin{pmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \vdots \\ \mathbf{x}_n^T \end{pmatrix}$$

We wish to test

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

Let $\overline{\mathbf{x}}$ and $\mathbf{S}$ be the MLE's of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ respectively. Thus

$$\overline{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i \quad \text{and} \quad \mathbf{S} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})^T.$$

We shall represent the usual *unbiased* estimate of $\mathbf{\Sigma}$ by

$$\mathbf{S}_U = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})^T.$$

9.2.1 UIT Approach

Consider, in the first instance, the univariate case, where $x_1, \ldots, x_n$ is a random sample of n observations drawn from the normal population $N(\mu, \sigma^2)$, where μ and σ^2 are unknown. We wish to test

$$H_0 : \mu = \mu_0 \text{ vs } H_1 : \mu \neq \mu_0.$$

In this *univariate* case, the test statistic is

$$t = \frac{(\overline{x} - \mu_0)}{s/\sqrt{n}} \sim \mathrm{t}(n-1) \quad \text{under } H_0.$$

where $\overline{x} = \dfrac{1}{n}\sum_{i=1}^{n} x_i$ and $s^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$, and the decision rule is

$$\begin{cases} \text{Reject } H_0 & \text{if } |t| > t_{\alpha/2}(n-1) \\ \text{Fail to reject (`accept') } H_0 & \text{if } |t| \leq t_{\alpha/2}(n-1) \end{cases}$$

Equivalently,

$$\begin{cases} \text{Reject } H_0 & \text{if } t^2 > t^2_{\alpha/2}(n-1) \\ \text{Fail to reject (`accept') } H_0 & \text{if } t^2 \leq t^2_{\alpha/2}(n-1) \end{cases} \tag{1}$$

where

$$t^2 = \frac{n(\overline{x} - \mu_0)^2}{s^2} \sim \mathrm{F}(1, n-1)$$

Note: s^2 is the <u>unbiased</u> estimate of σ^2 using divisor $(n-1)$. Let,

$$s_b^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{n-1}{n}s^2$$

so that s_b^2 is the maximum likelihood estimate (MLE) of σ^2.

Then, in terms of s_b^2,

$$t^2 = \frac{(n-1)(\overline{x} - \mu_0)^2}{s_b^2}. \tag{2}$$

Can we use our univariate test to guide the multivariate one?

If we form $y = \mathbf{a}^T\mathbf{x}$, an arbitrary linear compound of $\mathbf{x}$, then $y \sim \mathrm{N}(\mathbf{a}^T\boldsymbol{\mu}, \mathbf{a}^T\boldsymbol{\Sigma}\mathbf{a})$, and we can test

$$H_{0\mathbf{a}} : \mathbf{a}^T\boldsymbol{\mu} = \mathbf{a}^T\boldsymbol{\mu}_0 \text{ vs } H_{1\mathbf{a}} : \mathbf{a}^T\boldsymbol{\mu} \neq \mathbf{a}^T\boldsymbol{\mu}_0$$

using the statistic

$$t(\mathbf{a}) = \frac{\sqrt{n-1}\mathbf{a}^T(\overline{\mathbf{x}} - \boldsymbol{\mu}_0)}{\sqrt{\mathbf{a}^T\mathbf{S}\mathbf{a}}}$$

where the 'acceptance region' (from (1) and (2)) has the form

$$t^2(\mathbf{a}) = \frac{(n-1)[\mathbf{a}^T(\overline{\mathbf{x}} - \boldsymbol{\mu}_0)]^2}{\mathbf{a}^T\mathbf{S}\mathbf{a}} \leq c^2$$

Now, the original multivariate hypothesis H_0 is true if and only if $\mathbf{a}^T\boldsymbol{\mu} = \mathbf{a}^T\boldsymbol{\mu}_0$ holds for all non-null $\mathbf{a}$. i.e. $H_0 = \bigcap_{\mathbf{a}} H_{0\mathbf{a}}$.

This implies that acceptance of H_0 is equivalent to accepting all $H_{0\mathbf{a}}$. Thus, the *multivariate acceptance region* has the form

$$\bigcap_{\mathbf{a}}\{t^2(\mathbf{a}) \leq c^2\}.$$

This can be rewritten as

$$\{\max_{\mathbf{a}} t^2(\mathbf{a}) \leq c^2\}.$$

Thus, the test statistic we require has the form

$$\max_{\mathbf{a}} \frac{(n-1)[\mathbf{a}^T(\overline{\mathbf{x}} - \boldsymbol{\mu}_0)]^2}{\mathbf{a}^T\mathbf{S}\mathbf{a}}$$

Since $t^2(\mathbf{a})$ is dimensionless and unaffected by a change of scale of the elements of $\mathbf{a}$ we can remove this indeterminacy by imposing the constraint

$$\mathbf{a}^T\mathbf{S}\mathbf{a} = 1.$$

Thus, we require to

$$\text{Maximize} \quad (n-1)[\mathbf{a}^T(\overline{\mathbf{x}} - \boldsymbol{\mu}_0)]^2 \quad \text{subject to} \quad \mathbf{a}^T\mathbf{S}\mathbf{a} = 1, \tag{3}$$

or equivalently

$$\text{Maximize} \quad (n-1)\left[\mathbf{a}^T(\overline{\mathbf{x}}-\boldsymbol{\mu}_0)\right]\left[(\overline{\mathbf{x}}-\boldsymbol{\mu}_0)^T\mathbf{a}\right] \quad \text{subject to} \quad \mathbf{a}^T\mathbf{S}\mathbf{a}=1. \tag{4}$$

Forming the Lagrangian, we have

$$L(\mathbf{a},\lambda) \;=\; (n-1)\left[\mathbf{a}^T(\overline{\mathbf{x}}-\boldsymbol{\mu}_0)(\overline{\mathbf{x}}-\boldsymbol{\mu}_0)^T\mathbf{a}\right] \;-\; \lambda(\mathbf{a}^T\mathbf{S}\mathbf{a}-1)$$

so that,

$$\frac{\partial L}{\partial \mathbf{a}} = 2(n-1)(\overline{\mathbf{x}}-\boldsymbol{\mu}_0)(\overline{\mathbf{x}}-\boldsymbol{\mu}_0)^T\mathbf{a} - 2\lambda\mathbf{S}\mathbf{a}$$

and equating to zero gives

$$[(n-1)(\overline{\mathbf{x}}-\boldsymbol{\mu}_0)(\overline{\mathbf{x}}-\boldsymbol{\mu}_0)^T - \lambda\mathbf{S}]\mathbf{a} \;=\; 0 \tag{5}$$

for the maximizing $\mathbf{a}$. Equation (5) implies that λ is the only non-zero root of the rank 1 matrix

$$(n-1)\mathbf{S}^{-1}(\overline{\mathbf{x}}-\boldsymbol{\mu}_0)(\overline{\mathbf{x}}-\boldsymbol{\mu}_0)^T$$

Clearly,

$$\begin{aligned}
\lambda &= \mathrm{tr}\left\{(n-1)\mathbf{S}^{-1}(\overline{\mathbf{x}}-\boldsymbol{\mu}_0)(\overline{\mathbf{x}}-\boldsymbol{\mu}_0)^T\right\} \\
&= \mathrm{tr}\left\{(n-1)(\overline{\mathbf{x}}-\boldsymbol{\mu}_0)^T\mathbf{S}^{-1}(\overline{\mathbf{x}}-\boldsymbol{\mu}_0)\right\}
\end{aligned}$$

$$\Rightarrow \lambda = (n-1)(\overline{\mathbf{x}}-\boldsymbol{\mu}_0)^T\mathbf{S}^{-1}(\overline{\mathbf{x}}-\boldsymbol{\mu}_0) \tag{6}$$

Also, premultiplying (5) by $\mathbf{a}^T$ we have

$$\begin{aligned}
\lambda &= \frac{(n-1)\mathbf{a}^T(\overline{\mathbf{x}}-\boldsymbol{\mu}_0)(\overline{\mathbf{x}}-\boldsymbol{\mu}_0)^T\mathbf{a}}{\mathbf{a}^T\mathbf{S}\mathbf{a}} \\
&= \frac{(n-1)[\mathbf{a}^T(\overline{\mathbf{x}}-\boldsymbol{\mu}_0)]^2}{\mathbf{a}^T\mathbf{S}\mathbf{a}} \qquad (7) \\
&= t^2(\mathbf{a}) \quad \text{for the maximizing } \mathbf{a}.
\end{aligned}$$

Thus from (6) and (7) we have that

$$\begin{aligned}
\max_{\mathbf{a}} t^2(\mathbf{a}) &= \lambda \\
&= (n-1)(\overline{\mathbf{x}} - \boldsymbol{\mu}_0)^T \mathbf{S}^{-1}(\overline{\mathbf{x}} - \boldsymbol{\mu}_0) \\
&= T^2
\end{aligned} \tag{8}$$

and hence the 'acceptance' region has the form $\{T^2 \leq c^2\}$. The form of T^2 is analogous to the univariate

$$t^2 = \frac{(n-1)(\overline{x} - \mu_0)^2}{s_b^2} = \frac{n(\overline{x} - \mu_0)^2}{s^2}$$

and reduces to t^2 when $p = 1$. (Under H_0, T^2 has a Hotelling T^2 distribution - see later).

[It is easily verified that the maximized $t^2(\mathbf{a})$ is achieved for $\mathbf{a}^* = \mathbf{S}^{-1}(\overline{\mathbf{x}} - \boldsymbol{\mu})$. Do as an exercise.]

9.2.2 Generalized Likelihood Ratio Test (LRT) Construction

Recall, from equation (17) in Chapter 7, that given the data matrix $\mathbf{X}$, the log-likelihood function for unknown $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ is

$$\ell(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \mathbf{X}) = -\frac{np}{2}\ln(2\pi) - \frac{n}{2}\ln|\boldsymbol{\Sigma}| - \frac{n}{2}\mathrm{tr}(\boldsymbol{\Sigma}^{-1}S) - \frac{n}{2}(\overline{\mathbf{x}} - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\overline{\mathbf{x}} - \boldsymbol{\mu}). \tag{9}$$

Evaluating this expression at the maximum likelihood estimates $\widehat{\boldsymbol{\Sigma}} = S$ and $\hat{\boldsymbol{\mu}} = \overline{\mathbf{x}}$ we obtain

$$\ell^*(\overline{\mathbf{x}}, \mathbf{S}; \mathbf{X}) = -\frac{n}{2}\{p(1 + \ln(2\pi)) + \ln|\mathbf{S}|\}.$$

We are required to test

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

So under H_1, the maximum likelihood estimates are $\widehat{\boldsymbol{\Sigma}} = S$ and $\hat{\boldsymbol{\mu}} = \overline{\mathbf{x}}$, with

$$\ell_1^* = -\frac{n}{2}\{p(1 + \ln(2\pi)) + \ln|\mathbf{S}|\}.$$

Under H_0 it can easily be shown that the m.l.e. of $\mathbf{\Sigma}$ is given by

$$\mathbf{S} + (\overline{\mathbf{x}} - \boldsymbol{\mu}_0)(\overline{\mathbf{x}} - \boldsymbol{\mu}_0)^T. \tag{10}$$

[See Mardia, Kent and Bibby, p.125], and of course $\hat{\boldsymbol{\mu}} = \boldsymbol{\mu}_0$.

Using these estimates and equation (9) it can be shown (see Mardia, Kent and Bibby) that

$$\ell_0^* = -\frac{n}{2}\left[p(1 + \ln(2\pi)) + \ln|\mathbf{S}| + \ln(1 + (\mathbf{x} - \boldsymbol{\mu}_0)^T \mathbf{S}^{-1}(\overline{\mathbf{x}} - \boldsymbol{\mu}_0))\right].$$

Now, the likelihood ratio is given by

$$\lambda \;=\; L_0^*/L_1^*,$$

so that

$$-2\ln(\lambda) \;=\; 2(\ell_1^* - \ell_0^*) \;=\; n\,\ln[1 + (\overline{\mathbf{x}} - \boldsymbol{\mu}_0)^T \mathbf{S}^{-1}(\overline{\mathbf{x}} - \boldsymbol{\mu}_0)].$$

Thus, the LRT statistic also depends on

$$T^2 = (n-1)(\overline{\mathbf{x}} - \boldsymbol{\mu}_0)^T \mathbf{S}^{-1}(\overline{\mathbf{x}} - \boldsymbol{\mu}_0) \tag{11}$$

and again large values of T^2 are evidence against H_0.

9.3 Further Distributional Theory related to $N_p(\boldsymbol{\mu}, \mathbf{\Sigma})$

9.3.1 Wishart Distribution

This distribution is the multivariate analogue of the χ^2 distribution, or more precisely of the $\sigma^2\chi^2$ distribution. Suppose that $\mathbf{X}_r \sim N_p(\mathbf{0}, \mathbf{\Sigma}),\ r = 1, \ldots, f$, are mutually independent where $\mathbf{\Sigma}$ is of full rank. Then the random $p \times p$ matrix

$$\mathbf{W} = \sum_{r=1}^{f} \mathbf{X}_r \mathbf{X}_r^T$$

is said to have a CENTRAL WISHART distribution on f degrees of freedom, and is written as $\mathbf{W} \sim \mathbf{W}_p(f, \mathbf{\Sigma})$.

We can also define the NON-CENTRAL WISHART distribution. If $\mathbf{X}_r \sim N_p(\boldsymbol{\mu}_r, \mathbf{\Sigma}),\ r = 1, \ldots, f$, are mutually independent. Then the random matrix

$$\mathbf{W} = \sum_{r=1}^{f} \mathbf{X}_r \mathbf{X}_r^T$$

is said to have a NON-CENTRAL WISHART distribution on f degrees of freedom, and is written as $\mathbf{W} \sim W_p(f, \mathbf{\Sigma}; \mathbf{M})$, where $\mathbf{M}$ is the $f \times p$ matrix given by $\mathbf{M} = \left[\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_f\right]^T$.

Remarks 9.1

(i) For $p=1$, $\mu_r=0$, $r=1,\ldots,f$, $\mathbf{\Sigma}=\sigma^2$, say, then $X_r \sim NID(0,\sigma^2)$, and so, by definition

$$W=\sum_{r=1}^{f} X_r^2 \sim W_1(f,\sigma^2).$$

On the other hand, $\dfrac{X_r}{\sigma} \sim NID(0,1)$, which implies that

$$\frac{1}{\sigma^2}\sum_{r=1}^{f} X_r^2 \sim \chi_f^2 \;\Rightarrow\; W=\sum_{r=1}^{f} X_r^2 \sim \sigma^2\chi_f^2.$$

(ii) If $\mathbf{W} \sim W_p(f,\mathbf{\Sigma};\mathbf{M})$, then

$$E[\mathbf{W}] = f\mathbf{\Sigma} + \mathbf{M}^T\mathbf{M}$$

Proof
From the results of Chapter 7 it is immediate that

$$E[\mathbf{X}_r\mathbf{X}_r^T] = \mathbf{\Sigma} + \boldsymbol{\mu}_r\boldsymbol{\mu}_r^T.$$

Hence

$$E[\mathbf{W}] = \sum_{r=1}^{f} E[\mathbf{X}_r\mathbf{X}_r^T] = \sum_{r=1}^{f}\left\{\mathbf{\Sigma} + \boldsymbol{\mu}_r\boldsymbol{\mu}_r^T\right\}$$
$$= f\mathbf{\Sigma} + \sum_{r=1}^{f}\boldsymbol{\mu}_r\boldsymbol{\mu}_r^T = f\mathbf{\Sigma} + \mathbf{M}^T\mathbf{M}$$

□

(iii) Suppose $\mathbf{W}_1 \sim W_p(f_1,\mathbf{\Sigma};\mathbf{M}_1)$ and $\mathbf{W}_2 \sim W_p(f_2,\mathbf{\Sigma};\mathbf{M}_2)$ independently.
Then

$$\mathbf{W}_1+\mathbf{W}_2 \sim W_p(f_1+f_2,\mathbf{\Sigma};\mathbf{M})$$

where $\mathbf{M} = \begin{bmatrix}\mathbf{M}_1\\ \mathbf{M}_2\end{bmatrix}$.

(iv) Suppose $W \sim W_p(f,\mathbf{\Sigma};\mathbf{M})$ and $\mathbf{C}$ is a $q\times p$ matrix of constants.
Then

$$\mathbf{C}\mathbf{W}\mathbf{C}^T \sim W_q(f,\mathbf{C}\mathbf{\Sigma}\mathbf{C}^T;\mathbf{M}\mathbf{C}^{\mathbf{T}}).$$

Proposition 9.2 (Distribution of sample covariance matrix)
Suppose $\mathbf{X}_r \sim N_p(\boldsymbol{\mu},\mathbf{\Sigma})$, $r=1,\ldots,n$, are mutually independent. Then

$$n\mathbf{S} \sim W_p(n-1,\mathbf{\Sigma}),$$

where $\mathbf{S}$ is the *maximum likelihood* estimate of $\mathbf{\Sigma}$. Correspondingly

$$(n-1)\mathbf{S}_U \sim W_p(n-1,\mathbf{\Sigma}).$$

Proof
From Property 4 of the estimators $\mathbf{S}$ and $\mathbf{S}_U$ given in Chapter 7 we have that the corresponding SSP matrix A, where

$$A = \sum_{r=1}^{n}(\mathbf{X}_r - \overline{\mathbf{X}})(\mathbf{X}_r - \overline{\mathbf{X}})^T,$$

can be written as

$$A = \sum_{i=1}^{n-1} \mathbf{Z}_i \mathbf{Z}_i^T,$$

where the $\mathbf{Z}_i$ are iid $N_p(\mathbf{0}, \boldsymbol{\Sigma})$. Thus, by definition,

$$A \sim W_p(n-1, \boldsymbol{\Sigma}),$$

from which the results follow.

9.3.2 Hotelling T^2–distribution

This *univariate* distribution is a multivariate extension of the Student-t and F distributions.

Definition 9.3 (Hotelling T^2)
If $\mathbf{Y} \sim N_p(\mathbf{0}, \mathbf{I})$ and *independently* $\mathbf{W} \sim W_p(f, \mathbf{I})$, then

$$T^2 = f\mathbf{Y}^T\mathbf{W}^{-1}\mathbf{Y}$$

is said to have a Hotelling T^2-distribution. We write $T^2 \sim T_p^2(f)$.

Theorem 9.4 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

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

Proof
Let $\mathbf{X} = \boldsymbol{\Sigma}^{-\frac{1}{2}}(\mathbf{Y} - \boldsymbol{\mu})$ and $\mathbf{W} = \boldsymbol{\Sigma}^{-\frac{1}{2}}\mathbf{A}\boldsymbol{\Sigma}^{-\frac{1}{2}}$, so that,

$$\mathbf{X} \sim N_p(\mathbf{0}, \mathbf{I}) \qquad (*)$$

By definition $\mathbf{A} = \sum_{i=1}^{m} \mathbf{Z}_i\mathbf{Z}_i^T$ where $\mathbf{Z}_i$ are iid $N_p(\mathbf{0}, \boldsymbol{\Sigma})$, so

$$\mathbf{W} = \sum_{i=1}^{m}(\boldsymbol{\Sigma}^{-\frac{1}{2}}\mathbf{Z}_i)(\boldsymbol{\Sigma}^{-\frac{1}{2}}\mathbf{Z}_i)^T$$

and

$$\boldsymbol{\Sigma}^{-\frac{1}{2}}\mathbf{Z}_i \sim N_p(\mathbf{0}, \mathbf{I})$$

$$\Rightarrow \quad \mathbf{W} \sim \mathrm{W}_p(m, \mathbf{I}) \qquad (**)$$

From $(*)$ and $(**)$ we have $m\mathbf{X}^T\mathbf{W}^{-1}\mathbf{X} \sim \mathrm{T}_p^2(m)$, by definition. This implies that

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

Note: If $p = 1$, then $Y \sim \mathrm{N}(\mu, \sigma^2)$ and $\dfrac{A}{\sigma^2} \sim \chi_m^2$ (so $E\left[\frac{1}{m}A\right] = \sigma^2$).

Then Theorem 9.4 tells that

$$\left(\frac{Y-\mu}{\sigma}\right)^2 \Big/ \frac{A}{\sigma^2 m} = \frac{(Y-\mu)^2}{A/m} \sim T_1^2(m) \ \ (= \ t^2(m) \ = F(1,m)).$$

We can now identify the distribution required to carry out the one-sample multivariate hypothesis test $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ vs $H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$. Recall that the test statistic, identified via both the UIT and LRT approaches to multivariate hypothesis testing, is

$$T^2 = (n-1)(\overline{\mathbf{X}}-\boldsymbol{\mu}_0)^T\mathbf{S}^{-1}(\overline{\mathbf{X}}-\boldsymbol{\mu}_0),$$

where large values are evidence against H_0.

Corollary 9.5 If $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are iid $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then

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

Proof
We have $\overline{\mathbf{X}} \sim N_p(\boldsymbol{\mu}, \frac{1}{n}\boldsymbol{\Sigma})$ and $n\mathbf{S} \sim \mathrm{W}_p(n-1, \boldsymbol{\Sigma})$.

Now, $\sqrt{n}\overline{\mathbf{X}} \sim N_p(\sqrt{n}\boldsymbol{\mu}, \boldsymbol{\Sigma})$, so that, from Theorem 9.4,

$$(n-1)[\sqrt{n}(\overline{\mathbf{X}}-\boldsymbol{\mu})^T n^{-1}\mathbf{S}^{-1}\sqrt{n}(\overline{\mathbf{X}}-\boldsymbol{\mu})] \sim \mathrm{T}_p^2(n-1)$$

i.e.

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

Theorem 9.6 If $m > (p-1)$, then

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

Proof: [See Mardia, Kent and Bibby, p74].

[**Note**: $\mathrm{T}_1^2(m) = \mathrm{F}(1,m) = \mathrm{t}^2(m)$.]

9.4 One-sample Test for the Mean Vector: an Example

Suppose we have observations $\mathbf{x}_1, \ldots, \mathbf{x}_n$, a random sample from $\mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The population parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are unknown. We wish to test

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

Under H_0

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

is an observation drawn from the distribution $\mathrm{T}_p^2(n-1)$, and a large value for T^2 is evidence against H_0. How large?

At significance level α, the rejection (critical) region for the test is given by

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

from Theorem 9.6.

Also note 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 the maximizing $\mathbf{a}$, is given by $\mathbf{a}^* = \mathbf{S}^{-1}(\overline{\mathbf{x}} - \boldsymbol{\mu}_0)$.

Example The following artificial data refer to a random sample of six approximately 2-year old boys from a high altitude region in Asia. The variables recorded on each member of the sample were Height (X_1), Chest Circumference (X_2), and Middle Upper Arm Circumference (X_3), measured in c.m.

Individual	Height (cm)	Chest circumference (cm)	MUAC (cm)
1	78	60.6	16.5
2	76	58.1	12.5
3	92	63.2	14.5
4	81	59.0	14.0
5	81	60.8	15.5
6	84	59.5	14.0

For lowland children of the same age in the same country, the height, chest and MUAC means are considered to be 87.8, 58.4 and 15.9 cm respectively. We wish to test the hypothesis that the highland boys have the same means.

Let $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3)'$ be a 3×1 vector, where $\mu_1 = E[X_1]$, $\mu_2 = E[X_2]$, $\mu_3 = E[X_3]$. We wish to test the hypothesis

$$H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0 = (87.8, 58.4, 15.9)^T.$$

We use R to carry out this test.

```
#One-sample Hotelling T2-test

> b.height <- c(78, 76, 92, 81, 81, 84)
> b.chest <- c(60.6, 58.1, 63.2, 59, 60.8, 59.5)
> b.muac <- c(16.5, 12.5, 14.5, 14, 15.5, 14)
> boys <- data.frame(b.height, b.chest, b.muac)
> boys
  b.height b.chest b.muac
1       78    60.6   16.5
2       76    58.1   12.5
3       92    63.2   14.5
4       81    59.0   14.0
5       81    60.8   15.5
6       84    59.5   14.0

> S <- var(boys)  #Note this is the unbiased estimate
> S
         b.height b.chest b.muac
b.height    31.60   8.040   0.50
b.chest      8.04   3.172   1.31
b.muac       0.50   1.310   1.90

> S.inv <- solve(S)
> S.inv
           b.height    b.chest    b.muac
b.height  0.1863001 -0.6318911  0.386646
b.chest  -0.6318911  2.5840072 -1.615318
b.muac    0.3866460 -1.6153179  1.538286

> n <- 6
> p <- 3

> m.boys <- apply(boys, 2, mean)
> m.boys
b.height  b.chest   b.muac
    82.0     60.2     14.5

> mu0 <- c(87.8, 58.4, 15.9)
> mu0
[1] 87.8 58.4 15.9

> T2 <- n * (t(m.boys - mu0) %*% S.inv %*% (m.boys - mu0))
> T2
         [,1]
[1,] 271.6115
```

```
> T2 <- drop(T2)
> T2
[1] 271.6115

> Fobs <- (n - p)/((n - 1) * p) * T2
> Fobs
[1] 54.32229

> prob <- pf(Fobs, p, (n - p), lower.tail = F)
> prob
[1] 0.004103264
```

Thus H_0 is emphatically rejected. The mean vector for highland boys is not the same as the mean vector for lowland children.

It might occur to some researcher that, instead of carrying out the multivariate test, we might just carry out three separate univariate t-tests, one for each of the variables X_1, X_2 and X_3. e.g.

$$H_0 : \mu_1 = \mu_{10} = 87.8$$
$$H_0 : \mu_2 = \mu_{20} = 58.4$$
$$H_0 : \mu_3 = \mu_{30} = 15.9$$

Recall the univariate t statistic

$$t = \frac{\overline{x} - \mu_0}{s/\sqrt{n}} \sim t(n-1) \quad \text{under } H_0.$$

These can be calculated in a variety of ways in R. One such way is given below.

```
#Univariate one sample t-tests
> m.boys - mu0
b.height  b.chest   b.muac
    -5.8      1.8     -1.4

> sds <- apply(boys, 2, sd)
> sds
b.height  b.chest   b.muac
5.621388 1.781011 1.378405

> diagse <- diag(sds/sqrt(n))
> diagse
         [,1]      [,2]      [,3]
[1,] 2.294922 0.0000000 0.0000000
[2,] 0.000000 0.7270947 0.0000000
[3,] 0.000000 0.0000000 0.5627314

> invdiagse <- solve(diagse)
> invdiagse
          [,1]     [,2]     [,3]
[1,] 0.4357447 0.000000 0.000000
[2,] 0.0000000 1.375337 0.000000
[3,] 0.0000000 0.000000 1.777047

> univtstats <- invdiagse %*% (m.boys - mu0)
```

```
> univtstats
          [,1]
[1,] -2.527319
[2,]  2.475606
[3,] -2.487865

> qt(0.025, (n - 1))
[1] -2.570582

> probs <- 2 * pt(abs(univtstats), (n - 1), lower.tail = F)
> probs
           [,1]
[1,] 0.05270197
[2,] 0.05614275
[3,] 0.05530549
```

None of the three hypotheses can be rejected at the 5% level. The multivariate result and the univariate results seem to contradict each other. Examining the sample correlations and the mean differences we can see why.

```
> m.boys - mu0
b.height  b.chest   b.muac
    -5.8      1.8     -1.4

> cor(boys)
            b.height   b.chest      b.muac
b.height 1.00000000 0.8030562 0.06452822
b.chest  0.80305616 1.0000000 0.53361485
b.muac   0.06452822 0.5336148 1.00000000
```

For each individual measurement the difference is not significant, but the sample correlation between the height and chest circumference, 0.803, tells us that these two measurements are strongly positively correlated. However, the sample has a *smaller* mean measurement for height ($\overline{x}_1 - \mu_{01} = -5.8$) and a *larger* mean measurement for chest circumference ($\overline{x}_2 - \mu_{02} = 1.8$). Similarly for chest circumference and MUAC. These differences go against the correlation structure and are sufficiently large to make it highly unlikely that the sample of boys comes from the same population as the lowland children.

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

```
> S.inv %*% (m.boys - mu0)
              [,1]
b.height -2.759249
b.chest  10.577627
b.muac   -7.303719

> cor(boys)
           b.height   b.chest     b.muac
b.height 1.00000000 0.8030562 0.06452822
b.chest  0.80305616 1.0000000 0.53361485
b.muac   0.06452822 0.5336148 1.00000000
```

```
> m.boys - mu0
 b.height b.chest b.muac 
b.height  b.chest   b.muac 
    -5.8      1.8     -1.4 
```