$$N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}} \, e^{\frac{-(\boldsymbol{X}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})}{2}}$$

General Classification Rules

Expected Cost of Misclassification (ECM)

$$\begin{aligned} ECM &= c(2|1)\, p_1 \int_{R_2} f_1(\boldsymbol{x})d\boldsymbol{x} \,+\, c(1|2)\, p_2 \int_{R_1} f_2(\boldsymbol{x})d\boldsymbol{x} \\ &= c(2|1)\, p_1\, p(2|1) + c(1|2)\, p_2\, p(1|2) \end{aligned} \tag{1}$$

Total Probability of Misclassifiaction (TPM)
TPM is equivalent to ECM with equal costs.

$$TPM = p_1\, p(2|1) + p_2\, p(1|2) \tag{2}$$

ECM Minimization Rule (general)

Classify observation $\boldsymbol{x}$ as π_1 if: $\frac{f_1(\boldsymbol{x})}{f_2(\boldsymbol{x})} > \frac{c(1|2)\, p_2}{c(2|1)\, p_1}$

Allocation by Posterior Probabilities
Allocate observation $\boldsymbol{x}$ to population (π_1 or π_2) with the largest $p(\pi_i|\boldsymbol{x})$
Classify observation $\boldsymbol{x}$ as π_1 if:

$$p(\pi_1|\boldsymbol{x}) > p(\pi_2|\boldsymbol{x}), \quad where \;\; p(\pi_i|\boldsymbol{x}) = \frac{p_i\, f_i(\boldsymbol{x})}{p_1\, f_1(\boldsymbol{x}) + p_2\, f_2(\boldsymbol{x})} \tag{3}$$

is equivalent to TPM.

Classification with MV Normal Populations ($\Sigma_1 = \Sigma_2 = \Sigma$. Linear)

$$\frac{f_1(\boldsymbol{x})}{f_2(\boldsymbol{x})} = exp[(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}^{-1}\boldsymbol{x} - \frac{1}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)] \tag{4}$$

By taking natural logarithm, we get the ECM minimization rule.
Classify observation $\boldsymbol{x}$ as π_1 if:

$$(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}^{-1}\boldsymbol{x} - \frac{1}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) \geq ln(\frac{c(1|2)\, p_2}{c(2|1)\, p_1}) \tag{5}$$

The sample version is obtained by replacing $\boldsymbol{\mu_i}$, $\boldsymbol{\Sigma}$ with $\bar{\boldsymbol{x}}_{\boldsymbol{i}}$, $\boldsymbol{S}_{pool} = \frac{n_1-1}{n_1+n_2-2}\boldsymbol{S}_1 + \frac{n_2-1}{n_1+n_2-2}\boldsymbol{S}_2$, respectively.

Fisher's Discriminant Function

Maximize $\frac{(\bar{y}_1 - \bar{y}_2)^2}{S_y^2}$ by choosing the linear combination $\bar{y}_i = a^T \bar{\boldsymbol{x}}_i$,
where $a^T = (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)^T \boldsymbol{S}_{pool}^{-1}$.
Classify observation $\boldsymbol{x}_0$ as π_1 if:

$$\begin{aligned} \hat{y}_0 = (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)^T \boldsymbol{S}_{pool}^{-1}\boldsymbol{x}_0 &> \frac{1}{2}[a^T \bar{\boldsymbol{x}}_1 + a^T \bar{\boldsymbol{x}}_2] = \frac{1}{2}a^T(\bar{\boldsymbol{x}}_1 + \bar{\boldsymbol{x}}_2) \\ &= \frac{1}{2}(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)^T \boldsymbol{S}_{pool}^{-1}(\bar{\boldsymbol{x}}_1 + \bar{\boldsymbol{x}}_2) \end{aligned} \tag{6}$$

equivalent to Equation (5) with $ln(\frac{c(1|2)\, p_2}{c(2|1)\, p_1}) = 0$ (equal costs, equal priors).

Classification with MV Normal Populations ($\Sigma_1 \neq \Sigma_2$. Quadratic)

Classify observation $\boldsymbol{x}$ as π_1 if:

$$\begin{aligned} \frac{-1}{2}\boldsymbol{x}^T(\boldsymbol{\Sigma}_1^{-1} - \boldsymbol{\Sigma}_2^{-1})\boldsymbol{x} + (\boldsymbol{\mu}_1^T\boldsymbol{\Sigma}_1^{-1} - \boldsymbol{\mu}_2^T\boldsymbol{\Sigma}_2^{-1})\boldsymbol{x} - k &> ln(\frac{c(1|2)\, p_2}{c(2|1)\, p_1}), \\ where \;\; k = \frac{1}{2}ln(\frac{|\boldsymbol{\Sigma}_1|}{|\boldsymbol{\Sigma}_2|}) + \frac{1}{2}(\boldsymbol{\mu}_1^T\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu}_1 &- \boldsymbol{\mu}_2^T\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\mu}_2) \end{aligned} \tag{7}$$

Evaluate Classification Functions

Actual Error Rate (AER)

$$AER = p_1 \int_{\hat{R}_2} f_1(\boldsymbol{x})d\boldsymbol{x} + p_2 \int_{\hat{R}_1} f_2(\boldsymbol{x})d\boldsymbol{x},$$

$$where \quad \hat{R}_1 : (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)^T \boldsymbol{S}_{pool}\boldsymbol{x} - \frac{1}{2}(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)^T \boldsymbol{S}_{pool}(\bar{\boldsymbol{x}}_1 + \bar{\boldsymbol{x}}_2) > ln(\frac{c(1|2)\,p_2}{c(2|1)\,p_1})$$

(8)

Apparent error rate (APER) is just the misclassification rate in the training set. $APER = \frac{n_{1M} + n_{2M}}{n_1 + n_2}$

Lachenbruch's Holdout Procedure

Lachenbruch's Holdout Procedure gives a nearly unbiased estimate of $E(ARE)$ for moderate sample size. The procedure is as follow:

1. Start with π_1 group. Omit 1 obs.(holdout) and then construct a classification function based on the remaining $n_1 - 1$ obs.
2. Classify the omitted obs. with the function constructed in step 1.
3. Repeat 1. and 2. untill all obs. in π_1 are classified. ($n_{1M}^{(H)}$: number of holdout obs. misclassified in π_1)
4. Repeat 1. to 3. for π_2 obs.

$$\hat{P}(2|1) = \frac{n_{1M}^{(H)}}{n_1}, \quad \hat{P}(1|2) = \frac{n_{2M}^{(H)}}{n_2}, \quad \hat{E}(AER) = \frac{n_{1M}^{(H)} + n_{2M}^{(H)}}{n_1 + n_2}$$

(9)

Clustering

Similarity Coefficients

	k = 1	k = 0	
i = 1	a	b	a + b
i = 0	c	d	c + d
	a + c	b + d	p (or n)

Item i, k measured by **p binary variables**. **Similarity between item i and item k** can be defined as $\frac{a+d}{p}$, $\frac{2(a+d)}{2(a+d)+b+c}$, $\frac{a}{p}$, etc.

Variable i, k with **n items**. **Similarity between variable i and variable k** can be defined as the **correlation** between the two variables:
$r = \frac{ad-bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$

Linkage Method (Hierarchical Clustering)

Clustering N objects with agglomerative hierachical clustering algorithm:

1. Start with N clusters and an $N \times N$ symmetric matrix of distances $\boldsymbol{D} = d_{ik}$.
2. Search for the nearest (most similar) pair of clusters in the distance matrix. Let the distance between these two clusters (U and V) be d_{UV}.
3. Merge cluster U and V into cluster (UV). Update the distance matrix by removing rows and columns corresponding to U and V, and add a row and a column giving the distances between (UV) and the remaining clusters.
4. Record the identity of the merged clusters and the distance at which the clusters merged. Repeat until all objects merged into 1 cluster.

Single Linkage
$d_{(UV)W} = min\{d_{UW}, d_{VW}\}$

Complete Linkage
$d_{(UV)W} = max\{d_{UW}, d_{VW}\}$

Average Linkage
$d_{(UV)W} = \frac{\sum_i \sum_k d_{ik}}{N_{(UV)} N_W}$

Ward's Method
For a given k, ESS_k is the sum of squared deviations of every item from the mean (centroid) in cluster k. $ESS = \sum_{k=1}^{K} ESS_k$. For every step, the two clusters whose combination results in the smallest increase in ESS are joined.

K-means Method (Nonhierarchical Clustering)

K-means Method can **only be used to group items**.

Algorithm:

1. Partition items into K initial clusters (randomly).
2. Proceed through the list of items. Assign an item to the cluster whose mean (centroid) is nearest. Recalculate the centroids for **the cluster receiving the item** and the **cluster losing the item**.
3. Repeat step 2 until ni more reassignment takes place.

Formulas for updating cluster centroids:

For the ith coordinate, i = 1, 2, ..., p,

$$\bar{x}_{i,new} = \frac{n\bar{x}_i + x_{ji}}{n+1} \qquad \textit{if the jth item is added to this group}$$

$$\bar{x}_{i,new} = \frac{n\bar{x}_i - x_{ji}}{n-1} \qquad \textit{if the jth item is removed from this group}$$

(10)

Comapare 1 Population with a constant vector

Hotelling's T²

Let $\boldsymbol{X}_1, \boldsymbol{X}_2, ..., \boldsymbol{X}_n$ be i.i.d. from $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

$$T^2 = (\bar{\boldsymbol{X}} - \boldsymbol{\mu})^T (\frac{\boldsymbol{S}}{n})^{-1} (\bar{\boldsymbol{X}} - \boldsymbol{\mu})$$

$$= n(\bar{\boldsymbol{X}} - \boldsymbol{\mu})^T \boldsymbol{S} (\bar{\boldsymbol{X}} - \boldsymbol{\mu}) \quad \sim \quad \frac{(n-1)p}{n-p} F_{p,n-p}$$

(11)

, regardless of the true value of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.
$\bar{\boldsymbol{x}} = \frac{1}{n} \sum_{j=1}^{n} \boldsymbol{x}_j$, $\boldsymbol{S} = \frac{1}{n-1} \sum_{j=1}^{n} (\boldsymbol{x}_j - \bar{\boldsymbol{x}})(\boldsymbol{x}_j - \bar{\boldsymbol{x}})^T$.

Testing H₀: $\boldsymbol{\mu} = \boldsymbol{\mu_0}$

Calculate T² = $n(\bar{\boldsymbol{x}} - \boldsymbol{\mu}_0)^T \boldsymbol{S} (\bar{\boldsymbol{x}} - \boldsymbol{\mu}_0)$. Reject H₀ if T² > $\frac{(n-1)p}{n-p} F_{p,n-p}(\alpha)$.

Confidence Region (ellipsoid)

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

(12)

Axes of confidence ellipsoid

$$\pm \sqrt{\lambda_i} \sqrt{\frac{(n-1)p}{(n-p)n} F_{p,n-p}(\alpha)} \; \boldsymbol{e}_i \quad \boldsymbol{S}\boldsymbol{e}_i = \lambda_i \boldsymbol{e}_i, \; i = 1, 2, ..., p$$

(13)

Simultaneous Confidence Intervals

Let $\boldsymbol{X}_1, \boldsymbol{X}_2, ..., \boldsymbol{X}_n$ be i.i.d. from $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then **for all $\boldsymbol{a}$ simultaneously**, the interval between the values $\boldsymbol{a}^T \pm \sqrt{\frac{(n-1)p}{n-p} F_{p,n-p}(\alpha) \boldsymbol{a}^T \frac{\boldsymbol{S}}{n} \boldsymbol{a}}$ will contain $\boldsymbol{a}^T \boldsymbol{\mu}$ with probability $1 - \alpha$. e.g., for $\boldsymbol{a}^T = [1 \; 0 \; 0 \; \cdots \; 0]$,

$$\bar{x}_1 - \sqrt{\frac{(n-1)p}{n-p} F_{p,n-p}(\alpha) \frac{S_{11}}{n}} < \mu_1 < \bar{x}_1 + \sqrt{\frac{(n-1)p}{n-p} F_{p,n-p}(\alpha) \frac{S_{11}}{n}}$$

(14)

Bonferroni Method

$$\bar{x}_i - t_{n-1}(\frac{\alpha}{2m}) \sqrt{\frac{S_{ii}}{n}} < \mu_i < \bar{x}_i + t_{n-1}(\frac{\alpha}{2m}) \sqrt{\frac{S_{ii}}{n}} \qquad i = 1, 2, ..., m$$

(15)

Compare 2 Population Means

Paired Comaprisons (Multivariate paired t-test)

Let $\boldsymbol{D}_j = \begin{pmatrix} D_{j1} \\ D_{j2} \\ \vdots \\ D_{jp} \end{pmatrix} = \begin{pmatrix} X_{1j1} - X_{2j1} \\ X_{1j2} - X_{2j2} \\ \vdots \\ X_{1j3} - X_{2j3} \end{pmatrix}$. j means the jth subject, X_{2j3} is the response of the **jth subject** on **varable 3** in the **2nd treatment**.

D_{j2} is the **difference in the response** on the **2nd variable** between the 2 treatments.

Let $E(\boldsymbol{D_j}) = \boldsymbol{\delta}$, $Cov(\boldsymbol{D_j}) = \boldsymbol{\Sigma}_d$, and $\boldsymbol{D_1}, \boldsymbol{D_2}, ..., \boldsymbol{D_n}$ be i.i.d. from $N_p(\boldsymbol{\delta}, \boldsymbol{\Sigma}_d)$, then

$$T^2 = (\bar{\boldsymbol{D}} - \boldsymbol{\delta})^T (\frac{\boldsymbol{S_d}}{n})^{-1}(\bar{\boldsymbol{D}} - \boldsymbol{\delta})$$

$$= n(\bar{\boldsymbol{D}} - \boldsymbol{\delta})^T \boldsymbol{S_d}(\bar{\boldsymbol{D}} - \boldsymbol{\delta}) \quad \sim \quad \frac{(n-1)p}{n-p}F_{p,n-p} \tag{16}$$

If n, n-p are large, $T^2 \to \chi_p^2$

Simultaneous C.I. and Bonferroni intervals follows directly from Equation (14) & (15) by replacing $\boldsymbol{x}_i$ with $\boldsymbol{D}_i$.

Two Independent Populations ($\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}$)

Let $\bar{\boldsymbol{x}}_1 = \frac{1}{n_1}\sum_{j=1}^{n_1} \boldsymbol{x}_{1j}$, $\boldsymbol{S}_1 = \frac{1}{n_1-1}\sum_{j=1}^{n_1}(\boldsymbol{x}_{1j} - \bar{\boldsymbol{x}}_1)(\boldsymbol{x}_{1j} - \bar{\boldsymbol{x}}_1)^T$, where $\boldsymbol{x}_{11}, \boldsymbol{x}_{12}, \dots, \boldsymbol{x}_{1n_1} \sim N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$. Then, $\boldsymbol{S}_{pool}$ estimates $\boldsymbol{\Sigma}$ (see formulas for $\boldsymbol{S}_{pool}$ under eq (5)).

Test H₀: $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \boldsymbol{\delta}_0$

$$T^2 = (\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2 - \boldsymbol{\delta})^T [(\frac{1}{n_1} + \frac{1}{n_2})\boldsymbol{S}_{pool}]^{-1}(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2 - \boldsymbol{\delta}) \quad \sim \quad \frac{(n_1+n_2-2)p}{n_1+n_2-p-1}F_{p,n_1+n_2-p-1} \tag{17}$$

Simultaneous Intervals

Let $c^2 = \frac{(n_1+n_2-2)p}{n_1+n_2-p-1}F_{p,n_1+n_2-p-1}(\alpha)$.

$$\boldsymbol{a}^T(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2) \pm c\sqrt{\boldsymbol{a}^T(\frac{1}{n_1} + \frac{1}{n_2})\boldsymbol{S}_{pool}\boldsymbol{a}} \quad or \quad (\bar{X}_{1i} - \bar{X}_{2i}) \pm c\sqrt{(\frac{1}{n_1} + \frac{1}{n_2})S_{ii,pool}} \quad i = 1, 2, ..., p \tag{18}$$

The Bonferroni method is done by replacing c with $t_{n_1+n_2-2}(\frac{\alpha}{2p})$ in the second part of the equation above.

One-way MANOVA

$$\boldsymbol{X}_{\ell j} - \boldsymbol{\mu} = \boldsymbol{\tau}_\ell + \boldsymbol{e}_{\ell j}, \quad j = 1, 2, ..., n \quad \ell = 1, 2, ..., g$$

$$\sum_{\ell=1}^{g}\sum_{j=1}^{n_\ell}(\boldsymbol{x}_{\ell j} - \bar{\boldsymbol{x}})(\boldsymbol{x}_{\ell j} - \bar{\boldsymbol{x}})^T = \sum_{\ell=1}^{g}n_\ell(\bar{\boldsymbol{x}}_\ell - \bar{\boldsymbol{x}})(\bar{\boldsymbol{x}}_\ell - \bar{\boldsymbol{x}})^T + \sum_{\ell=1}^{g}\sum_{j=1}^{n_\ell}(\boldsymbol{x}_{\ell j} - \bar{\boldsymbol{x}}_\ell)(\boldsymbol{x}_{\ell j} - \bar{\boldsymbol{x}}_\ell)^T \tag{19}$$

Source	SSP (matrix)	d.f.
Treatment	B	g - 1
Residual	W	$\sum_{\ell=1}^{g} n_\ell - g$
Total	B + W	$\sum_{\ell=1}^{g} n_\ell - 1$

$\Lambda^* = \frac{|W|}{|B+W|}$

Simultaneous C.I. (Bonferroni)

Test $\tau_{ki} - \tau_{\ell i}$ (treatment k, ℓ on response variable i)

$$C.I.: \quad \bar{x}_{ki} - \bar{x}_{\ell i} \pm t_{n-g}(\frac{\alpha}{pg(g-1)})\sqrt{\frac{w_{ii}}{n-g}(\frac{1}{n_k} + \frac{1}{n_\ell})}, \quad i = 1, 2, ..., p \quad w_{ii} : ith\ diagonal\ element\ of\ W \tag{20}$$