EJERCICIO 1

a. Construir e interpretar la función discriminante

$$\vec{X}_1 \sim N_2\left[\begin{pmatrix}0\\0\end{pmatrix};\begin{pmatrix}2 & 1.6\\1.6 & 2\end{pmatrix}\right]$$

$$\vec{X}_2 \sim N_2\left[\begin{pmatrix}1\\1\end{pmatrix};\begin{pmatrix}2 & 1.6\\1.6 & 2\end{pmatrix}\right]$$

$$\rho_{12} = \frac{\sigma_{12}}{\sigma_1 * \sigma_2} \Rightarrow \sigma_{12} = 1.6$$

Dado que por letra ambas poblaciones son normales y con igual matriz de varianzas y covarianzas, se construirá un discriminante lineal.

$$f_{\vec{X}_i} = \frac{1}{2\pi * |\mathbf{\Sigma}|^{0.5}} * exp\left\{-\frac{1}{2}(\vec{x}-\vec{\mu}_i)'\mathbf{\Sigma}^{-1}(\vec{x}-\vec{\mu}_i)\right\}$$

La partición óptima es:

$$clasificar\ en\ la\ población\ P_2 \Leftrightarrow \frac{f_{\vec{X}_2}(\vec{x}) * \pi_2}{c(2|1)} > \frac{f_{\vec{X}_1}(\vec{x}) * \pi_1}{c(1|2)}$$

$$donde\ \pi_i\ es\ la\ probabilidad\ a\ priori\ de\ pertenecer\ a\ la\ población\ i\ y$$

$$c(i|j)\ es\ el\ costo\ de\ clasificar\ en\ el\ grupo\ i\ una\ observación\ proveniente\ de\ la\ población\ j$$

Dado que no se informan costos distintos, estos se asumen iguales, y dado que no se informan probabilidades a priori distintas para cada caso, estas se asumen iguales, por lo que la partición óptima es:

$$\boxed{clasificar\ en\ la\ población\ P_2 \Leftrightarrow f_{\vec{X}_2}(\vec{x}) * \pi_2 > f_{\vec{X}_1}(\vec{x}) * \pi_1}$$

Como ambos términos son siempre positivos, podemos tomar logaritmos, y dado que ambas poblaciones tienen igual matriz de varianzas y covarianzas, podemos concentrarnos únicamente en las distancias de Mahalanobis. Por lo tanto, la regla será:

$$clasificar\ en\ la\ población\ P_2 \Leftrightarrow \pi_1 * \mathcal{D}_1^2(\vec{x}) > \pi_2 * \mathcal{D}_2^2(\vec{x})$$

Por lo tanto, las funciones de discriminantes serán:

$$P\left(i\epsilon g|\vec{X}=\vec{x}\right) = Max\ \left[\pi_g * ex\,p\left\{-\frac{1}{2}D_{ig}^2\right\}\right]$$

Aplicando logaritmos

$$log(\pi_g) - \frac{1}{2}D_{ig}^2 = log(\pi_g) - \frac{1}{2}\left\{(\vec{x}_i-\vec{\mu}_g)'\mathbf{\Sigma}_g^{-1}(\vec{x}_i-\vec{\mu}_g)\right\} =$$

$$= log(\pi_g) - \frac{1}{2}\vec{x}_i'\mathbf{\Sigma}_g^{-1}\vec{x}_i + \frac{1}{2}\vec{x}_i'\mathbf{\Sigma}_g^{-1}\vec{\mu}_g + \frac{1}{2}\vec{\mu}_g'\mathbf{\Sigma}_g^{-1}\vec{x}_i - \frac{1}{2}\vec{\mu}_g'\mathbf{\Sigma}_g^{-1}\vec{\mu}_g$$

Si las probabilidades a priori, las medias, o las matrices de varianzas y covarianzas no fueran conocidas, habría que remplazarlas por sus estimaciones: $\pi;\ \vec{\mu};\ \Sigma\ \rightarrow\ \hat{\pi};\ \vec{\bar{x}}, \mathbf{S}$

El score de la observación i en la función discriminante lineal para el grupo g es:

$$L_{ig} = \log(\pi_g) + \vec{\bar{x}}_g' \mathbf{S}_g^{-1} \vec{x}_i - \frac{1}{2} \vec{\bar{x}}_g' \mathbf{S}_g^{-1} \vec{\bar{x}}_g$$

Asumiendo $\pi_i = 0.5 \quad con\ i = 1,2$

Para el grupo 1

$$L_{i1} = log(0.5) + (0\ 0) \begin{bmatrix} 2 & 1.6 \\ 1.6 & 2 \end{bmatrix}^{-1} \vec{x}_i - \frac{1}{2} (0\ 0) \begin{bmatrix} 2 & 1.6 \\ 1.6 & 2 \end{bmatrix}^{-1} \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Rightarrow$$

$$\Rightarrow \boxed{L_{i1} = \log(0.5) \cong -0.69}$$

Para el grupo 2

$$L_{i2} = log(0.5) + (1\ \ 1) \begin{bmatrix} 2 & 1.6 \\ 1.6 & 2 \end{bmatrix}^{-1} \vec{x}_i - \frac{1}{2} (1\ 1) \begin{bmatrix} 2 & 1.6 \\ 1.6 & 2 \end{bmatrix}^{-1} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \Rightarrow$$

$$\Rightarrow \boxed{L_{i2} \cong \begin{pmatrix} \frac{5}{18} & \frac{5}{18} \end{pmatrix} \vec{x}_i - 0.56}$$

b. Ahora las varianzas son 2 y 4

$$\vec{X}_1 \sim N_2 \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}; \begin{pmatrix} 2 & 2.26 \\ 2.26 & 4 \end{pmatrix} \right]$$

$$\vec{X}_2 \sim N_2 \left[\begin{pmatrix} 1 \\ 1 \end{pmatrix}; \begin{pmatrix} 2 & 2.26 \\ 2.26 & 4 \end{pmatrix} \right]$$

$$\rho_{12} = \frac{\sigma_{12}}{\sigma_1 * \sigma_2} \Rightarrow \sigma_{12} = 2.26$$

Siguiendo el mismo razonamiento obtenemos que:

$$\boxed{L_{i1} = \log(0.5) \cong -0.69}$$

$$\boxed{L_{i2} \cong (0.278\ \ 0.278)\ \vec{x}_i - 0.16}$$

c. Ahora $\pi_1 = 0.7$ y $\pi_2 = 0.3$

$$\boxed{L_{i1} = \log(0.7) \cong -0.36}$$

$$L_{i2} = log(0.3) + (1\ \ 1) \begin{bmatrix} 2 & 2.26 \\ 2.26 & 4 \end{bmatrix}^{-1} \vec{x}_i - \frac{1}{2} (1\ 1) \begin{bmatrix} 2 & 2.26 \\ 2.26 & 4 \end{bmatrix}^{-1} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \Rightarrow$$

$$\Rightarrow \boxed{L_{i2} \cong (1\ \ 1) \begin{bmatrix} 2 & 2.26 \\ 2.26 & 4 \end{bmatrix}^{-1} \vec{x}_i - 0.67}$$

EJERCICIO 2

1. ¿Qué indica el test de multinormalidad?

Test de Mardia de multinormalidad:

$$H_0)\ \vec{X} \sim N_p\left(\vec{\mu}_p;\ \mathbf{\Sigma}_p\right) \quad Versus \quad H_1)\ \vec{X} \nsim N_p\left(\vec{\mu}_p;\ \mathbf{\Sigma}_p\right)$$

Debido a que la prueba busca testear dos momentos (simetría y kurtosis) el test puede a su vez separase en dos partes:

$$SIMETRÍA\text{: } H_0)\ \beta_{1;p} = 0 \quad Versus \quad H_1)\ \beta_{1;p} \neq 0$$

$$\beta_{1;p} = E\left\{\left[(\vec{x}_i - \vec{\mu})'\mathbf{\Sigma}^{-1}\left(\vec{x}_j - \vec{\mu}\right)\right]^3\right\}$$

$$RC = \left\{muestras/\ \kappa_1 = \frac{n}{6} * \hat{\beta}_{1;p} > \chi^2_{\frac{p(p+1)(p+2)}{6}}(1-\alpha)\right\}$$

$$\hat{\beta}_{1;p} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[\left(\vec{x}_i - \bar{\vec{x}}\right)'\mathbf{S}^{-1}\left(\vec{x}_j - \bar{\vec{x}}\right)\right]^3$$

$$KURTOSIS\text{: } H_0)\ \beta_{2;p} = p(p+2) \quad Versus \quad H_1)\ \beta_{2;p} \neq p(p+2)$$

$$\beta_{2;p} = E\{[(\vec{x} - \vec{\mu})'\mathbf{\Sigma}^{-1}(\vec{x} - \vec{\mu})]^2\}$$

$$RC = \left\{muestras \Big/ \kappa_2 = \frac{\hat{\beta}_{2;p} - p(p+2)}{\sqrt{\frac{8p(p+2)}{n}}} > Z(1-\alpha)\right\}$$

$$\hat{\beta}_{2;p} = \frac{1}{n}\sum_{i=1}^{n}\left[\left(\vec{x}_i - \bar{\vec{x}}\right)'\mathbf{S}^{-1}\left(\vec{x}_i - \bar{\vec{x}}\right)\right]^2$$

Estos test funcionan de forma asintótica, aunque pueden utilizarse las tablas de Mardia para muestras pequeñas.

Para el grupo 0: con $p - valor\ = 0.055\ > 0.05 = \alpha$ no se rechaza H0 para la prueba de simetría, y con $p - valor\ = 0.078\ > 0.05 = \alpha$ no se rechaza H0 para la prueba de kurtosis. Por lo tanto, no existe evidencia estadística suficiente para afirmar que el grupo 0 no provenga de una distribución multinormal.

Para el grupo 1: con $p - valor\ = 0.152\ > 0.05 = \alpha$ no se rechaza H0 para la prueba de simetría, y con $p - valor\ = 0.111\ > 0.05 = \alpha$ no se rechaza H0 para la prueba de kurtosis. Por lo tanto, no existe evidencia estadística suficiente para afirmar que el grupo 1 no provenga de una distribución multinormal.

Dados los resultados del test de multinormalidad, es válida la utilización de un discriminante probabilístico basado en la distribución multinormal (lineal o cuadrático dependiendo del test de igualdad de matrices varianzas y covarianzas).

Para poder determinar si el discriminante a utilizar debe ser lineal o cuadrático, debe realizase el test de Box M de igualdad de matrices de varianzas y covarianzas.

$$H_0)\ \mathbf{\Sigma}_0 = \mathbf{\Sigma}_1 = \mathbf{\Sigma} \quad Versus \quad H_1)\ \mathbf{\Sigma}_0 \neq \mathbf{\Sigma}_1$$

2. La regla discriminante lineal es aplicable en caso donde los grupos provienen de poblaciones multinormales con igual forma de la matriz de varianzas y covarianzas.

3. El cuadro aporta la información sobre los valores de los coeficientes asociados a las distintas variables de la función discriminante. La misma puede ser utilizada para clasificar las observaciones entre grupo 0 (no quebró la empresa), y grupo 1 (quebró la empresa).

La función discriminante en el grupo 0 es:

$$L_{i0} = -7.39 + 5.36 * Fc.dt_i - 9.99 * In.At_i + 3.30 * Ac.pc_i + 10.07 * AC.VN_i$$

La función discriminante del grupo 1 es:

$$L_{i1} = -5.20 + 4.08 * Fc.dt_i - 18.59 * In.At_i + 1.62 * Ac.pc_i + 12.24 * AC.VN_i$$

4. En cross-validation lo que se busca es, dejar una observación de lado, calcular la función discriminante, y utilizarla para clasificar la observación que fue separada. El proceso se repite con cada observación. Para muestras extremadamente grandes, pueden tomarse grupos de observaciones. El error de cross validation es la proporción de observaciones que fueron mal clasificadas utilizando el método de cross validation.

El error del grupo 0 fue:

$$e_{0,c} = \frac{2}{25} = 0.08$$

Mientras que el error del grupo 1 fue:

$$e_{1,c} = \frac{4}{21} = 0.19$$

El error total es entonces:

$$e_c = \pi_0 * e_{0,c} + \pi_1 * e_{1,c} = \frac{25}{25+21} * 0.08 + \frac{21}{25+21} * 0.19 = \frac{3}{23} \cong 0.13$$

5.

6. Los datos de la nueva empresa deberían evaluarse en las funciones discriminantes de ambos grupos, y asignar la empresa al grupo con menor score.

EJERCICIO 3

Los resultados del test de Mardia[1] rechazan la presencia de multinormalidad en ambos grupos incluso con $\alpha = 0.01$. Por lo tanto, si se desea realizar un AD probabilístico, se debe realizar uno discriminante logístico (dado que no es correcto asumir la multinormalidad necesaria para los AD lineal y cuadrático). Podría buscarse sino, construir un discriminante factorial, donde no se asuma ninguna función de distribución de los datos.

En el caso del determinante factorial lo que se busca es encontrar combinaciones lineales de los datos ($Z = Xu$) que maximicen la discriminación de los datos en k grupos.

EJERCICIO 4

Parte 1: derivar el estadístico de razón de verosimilitud para la prueba de igualdad de medias.

La matriz de información tiene la siguiente forma:

$$\mathbb{X} = \begin{bmatrix} x_{111} & x_{121} & \cdots & x_{1p1} \\ x_{211} & x_{221} & \cdots & x_{2p1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n_1 11} & x_{n_1 21} & \cdots & x_{n_1 p1} \\ x_{112} & x_{122} & \cdots & x_{1p2} \\ x_{212} & x_{222} & \cdots & x_{2p2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n_2 11} & x_{n_2 21} & \cdots & x_{n_2 p1} \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ x_{11k} & x_{12k} & \cdots & x_{1pk} \\ x_{21k} & x_{22k} & \cdots & x_{2pk} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n_k 1k} & x_{n_k 2k} & \cdots & x_{n_k pk} \end{bmatrix} = \begin{bmatrix} \mathbb{X}_1 \\ \mathbb{X}_2 \\ \vdots \\ \mathbb{X}_k \end{bmatrix}$$

$$con\ i = 1, \dots, n \quad j = 1, \dots, p \quad g = 1, \dots, k\ con\ n = \sum_{g=1}^{k} n_g$$

Donde cada $\mathbb{X}_g$ representa la sub-matriz que contiene los datos provenientes de la población g

$$X_g \sim N_p\left(\mu_g; \, \mathbf{\Sigma}_g\right) \ \ con\ i = 1, \dots, n\ y\ g = 1, \dots, k$$

La función de densidad multivariada es:

$$f_X(x) = (2\pi)^{-p/2} * |\mathbf{\Sigma}|^{-1/2} * exp\left\{-\frac{1}{2}(x-\mu)'\mathbf{\Sigma}^{-1}(x-\mu)\right\}$$

Asumiendo que $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 = \cdots = \mathbf{\Sigma}_k = \mathbf{\Sigma}$

La prueba puede plantearse de la siguiente forma:

[1] Ver planteo del test en el Ejercicio 1

$$H0)\ \vec{\mu}_1 = \vec{\mu}_2 = \cdots = \vec{\mu}_k = \vec{\mu}_0 \quad Versus \quad H1)\ \exists \vec{\mu}_g \neq \vec{\mu}_0$$

La verosimilitud bajo H0 cierta y asumiendo Σ conocida será:

$$L(\vec{\mu}_0;\ \mathbf{\Sigma}|\mathbb{X}) = \prod_{i=1}^{n} (2\pi)^{-p/2} * |\mathbf{\Sigma}|^{-1/2} * exp\left\{-\frac{1}{2}(\vec{x}_i - \vec{\mu}_0)'\mathbf{\Sigma}^{-1}(\vec{x}_i - \vec{\mu}_0)\right\}$$

La cual puede expresarse en término logarítmicos, y remplazando la matriz de varianzas y covarianzas por su expresión máximo verosímil como:

$$l(\vec{\mu}_0;\ \mathbf{\Sigma}|\mathbb{X}) = -\frac{n}{2} log|\mathbf{S}| - \frac{np}{2}$$

La verosimilitud bajo H1 cierta y asumiendo Σ conocida será:

$$L(\vec{\mu}_1, \vec{\mu}_2, \dots, \vec{\mu}_k;\ \mathbf{\Sigma}|\mathbb{X}) = \prod_{i=1}^{n_1} f_{\mathbb{X}_1}(x_{i1}) * \prod_{i=1}^{n_2} f_{\mathbb{X}_2}(x_{i2}) * \dots * \prod_{i=1}^{n_k} f_{\mathbb{X}_k}(x_{ik}) =$$

$$= \prod_{g=1}^{k} \prod_{i=1}^{n_g} (2\pi)^{-p/2} * |\mathbf{\Sigma}|^{-1/2} * exp\left\{-\frac{1}{2}(\vec{x}_{ig} - \vec{\mu}_g)'\mathbf{\Sigma}^{-1}(\vec{x}_{ig} - \vec{\mu}_g)\right\}$$

La cual se maximiza cuando: $\bar{\vec{x}}_g = \vec{\mu}_g$

$$= \prod_{g=1}^{k} \prod_{i=1}^{n_g} (2\pi)^{-p/2} * |\mathbf{\Sigma}|^{-1/2} * exp\left\{-\frac{1}{2}(\vec{x}_{ig} - \bar{\vec{x}}_g)'\mathbf{\Sigma}^{-1}(\vec{x}_{ig} - \bar{\vec{x}}_g)\right\} =$$

$$= (2\pi)^{-np/2} * |\mathbf{\Sigma}|^{-n/2} * exp\left\{-\frac{1}{2}\sum_{g=1}^{k}\sum_{i=1}^{n_g}(\vec{x}_{ig} - \bar{\vec{x}}_g)'\mathbf{\Sigma}^{-1}(\vec{x}_{ig} - \bar{\vec{x}}_g)\right\}$$

Teniendo en cuenta que:

$$\sum_{g=1}^{k}\sum_{i=1}^{n_g}\left[(\vec{x}_{ig} - \bar{\vec{x}}_g)'\mathbf{\Sigma}^{-1}(\vec{x}_{ig} - \bar{\vec{x}}_g)\right] = tr\left\{\sum_{g=1}^{k}\sum_{i=1}^{n_g}\left[(\vec{x}_{ig} - \bar{\vec{x}}_g)'\mathbf{\Sigma}^{-1}(\vec{x}_{ig} - \bar{\vec{x}}_g)\right]\right\} =$$

$$= \sum_{g=1}^{k}\sum_{i=1}^{n_g} tr\left[\mathbf{\Sigma}^{-1}(\vec{x}_{ig} - \bar{\vec{x}}_g)'(\vec{x}_{ig} - \bar{\vec{x}}_g)\right] = tr(\mathbf{\Sigma}^{-1}\mathbf{W})$$

Si llamamos W a la matriz de suma de cuadrados dentro de los grupos:

$$W = \sum_{g=1}^{k}\sum_{i=1}^{n_g}(\vec{x}_{ig} - \bar{\vec{x}}_g)(\vec{x}_{ig} - \bar{\vec{x}}_g)'$$

Con esto, la función de verosimilitud bajo H1 nos queda:

$$L(\vec{\mu}_1, \vec{\mu}_2, \dots, \vec{\mu}_k;\ \mathbf{\Sigma}|\mathbb{X}) = (2\pi)^{-np/2} * |\mathbf{\Sigma}|^{-n/2} * exp\left\{-\frac{n}{2}tr\left(\mathbf{\Sigma}^{-1}W/_n\right)\right\}$$

O, en términos logarítmicos:

$$l(\vec{\mu}_1, \vec{\mu}_2, \dots, \vec{\mu}_k;\ \mathbf{\Sigma}|\mathbb{X}) = -\frac{np}{2}log(2\pi) - \frac{n}{2}\log(|\mathbf{\Sigma}|) - \frac{n}{2} * tr\left(\frac{\mathbf{\Sigma}^{-1}W}{n}\right)$$

La matriz de varianzas y covarianzas común cuando la media es distinta (es decir, bajo H_1), se estima de la siguiente forma:

$$\mathbf{S}_w = \frac{\mathbf{W}}{n}$$

Por lo que obtenemos:

$$l(\vec{\mu}_1, \vec{\mu}_2, \dots, \vec{\mu}_k;\ \mathbf{\Sigma}|\mathbb{X}) = -\frac{np}{2} - \frac{n}{2}\log(|\mathbf{S}_w|)$$

Recordando que el cociente de verosimilitudes adquiere la siguiente forma:

$$RV = \frac{L(H_0)}{L(H_1)}$$

O, expresada en forma logarítmica:

$$RV = -2[l(H_0) - l(H_1)]$$

Obtenemos que:

$$RV = -2\left[-\frac{n}{2}log(|\mathbf{S}|) - \frac{np}{2} + \frac{np}{2} + \frac{n}{2}\log(|\mathbf{S}_{pl}|)\right] =$$

$$= -n\left[-log|\mathbf{S}| + \log(|\mathbf{S}_{pl}|)\right] = n\left[log(\mathbf{S}) - \log(|\mathbf{S}_{pl}|)\right] = n * log\left[\frac{|\mathbf{S}|}{|\mathbf{S}_{pl}|}\right]$$

El cociente distribuye chi-cuadrado:

$$RV = n * log\left[\frac{|\mathbf{S}|}{|\mathbf{S}_{pl}|}\right] \sim \chi^2_g$$

Donde sus grados de libertad se obtienen como la diferencia entre las dimensiones de los espacios paramétricos de H_0 (es decir: Ω_0), y de H_1 (es decir: Ω).

$$g = \dim(\Omega) - \dim(\Omega_0) = \left[p + \frac{p(p+1)}{2}\right] - \left[kp + \frac{p(p+1)}{2}\right] = p(k-1)$$

Por lo tanto:

$$H0)\ \vec{\mu}_1 = \vec{\mu}_2 = \dots = \vec{\mu}_k = \vec{\mu}_0 \quad Versus \quad H1)\ \exists \vec{\mu}_g \neq \vec{\mu}_0$$

$$RC = \left\{muestras/RV > \chi^2_{p(k-1)}(1-\alpha)\right\}$$

$$RV = n * log\left[\frac{|\mathbf{S}|}{|\mathbf{S}_{pl}|}\right] \sim \chi^2_{p(k-1)}$$

Parte 2: derivar el estadístico de razón de verosimilitud para la prueba de igualdad de matices de varianzas y covarianzas.

La matriz de información tiene la siguiente forma:

$$\mathbb{X} = \begin{bmatrix} x_{111} & x_{121} & \cdots & x_{1p1} \\ x_{211} & x_{221} & \cdots & x_{2p1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n_1 11} & x_{n_1 21} & \cdots & x_{n_1 p1} \\ x_{112} & x_{122} & \cdots & x_{1p2} \\ x_{212} & x_{222} & \cdots & x_{2p2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n_2 11} & x_{n_2 21} & \cdots & x_{n_2 p1} \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ x_{11k} & x_{12k} & \cdots & x_{1pk} \\ x_{21k} & x_{22k} & \cdots & x_{2pk} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n_k 1k} & x_{n_k 2k} & \cdots & x_{n_k pk} \end{bmatrix} = \begin{bmatrix} \mathbb{X}_1 \\ \mathbb{X}_2 \\ \vdots \\ \mathbb{X}_k \end{bmatrix}$$

$$con\ i = 1, \dots, n \ \ j = 1, \dots, p \ \ g = 1, \dots, k \ con\ n = \sum_{g=1}^{k} n_g$$

Donde cada $\mathbb{X}_g$ representa la sub-matriz que contiene los datos provenientes de la población g

$$X_g \sim N_p\left(\mu_g;\ \mathbf{\Sigma}_g\right) \ \ con\ i = 1, \dots, n\ y\ g = 1, \dots, k$$

La función de densidad multivariada es:

$$f_X(x) = (2\pi)^{-p/2} * |\mathbf{\Sigma}|^{-1/2} * exp\left\{-\frac{1}{2}(x-\mu)'\mathbf{\Sigma}^{-1}(x-\mu)\right\}$$

La prueba a evaluar es:

$$H0)\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 = \cdots = \mathbf{\Sigma}_k = \mathbf{\Sigma}_0 \qquad Versus \qquad H1)\ \exists \mathbf{\Sigma}_g \neq \mathbf{\Sigma}_0$$

La verosimilitud bajo H0 cierta y asumiendo igualdad de medias será:

$$L(\vec{\mu};\ \mathbf{\Sigma}_0|\mathbb{X}) = \prod_{i=1}^{n} (2\pi)^{-p/2} * |\mathbf{\Sigma}_0|^{-1/2} * exp\left\{-\frac{1}{2}(\vec{x}_i - \vec{\mu})'\mathbf{\Sigma}_0{}^{-1}(\vec{x}_i - \vec{\mu})\right\}$$

La verosimilitud bajo H1 cierta y asumiendo Σ conocida será:

$$L(\vec{\mu};\ \mathbf{\Sigma}_1, \mathbf{\Sigma}_2, \dots, \mathbf{\Sigma}_k|\mathbb{X}) = \prod_{i=1}^{n_1} f_{\mathbb{X}_1}(x_{i1}) * \prod_{i=1}^{n_2} f_{\mathbb{X}_2}(x_{i2}) * \dots * \prod_{i=1}^{n_k} f_{\mathbb{X}_k}(x_{ik}) =$$

$$= \prod_{g=1}^{k} \prod_{i=1}^{n_g} (2\pi)^{-p/2} * |\mathbf{\Sigma}_g|^{-1/2} * exp\left\{-\frac{1}{2}(\vec{x}_{ig} - \vec{\mu})'\mathbf{\Sigma}_g{}^{-1}(\vec{x}_{ig} - \vec{\mu})\right\}$$

La cual se maximiza cuando: $\widehat{\mathbf{\Sigma}}_g = \mathbf{S}_g$

El cociente de verosimilitud será:

$$\frac{L(\vec{\mu};\ \mathbf{\Sigma}_0|\mathbb{X})}{L(\vec{\mu};\ \mathbf{\Sigma}_1, \mathbf{\Sigma}_2, \dots, \mathbf{\Sigma}_k|\mathbb{X})} = \frac{|\mathbf{\Sigma}_0|^{-n/2} * exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(\vec{x}_i - \vec{\mu})'\mathbf{\Sigma}_0{}^{-1}(\vec{x}_i - \vec{\mu})\right\}}{\prod_{g=1}^{k}\prod_{i=1}^{n_g} * |\mathbf{S}_g|^{-1/2} * exp\left\{-\frac{1}{2}(\vec{x}_{ig} - \vec{\mu})'\mathbf{S}_g{}^{-1}(\vec{x}_{ig} - \vec{\mu})\right\}}$$