



ANÁLISIS ESTADÍSTICO MULTIVARIADO

Análisis Multivariado

1 Análisis Discriminante

- Introducción
- Discriminación en poblaciones normales
- Función discriminante de Fisher
- Anexo: Errores de clasificación

2 Generalización para G poblaciones

- Test de hipótesis

Introducción

- Las técnicas multivariadas de discriminación y separación tienen que ver con poder distinguir distintos conjuntos de elementos y con poder asignar un nuevo elemento a alguno de los conjuntos definidos previamente.
- El objetivo de la **discriminación** está ligado a describir los aspectos diferenciales de los objetos que pertenecen a distintos grupos.
- El objetivo de la **clasificación** está relacionado a la posibilidad de ordenar objetos o elementos en dos o más grupos y reglas óptimas de clasificación.
- Las técnicas que veremos se suelen denominar de **clasificación supervisada** para indicar que conocemos una muestra de elementos correctamente clasificados que sirven de pauta o modelo para la clasificación de otros elementos.

Discriminación para 2 grupos o poblaciones

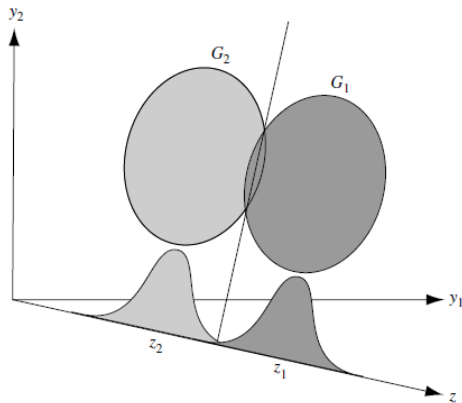


Figure 8.1. Two-group discriminant analysis.

Introducción

- Consideraremos un conjunto de elementos que pueden provenir de dos poblaciones distintas P_1 y P_2 .
- Supondremos que x es continua y que las funciones de densidad de ambas poblaciones f_1 y f_2 son conocidas.
- El problema que estudiaremos es el de clasificar un nuevo elemento a partir de los valores conocidos de x , x_0 , en alguna de las dos poblaciones.
- Si conocemos las probabilidades a priori π_1 y π_2 donde $\pi_1 + \pi_2 = 1$ de que este nuevo elemento provenga de cada una de estas dos poblaciones, su distribución de probabilidad será una distribución mezclada:

$$f(x) = \pi_1 f_1 + \pi_2 f_2$$

Reglas de clasificación

- Una vez observado x_0 podemos calcular las probabilidades a posteriori de que este elemento provenga de una de las dos poblaciones $P(i/x_0)$, ($i = 1, 2$). Aplicando el teorema de Bayes:

$$P(1/x_0) = \frac{P(x_0/1)\pi_1}{P(x_0/1)\pi_1 + P(x_0/2)\pi_2} \quad \text{donde} \quad P(x_0/i) = f_i(x_0)$$

Para la segunda población:

$$P(2/x_0) = \frac{P(x_0/2)\pi_2}{P(x_0/1)\pi_1 + P(x_0/2)\pi_2} \quad \text{donde}$$

Clasificaremos x_0 en la población P_1 si $\pi_1 f_1(x_0) > \pi_2 f_2(x_0)$.

Reglas de clasificación

- Como todo procedimiento estadístico, la clasificación de elementos puede estar sujeta a error.
- Los costos asociados a una clasificación equivocada pueden ser diferentes en cada población. Si clasificamos al elemento x_0 en el grupo 2, las posibles consecuencias son:
- Acertar, con probabilidad $P(2/x_0)$, en este caso el costo de la asignación es nulo
- Equivocarnos, con probabilidad $P(1/x_0)$ en cuyo caso el costo de la asignación es $c(2/1)$.

El costo promedio de la decisión d_2 : *clasificar* x_0 en P_2 será:

$$E(d_2) = c(2/1)P(1/x_0) + 0.P(2/x_0) = c(2/1)P(1/x_0)$$

El costo promedio de la decisión d_1 : *clasificar* x_0 en P_1 será:

$$E(d_1) = c(1/2)P(2/x_0) + 0.P(1/x_0) = c(1/2)P(2/x_0)$$

Reglas de clasificación

- Las regiones R_1 y R_2 que minimizan el costo esperado de una mala clasificación están definidas por los valores de x que satisfacen las siguientes desigualdades.

- Región 1:

$$\frac{f_1(x)}{f_2(x)} \geq \left(\frac{\pi_2}{\pi_1} \right) \left(\frac{c(1/2)}{c(2/1)} \right)$$

- Región 2:

$$\frac{f_1(x)}{f_2(x)} < \left(\frac{\pi_2}{\pi_1} \right) \left(\frac{c(1/2)}{c(2/1)} \right)$$

- En términos generales se compara el cociente de densidades con respecto al producto entre el cociente de costos y el cociente de probabilidades a priori.

Casos especiales

- Si las probabilidades a priori son iguales ($\pi_1 = \pi_2$) entonces:

$$R_1 : \frac{f_1(x)}{f_2(x)} \geq \left(\frac{c(1/2)}{c(2/1)} \right) \quad R_2 : \frac{f_1(x)}{f_2(x)} < \left(\frac{c(1/2)}{c(2/1)} \right)$$

- Si los costos de una mala clasificación son iguales ($c(1/2) = c(2/1)$) entonces:

$$R_1 : \frac{f_1(x)}{f_2(x)} \geq \left(\frac{\pi_2}{\pi_1} \right) \quad R_2 : \frac{f_1(x)}{f_2(x)} < \left(\frac{\pi_2}{\pi_1} \right)$$

- Si tanto las probabilidades a priori como los costos de una mala clasificación son iguales entonces:

$$R_1 : \frac{f_1(x)}{f_2(x)} \geq 1 \quad R_2 : \frac{f_1(x)}{f_2(x)} < 1$$

Clasificación de poblaciones normales

- Supondremos que f_1 y f_2 son dos distribuciones normales con distintos vectores de medias pero idéntica matriz de variancias y covariancias, que caracterizan a dos poblaciones P_1 y P_2 .
- La partición óptima es clasificar x_0 en la población P_1 si:

$$\frac{e^{-\frac{1}{2}(x-\mu_1)'\Sigma^{-1}(x-\mu_1)}}{e^{-\frac{1}{2}(x-\mu_2)'\Sigma^{-1}(x-\mu_2)}} \geq \left(\frac{c(1/2)}{c(2/1)} \right) \left(\frac{\pi_2}{\pi_1} \right)$$

- Como ambos términos son siempre positivos, tomando logaritmos y reordenando términos se obtiene:

$$(\mu_1 - \mu_2)'\Sigma^{-1}x - \frac{1}{2}(\mu_1 - \mu_2)'\Sigma^{-1}(\mu_1 + \mu_2) \geq \log \left[\left(\frac{c(1/2)}{c(2/1)} \right) \left(\frac{\pi_2}{\pi_1} \right) \right]$$

caso contrario clasificar x_0 en la población P_2

Clasificación de poblaciones normales

Otra forma de representar la regla de asignación anterior:

- Llamando D_i^2 a la distancia de Mahalanobis entre un punto observado y la media de la población i , clasificaremos x_0 en la población P_1 si se verifica:

$$(D_2^2 - D_1^2) \geq 2 \log \left[\left(\frac{c(1/2)}{c(2/1)} \right) \left(\frac{\pi_2}{\pi_1} \right) \right]$$

- Suponiendo que los costos asociados a una clasificación equivocada y las probabilidades a priori son iguales para ambas poblaciones, la regla de clasificación se transforma en *Clasificar x_0 en la Población 1 si $D_2^2 > D_1^2$* que es equivalente a clasificar x_0 en la población cuya media esté más próxima al punto, en base a la distancia de Mahalanobis.

Clasificación de poblaciones normales

- Consideremos nuevamente la regla de discriminación:

$$(\mu_1 - \mu_2)' \Sigma^{-1} x - \frac{1}{2} (\mu_1 - \mu_2)' \Sigma^{-1} (\mu_1 - \mu_2) \geq \log \left[\left(\frac{c(1/2)}{c(2/1)} \right) \left(\frac{\pi_2}{\pi_1} \right) \right]$$

Reordenando términos se obtiene:

$$(\mu_1 - \mu_2)' \Sigma^{-1} x = (\mu_1 - \mu_2)' \Sigma^{-1} \left(\frac{\mu_1 + \mu_2}{2} \right) + \log \left[\frac{c(1/2)\pi_2}{c(2/1)\pi_1} \right]$$

y definiendo $w = \Sigma^{-1}(\mu_1 - \mu_2)$ la frontera se puede escribir como:

$$w'x = w' \left(\frac{\mu_1 + \mu_2}{2} \right) + \log \left[\frac{c(1/2)\pi_2}{c(2/1)\pi_1} \right]$$

Discriminación para 2 grupos o poblaciones

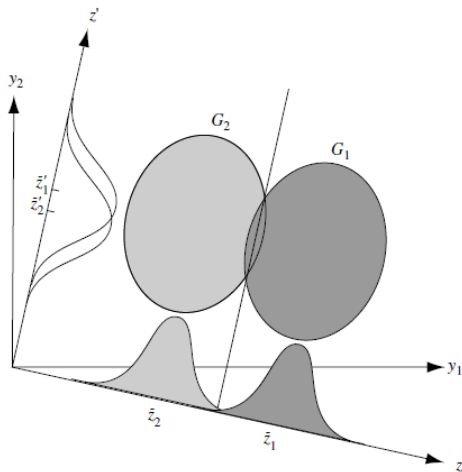


Figure 8.2. Separation achieved by the discriminant function.

Función discriminante de Fisher

- La idea de Fisher fue transformar las observaciones multivariadas x por valores univariados z de manera tal que los resultados provenientes de las poblaciones P_1 y P_2 estén lo más separados posibles.
- El criterio propuesto por Fisher es encontrar una variable escalar $z = \alpha'x$ tal que maximice la distancia entre las medias proyectadas con relación a la variabilidad resultante en la proyección. Intuitivamente, la escala z permitirá separar lo máximo posible ambos grupos.
- El enfoque planteado por Fisher no asume que las poblaciones sean normales, pero implícitamente asume la igualdad de la matriz de variancias y covariancias en ambas poblaciones.
- La media de la variable z en los grupos 1 y 2 se obtiene como la proyección del vector de medias sobre la dirección de α :

$$\bar{z}_1 = \alpha' \bar{x}_1, \quad \bar{z}_2 = \alpha' \bar{x}_2$$

Función discriminante de Fisher

- La variancia de la variable z será la misma en ambos grupos, $\alpha' V \alpha$ y se estimará con $s_z^2 = \alpha' S_w \alpha$
- Se desea elegir α de manera tal que la separación entre las medias m_1 y m_2 de ambas poblaciones sea máxima.

$$\text{separación} = \frac{|\bar{z}_1 - \bar{z}_2|}{s_z}, \text{ donde } s_z^2 = \frac{\sum_{j=1}^{n_1} (z_{1j} - \bar{z}_1)^2 + \sum_{j=1}^{n_2} (z_{2j} - \bar{z}_2)^2}{n_1 + n_2 - 2}$$

- La función objetivo resulta:

$$\phi = \left(\frac{\bar{z}_2 - \bar{z}_1}{s_z} \right)^2 = \left(\frac{[\alpha(\bar{x}_2 - \bar{x}_1)]^2}{\alpha' S_w \alpha} \right)$$

- En esta expresión α representa una dirección ya que ϕ es invariante ante multiplicaciones de α por cualquier constante.
- Encontrar la dirección de máxima discriminación es equivalente a resolver un problema de maximización estándar.

Función discriminante de Fisher

- Resultado: La combinación lineal $z = \alpha'x = (\bar{x}_1 - \bar{x}_2)'S_{pooled}^{-1}x$ maximiza la separación entre las poblaciones P_1 y P_2 .
- El máximo de la función objetivo resulta:

$$D^2 = (\bar{x}_1 - \bar{x}_2)'S_{pooled}^{-1}(\bar{x}_1 - \bar{x}_2)$$

- El criterio de discriminación de Fisher es equivalente al criterio de clasificación de mínimo costo que hallamos anteriormente.
- La regla completa:

$$w = (\bar{x}_1 - \bar{x}_2)'S_{pooled}^{-1} \left[x - \frac{1}{2}(\bar{x}_1 + \bar{x}_2) \right]$$

se suele denominar **función de clasificación de Anderson**.

Discriminación para 2 grupos o poblaciones

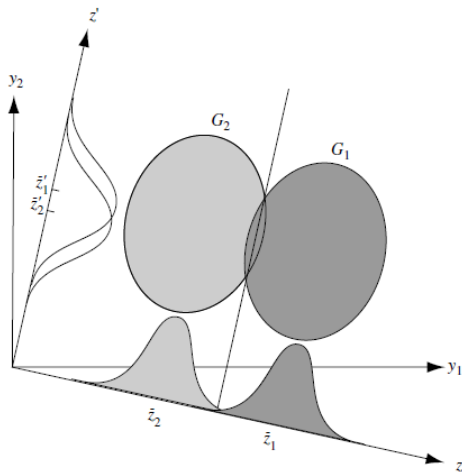


Figure 8.2. Separation achieved by the discriminant function.

Anexo: Error de la clasificación

- En la práctica, el cálculo de errores podría realizarse reemplazando en las expresiones que hallamos anteriormente los parámetros desconocidos por sus estimaciones muestrales.
- Sin embargo, este procedimiento puede subestimar las probabilidades de error al no tener en cuenta la incertidumbre en la estimación de parámetros.
- Existe una medida del desempeño de la clasificación que no depende de la distribución de x en las poblaciones de origen y se puede utilizar para cualquier procedimiento.
- Esta medida se denomina **tasa de error aparente** y se define como la fracción de observaciones en la muestra que han sido clasificadas erróneamente por la regla.
- La tasa de error aparente se puede calcular fácilmente a partir de la matriz de confusión que muestra la pertenencia real y estimada de cada observación de la muestra a una población.

Error de la clasificación

- En el caso de 2 grupos obtendríamos:

Población	P_1	P_2
P_1	n_{1C}	$n_{1M} = n_1 - n_{1C}$
P_2	$n_{2M} = n_2 - n_{2C}$	n_{2C}

donde n_{1C} es la cantidad de elementos que pertenecen a P_1 clasificados correctamente, n_{1M} es la cantidad de elementos que pertenecen a P_1 clasificados incorrectamente en la población 2, n_{2C} es la cantidad de elementos que pertenecen a P_2 clasificados correctamente y n_{2M} es la cantidad de elementos que pertenecen a P_2 clasificados incorrectamente en P_1 .

- El error aparente de la regla es:

$$APER = \frac{n_{1M} + n_{2M}}{n_1 + n_2}$$

Error de la clasificación

- Un inconveniente de este método es que tiende a subestimar las probabilidades de error: se utilizan los mismos datos para construir la regla de clasificación y para evaluarla.
- Una alternativa es dividir la muestra de datos en una submuestra de estimación y otra submuestra de validación.
- La tasa de error se calcula como la proporción de elementos mal clasificados en la muestra de validación.
- Aunque este procedimiento soluciona el posible sesgo que introduce el utilizar los mismos datos para determinar la regla y evaluarla, también tiene inconvenientes:
 - ▶ Se requieren muestras grandes.
 - ▶ La función evaluada no es la función de interés: deberían utilizarse casi todos los datos para determinar la regla de clasificación para no perder información valiosa.

Error de la clasificación

Un segundo enfoque es el **Procedimiento de Lachenbruch**:

- Comenzar con las observaciones de P_1 . Descartar una observación de este grupo y determinar la regla de clasificación en base a las $n_1 - 1$ observaciones de P_1 y las n_2 observaciones de P_2 .
- Clasificar la observación que se descartó usando la regla determinada en el punto 1.
- Repetir los pasos 1 y 2 hasta que todas las observaciones de P_1 hayan sido clasificadas. Se calcula el número de observaciones mal clasificadas en P_1 como $n_{1M}^{(H)}$.
- Repetir los pasos 1 a 3 para los elementos de P_2 . Se calcula el número de observaciones mal clasificadas en P_2 como $n_{2M}^{(H)}$.

Error de la clasificación

- Finalmente, las probabilidades condicionales de una mala clasificación se calculan como:

$$\hat{P}(2/1) = \frac{n_{1M}^{(H)}}{n_1} \quad \hat{P}(1/2) = \frac{n_{2M}^{(H)}}{n_2}$$

- La proporción total de elementos mal clasificados:

$$\frac{n_{1M}^{(H)} + n_{2M}^{(H)}}{n_1 + n_2}$$

- Este procedimiento se suele denominar **Jackknife** o de validación cruzada y permite una mejor estimación del error de clasificación. Si n es muy grande, este procedimiento puede resultar costoso desde el punto de vista del cálculo.

Generalización para G poblaciones

- La generalización para G poblaciones es simple: el objetivo es dividir el espacio muestral en G regiones A_1, A_2, \dots, A_G tales que si $x_0 \in A_g$ el punto se clasifica en P_g .
- Supondremos que los costos de clasificación son constantes y no dependen de la población en que se haya clasificado la observación.
- La región A_g se define por aquellos puntos con máxima probabilidad de ser generados por P_g , donde el producto de las probabilidades a priori y la verosimilitud sea máxima:

$$A_g = \{x \in E_x / \pi_g f_g(x) > \pi_i f_i(x), \forall i \neq g\}$$

Generalización para G poblaciones

- Si las probabilidades a priori son todas iguales ($\pi_i = 1/G, i = 1, \dots, G$) entonces:

$$R_k : \frac{f_k(x)}{f_i(x)} \geq \left(\frac{c(k/i)}{c(i/k)} \right) \quad \forall i \neq k$$

- Si los costos de una mala clasificación son iguales ($c(1/2) = c(2/1)$) entonces:

$$R_k : \frac{f_k(x)}{f_i(x)} \geq \left(\frac{\pi_i}{\pi_k} \right) \quad \forall i \neq k$$

- Si tanto las probabilidades a priori como los costos de una mala clasificación son iguales entonces:

$$R_k : \frac{f_k(x)}{f_i(x)} \geq 1 \quad \forall i \neq k$$

Criterio de Fisher para G poblaciones

- El objetivo es definir un vector $z = (z_1, \dots, z_r)$ de variables canónicas, donde $r = \min(G - 1, p)$ que se obtienen como combinación lineal de las p variables x originales, donde $z_i = u_i'x$ y que permitan resolver el problema de clasificación de un nuevo elemento en alguna de las G poblaciones posibles. Para ello:
 - ▶ Proyectamos las medias de las variables en los grupos sobre el espacio determinado por las r variables canónicas.
 - ▶ Proyectamos sobre el mismo espacio al punto x_0 a clasificar, llamando z_0 a la proyección
 - ▶ Clasificamos el punto en aquella población cuya media se encuentre más próxima. Clasificamos en la población i si:

$$(z_0 - \bar{z}_i)'(z_0 - \bar{z}_i) = \min_g (z_0 - \bar{z}_g)'(z_0 - \bar{z}_g)$$

Criterio de Fisher para G poblaciones

- Cuando hay varios grupos o poblaciones, mediremos la separación entre las medias por el cociente entre la variabilidad entre grupos **variabilidad explicada por los grupos** y la variabilidad dentro de los grupos **variabilidad no explicada o residual**.
- Para obtener las variables canónicas discriminantes comenzaremos buscando un vector u_1 (de norma unitaria) tal que los grupos proyectados sobre esta dirección tengan separación máxima.
- La proyección de la media de la población g sobre esta dirección es $\bar{z}_g = u_1' \bar{x}_g$ y la media de la proyección de todos los datos será $\bar{z}_T = u_1' \bar{x}_T$ donde \bar{x}_T es el vector $p \times 1$ que contiene las medias de las p variables para las n observaciones de la muestra uniendo todos los grupos.

Criterio de Fisher para G poblaciones

- Tomando como medida de la distancia entre las medias de los grupos proyectadas $\bar{z}_1, \dots, \bar{z}_G$ su variación total dada por

$$\sum_{g=1}^G n_g (\bar{z}_g - \bar{z}_T)^2$$

y comparando esta cantidad con la variabilidad dentro de los grupos:

$$\sum_{g=1}^G \sum_{i=1}^{n_g} (z_{ig} - \bar{z}_g)^2$$

La separación relativa entre las medias vendrá dada por el estadístico:

$$\phi = \frac{\sum_{g=1}^G n_g (\bar{z}_g - \bar{z}_T)^2}{\sum_{g=1}^G \sum_{i=1}^{n_g} (z_{ig} - \bar{z}_g)^2}$$

Criterio de Fisher para G poblaciones

- Si todos los datos provienen de la misma población y no existen grupos diferentes, el cociente ϕ tiene distribución F con $(G - 1)$ y $(n - G + 1)$ grados de libertad.
- La suma de cuadrados dentro de los grupos para los puntos proyectados se expresa como:

$$\begin{aligned} VNE &= \sum_{g=1}^G \sum_{i=1}^{n_g} (z_{ig} - \bar{z}_g)^2 \\ &= \sum_{g=1}^G \sum_{i=1}^{n_g} u' (x_{ig} - \bar{x}_g)(x_{ig} - \bar{x}_g)' u \\ &= u' W u \end{aligned}$$

donde W en general tendrá rango p y se define como:

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

Criterio de Fisher para G poblaciones

- La suma de cuadrados entre grupos para los puntos proyectados es:

$$\begin{aligned} VE &= \sum_{g=1}^G n_g (\bar{z}_g - \bar{z}_T)^2 \\ &= \sum_{g=1}^G n_g u' (\bar{x}_g - \bar{x}_T) (\bar{x}_g - \bar{x}_T)' u = u' B u \end{aligned}$$

donde B se define como:

$$B = \sum_{g=1}^G n_g (\bar{x}_g - \bar{x}_T) (\bar{x}_g - \bar{x}_T)' = \sum_{g=1}^G n_g a_g a_g'$$

- La matriz $B_{p \times p}$ es simétrica y se obtiene como suma de G matrices de rango 1 formadas por los vectores no independientes a_g , ya que $\sum_{g=1}^G n_g a_g = 0$. Esto implica que B es de rango $G - 1$.

Criterio de Fisher para G poblaciones

- En términos de las matrices W y B , el cociente a maximizar es:

$$\phi = \frac{\sum_{g=1}^G n_g (\bar{z}_g - \bar{z}_T)^2}{\sum_{g=1}^G \sum_{i=1}^{n_g} (z_{ig} - \bar{z}_g)^2} = \frac{u_1' B u_1}{u_1' W u_1}$$

- Derivando e igualando a cero obtenemos:

$$\frac{\partial \phi}{\partial u_1} = \frac{2B u_1 (u_1' W u_1) - 2(u_1' B u_1) W u_1}{(u_1' W u_1)^2} = 0$$

Entonces:

$$\begin{aligned} B u_1 &= W u_1 \left(\frac{u_1' B u_1}{u_1' W u_1} \right) \\ B u_1 &= \phi W u_1 \\ W^{-1} B u_1 &= \phi u_1 \end{aligned}$$

Criterio de Fisher para G poblaciones

- Elegir u_1 tal que maximice ϕ es equivalente a buscar el autovector asociado al autovalor mas grande de la matriz $W^{-1}B$.
- Este procedimiento se puede generalizar para encontrar más direcciones que maximicen la separación entre las poblaciones, con la condición de que cada nueva variable canónica $z_s = u'_s x$ esté no correlacionada con las anteriores.
- Los autovectores de la matriz $W^{-1}B$ no necesariamente serán ortogonales ya que no hay garantías de que la matriz sea simétrica.
- Por otro lado, el rango de $W^{-1}B$ será $r = \min(p, G - 1)$ y este es el número máximo de direcciones discriminantes que podremos encontrar.
- La matriz $W^{-1}B$ ha sido denominada por Rao como **matriz de distancias de Mahalanobis generalizada** ya que su traza es la suma de las distancias de Mahalanobis entre la media de cada grupo y la media total.

$$tr(W^{-1}B) = \sum_{g=1}^G (\bar{x}_g - \bar{x}_T)' (W/n_g)^{-1} (\bar{x}_g - \bar{x}_T)$$

Criterio de Fisher para G poblaciones

- Si las medias de las poblaciones están ubicadas sobre una recta entonces:

$$(\bar{x}_1 - \bar{x}_T) = k_2(\bar{x}_2 - \bar{x}_T) = \cdots = k_G(\bar{x}_G - \bar{x}_T) = c(\bar{x}_1 - \bar{x}_2)$$

Y la matriz B puede escribirse como:

$$B = \sum_{g=1}^G n_g (\bar{x}_g - \bar{x}_T)(\bar{x}_g - \bar{x}_T)' = k^* (\bar{x}_1 - \bar{x}_2)(\bar{x}_1 - \bar{x}_2)'$$

Siendo k^* una constante. El autovector de B asociado al único autovalor no nulo de $W^{-1}B$ es proporcional a $W^{-1}(\bar{x}_1 - \bar{x}_2)$.

Test de igualdad de medias

- Un enfoque alternativo: llamemos variabilidad total de los datos a la suma que mide los desvíos con respecto a la media general:

$$T = \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})'$$

- Vamos a descomponer la matriz $T = B + W$.
- La matriz W es la matriz de desvíos con respecto a la media de cada grupo que definimos antes. La matriz B medirá la variabilidad explicada por las diferencias entre las medias.

$$T = \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})' = \sum_{g=1}^G \sum_{j=1}^{n_g} (x_{jg} - \bar{x}_g + \bar{x}_g - \bar{x})(x_{jg} - \bar{x}_g + \bar{x}_g - \bar{x})'$$

$$T = \sum_{g=1}^G \sum_{j=1}^{n_g} (x_{jg} - \bar{x}_g)(x_{jg} - \bar{x}_g)' + \sum_{g=1}^G \sum_{j=1}^{n_g} (\bar{x}_g - \bar{x})(\bar{x}_g - \bar{x})'$$

Test de igualdad de medias

- La matriz de variabilidad explicada o sumas de cuadrados entre grupos resulta:

$$B = \sum_{g=1}^G n_g (\bar{x}_g - \bar{x})(\bar{x}_g - \bar{x})'$$

- Para plantear el test de igualdad de medias podemos comparar el tamaño de las matrices W y B , calculando el cociente $|B + W|/|W|$ cuya distribución se puede aproximar por la distribución F . Para tamaños de muestra moderados el test es similar al de razón de verosimilitudes que puede escribirse como:

$$\lambda' = m \log \frac{|T|}{|W|} = m \log \frac{|B + W|}{|W|} = m \log |I + W^{-1}B| = m \sum_i \log(1 + \lambda_i)$$

- Expresando con λ_i a los autovalores de la matriz $W^{-1}B$.

Test para determinar la cantidad de direcciones de discriminación

- Se considerará el estadístico anterior

$$\lambda = m \sum_i \log(1 + \lambda_i)$$

donde

$$\lambda_j = m \log(1 + \lambda_j)$$

$H_0) \exists$ 1 dirección de discriminación $H_1) \exists$ más de 1 dirección relevante

$$\lambda - \lambda_j \sim_{H_0} \chi^2_{(p-1)(G-2)}$$

$H_0) \exists$ m direcciones de discriminación $H_1) \exists$ más de m direcciones relevantes

$$\lambda - m \sum_i^m \log(1 + \lambda_i) \sim_{H_0} \chi^2_{(p-1)(G-m-1)}$$