MSc I: Statistical Analysis (Spring Term)
Multivariate Analysis

Solution to Exercise 3

1. (a) Union-Intersection Test (U.I.T.) Construction

Let $y = \mathbf{a}^T\mathbf{x}$, 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

$$H_{0\mathbf{a}} : \mathbf{a}^T\boldsymbol{\mu} = \mathbf{a}^T\boldsymbol{\mu}_0 \text{ v's } H_{1\mathbf{a}} : \mathbf{a}^T\boldsymbol{\mu} \neq \mathbf{a}^T\boldsymbol{\mu}_0$$

using the statistic

$$t(\mathbf{a}) = \frac{\sqrt{n}\mathbf{a}^T(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)}{\sqrt{\mathbf{a}^T\mathbf{S}\mathbf{a}}} \qquad \text{[c/f the univariate statistic } t = \frac{\overline{x} - \mu}{s/\sqrt{n}}\text{].}$$

where the acceptance region has the form

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

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. $H_0 = \cap H_{0\mathbf{a}}$.

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

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

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 maximize

$$n[\mathbf{a}^T(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)]^2 \tag{1}$$

subject to

$$\mathbf{a}^T\mathbf{S}\mathbf{a} = 1 \tag{2}$$

Forming the Lagrangian, we have

$$\begin{aligned} L(\mathbf{a}, \lambda) &= n[\mathbf{a}^T(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)]^2 - \lambda(\mathbf{a}^T\mathbf{S}\mathbf{a} - 1) \\ &= n\mathbf{a}^T(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)^T\mathbf{a} - \lambda(\mathbf{a}^T\mathbf{S}\mathbf{a} - 1) \end{aligned}$$

so that,

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

and equating to zero gives

$$[n(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)^T - \lambda\mathbf{S}]\mathbf{a} = 0 \tag{3}$$

which implies that λ is the only non-zero root of the rank 1 matrix

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

Clearly,

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

$$\Rightarrow \lambda = n(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)^T\mathbf{S}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0) \tag{4}$$

Also, premultiplying (3) by $\mathbf{a}^T$ we have

$$\begin{aligned} \lambda &= \frac{n\mathbf{a}^T(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)^T\mathbf{a}}{\mathbf{a}^T\mathbf{S}\mathbf{a}} \\ &= \frac{n[\mathbf{a}^T(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)]^2}{\mathbf{a}^T\mathbf{S}\mathbf{a}} \\ &= t^2(\mathbf{a}) \end{aligned} \tag{5}$$

Thus from (4) and (5) we have that

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

and hence the acceptance region has the form $\{T^2 \leq c^2\}$. That is, the U.I.T. procedure leads to Hotelling's T^2 statistic, as required. (Under H_0, T^2 has a Hotelling T^2 distribution).

(b) (i) The two-sample Hotelling T^2 statistic is

$$\begin{aligned}T^2 &= (\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2)^T \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1} \mathbf{S}^{-1}(\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2) \\ &= \frac{n_1 n_2}{n_1 + n_2}(\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2)^T \mathbf{S}^{-1}(\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2)\end{aligned}$$

which under the null hypothesis of no difference in the population mean vectors has a $T_p^2(n_1 + n_2 - 2)$ distribution.

Now,

$$\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2 = \begin{bmatrix} -1-4 \\ -2-5 \end{bmatrix} = \begin{bmatrix} -5 \\ -7 \end{bmatrix} \quad \text{and} \quad \mathbf{S}^{-1} = \frac{1}{13(22) - 6^2}\begin{bmatrix} 22 & -6 \\ -6 & 13 \end{bmatrix} = \frac{1}{250}\begin{bmatrix} 22 & -6 \\ -6 & 13 \end{bmatrix}$$

so that,

$$\begin{aligned}T^2 &= \frac{11 \times 12}{11 + 12} \times \frac{1}{250}\begin{bmatrix} -5, & -7 \end{bmatrix}\begin{bmatrix} 22 & -6 \\ -6 & 13 \end{bmatrix}\begin{bmatrix} -5 \\ -7 \end{bmatrix} \\ &= \frac{132}{23 \times 250}\begin{bmatrix} -68, & -61 \end{bmatrix}\begin{bmatrix} -5 \\ -7 \end{bmatrix} \\ &= \frac{132}{5250} \times 767 = 17.6077\end{aligned}$$

which, under the null hypothesis, has a $\mathrm{T}_2^2(21)$ distribution.

Hence,

$$F = \frac{21 - 2 + 1}{21 \times 2} T^2 = \frac{20}{42} 17.6077 = 8.38$$

which should be compared with an $F(2, 20)$ distribution. Looking at the tables, we see that the 0.005 (0.5%) and 0.001 (0.1%) quantiles of $F(2, 20)$ are 6.986 and 9.953 respectively, so that $0.001 < p < 0.005$. That is, there is strong evidence that the two population mean vectors are not the same.

[Note the 0.05 (5%) quantile of the same distribution is 3.493].

(ii) Fisher's linear discriminant function is $\mathbf{L}^T\mathbf{x}$, where $\mathbf{L} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$, and the allocation rule (assuming equal misclassification costs and equal prior probabilities) is to allocate to population 1 if

$$\mathbf{L}^T\mathbf{x} - \frac{1}{2}\mathbf{L}^T(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) \geq 0$$

Otherwise allocate to population 2. Substituting our sample estimates, we have

$$\hat{\mathbf{L}} = \mathbf{S}^{-1}(\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2) = \frac{1}{250}\begin{bmatrix}-68\\-61\end{bmatrix}$$

(from above, part(i)), and the allocation rule is allocate to population 1 if

$$\frac{1}{250}\begin{bmatrix}-68, & -61\end{bmatrix}\mathbf{x} \geq \frac{1}{2} \times \frac{1}{250}\begin{bmatrix}-68, & -61\end{bmatrix}\begin{bmatrix}3\\3\end{bmatrix} = -\frac{387}{500} = -0.774.$$

Now, for the given observation,

$$\hat{\mathbf{L}}\mathbf{x} = \frac{1}{250}\begin{bmatrix}-68, & -61\end{bmatrix}\begin{bmatrix}2\\1\end{bmatrix} = -\frac{1}{250} \times 197 = -0.788 < -0.774$$

so we classify this case into population 2.