ANALISIS ESTADISTICO MULTIVARIANTE Y REPRESENTACION CANONICA DE FUNCIONES ESTIMABLES

TESIS PARA OPTAR AL GRADO
DE DOCTOR EN CIENCIAS, SECCION
DE MATEMATICAS, PRESENTADA POR

CARLOS M. CUADRAS AVELLANA

UNIVERSIDAD DE BARCELONA

Recíprocamente, si $P = X' \Delta X D_1$, para un cierto D_1, entonces, tomando $\hat{D} = \Delta X D_1$, se verificará $P = X'\hat{D}$.

Además, $\hat{D}$ es único, pues si $P = X' \Delta X \bar{D}_1$, siendo $\bar{D}_1 = D_1 + V$, y $V'X' \Delta X = O$, entonces, por ser $S = T$, tendremos $V'X' = O$, y por lo tanto

$$\hat{\bar{D}} = \Delta X \bar{D}_1 = \Delta X(D_1 + V) = \Delta X D_1 = \hat{D}$$

c. q. d.

<u>Teorema 4.5.1</u>

Sea $\psi^* = P'\beta^* = p_1\beta_1^* + \dots + p_m\beta_m^*$ una función paramétrica estimable. Sea $\hat{\beta}$ una estimación LS de β^*.

1) El estimador lineal

$$\hat{\psi} = P'\hat{\beta} = p_1\hat{\beta}_1 + \dots + p_m\hat{\beta}_m$$

es un estimador insesgado y de dispersión mínima de ψ^*. (Teorema de Gauss-Markov).

2) Si $\hat{D} = (\hat{d}_1, \dots, \hat{d}_k)'$ verifica:

$$\hat{D} = \Delta X D_1, \quad \text{siendo} \quad P' = X' \Delta X D_1,$$

entonces

$$\psi^* = \sum_{i=1}^{k} \hat{d}_i m_i^*$$

es la expresión Δ-óptima de ψ^*.

<u>Demostración:</u>

Por ser ψ^* estimable, $P = X'D$, y por el lema anterior $P = X' \Delta X D_1$ para un cierto vector D_1.

Si $\hat{\beta}$ es una estimación LS de β^*,

$$P'\hat{\beta} = D_1' X' \Delta X = D_1' X' \Delta \bar{\pi}^* = \hat{D}' \bar{\pi}^*$$

siendo $\hat{D} = \Delta X D_1 = (\hat{d}_1, \dots, \hat{d}_k)'$, que por el lema anterior, es único.

Al ser $X'\hat{D} = X' \Delta X D_1 = P$, por el teorema 4.3.2 tendremos

$$\hat{\Psi} = P'\hat{\beta} = D'\bar{\Pi} = \sum_{i=1}^{k} \hat{d}_i \bar{\rho}_i$$

y Ψ^* admite la expresión

$$\Psi^* = \sum_{i=1}^{k} \hat{d}_i m_i^*.$$

Luego $\mathcal{E}^*(\hat{\Psi}) = \Psi^*$, y $\hat{\Psi}$ es estimador insesgado. Veamos que es de dispersión mínima, es decir, que la expresión

$$\Psi^* = \sum_{i=1}^{k} \hat{d}_i m_i^* \quad \text{es } \Delta\text{-óptima.}$$

Sea

$$\hat{\Psi}_1 = \sum_{i=1}^{k} d_i \bar{\rho}_i = D'\bar{\Pi}$$

otro estimador insesgado, siendo $P = X'D$. El producto escalar en N de $(\hat{\Psi}_1 - \hat{\Psi})$ y $\hat{\Psi}$ es:

$$(\hat{\Psi}_1 - \hat{\Psi})\cdot\hat{\Psi} = (\hat{D}' - D')\Delta^{-1}\hat{D} = (\hat{D}' - D')\Delta^{-1}\Delta X D_1 =$$

$$= (\hat{D}'X - D'X)D_1 = (P' - P')D_1 = 0$$

Luego: $\|\hat{\Psi}_1\|^2 = \|\hat{\Psi}_1 - \hat{\Psi} + \hat{\Psi}\|^2 = \|\hat{\Psi}_1 - \hat{\Psi}\|^2 + \|\hat{\Psi}\|^2 \geq \|\hat{\Psi}\|^2$

c. q. d.

Este teorema nos proporciona dos maneras de hallar la estimación óptima $\hat{\Psi}$ de la función paramétrica $\Psi^* = P'\beta^*$

a) Hallando la estimación $\hat{\beta}$ de los parámetros resolviendo las ecuaciones normales, y entonces $\hat{\Psi} = P'\hat{\beta}$.

b) Hallando D_1 tal que $P' = X'\Delta X D_1$, y a continuación $\hat{D} = \Delta X D_1$, que nos da la expresión Δ-óptima $\Psi^* = \sum_i \hat{d}_i m_i^*$. Entonces:

$$\hat{\Psi} = \hat{d}_1 \bar{\rho}_1 + \ldots + \hat{d}_k \bar{\rho}_k$$

5 HIPOTESIS LINEALES Y COMPARACION ESTADISTICA DE FUNCIONES ESTIMABLES

Se expone la estimación de la matriz de varianzas-covarianzas, siguiendo las notaciones de RAO (1965), el test estadístico de comparación de funciones estimables, basándonos en el principio de unión-intersección de ROY (1957), y algunos otros resultados de análisis multivariante

El contenido de este capítulo es materia bien conocida. Su inclusión se justifica por la necesidad que de él tendremos más adelante, y por el deseo de conseguir una completa unidad de exposición de la memoria.

Otras referencias: ANDERSON(1958), MORRISON(1967), RAO(1951), ROY(1953,1966).

5.1. PROPIEDADES DE LAS ESTIMACIONES LS

Sea $Y \in E$ una v.a. tal que $var(Y) = \sigma^2$. Para no perder grados de libertad, vamos a considerar aquí los $n = \sum n_i$ valores muestrales de un muestreo $\Delta = (n_1, \ldots, n_k)$ de Y. Indicaremos la muestra mY y el vector aleatorio $Y^{(n)}$, abreviadamente,

$$mY = (y_1, \ldots, y_n) \qquad Y^{(n)} = (Y^1 \ldots Y^n)$$

Sea E_Y^n el espacio vectorial de dimensión n,

$$E_Y^n = \langle Y^1, \ldots, Y^n \rangle$$

con base ortonormal, justificado por el hecho de que las v.a. $\{Y^i\}$ son estocásticamente independientes.

Consideremos ahora un espacio vectorial euclídeo E_n de dimensión n. Una muestra $mY = (y_1, \ldots, y_n)$ de cualquier v.a. Y, se puede identificar con un punto de E_n. El muestreo $m \in m_\Delta$, define, pues, una aplicación:

$$m : E_Y^n \longrightarrow E_n$$

$$m(Y^{(n)}) = mY \qquad \forall Y \in E$$

Asimismo, la esperanza matemática $\mathcal{E}$ define una aplicación:

$$\mathcal{E} : E_Y^n \longrightarrow E_n$$

$$(5.1.1) \qquad \mathcal{E}(Y^{(n)}) = X_a \beta_Y$$

Consideremos ahora el subespacio $F_r \subset E_n$, generado por los vectores columna de X_a. Tendremos,

$$\dim (F_r) = \text{rang } (X) = r$$

Observemos, finalmente, que para toda v.a. Y, la estructura métrica de los espacios E_Y^n y E_n, ambos con base ortonormal, es la misma, y se pueden identicar totalmente.

Teorema 5.1.1

Sea Y una v.a. de E. Se verifica:

1) $\mathcal{E}(Y^{(n)}) \in F_r$

2) Cualquiera que sea el muestreo $m \in m_\Delta$

$$(\mathcal{E}(Y^{(n)}) - mY) = (X_a \hat{\beta}_Y - mY)$$

es ortogonal al subespacio F_r.

Demostración:

1) Por (5.1.1), las componentes de la imagen de $Y^{(n)}$ por $\mathcal{E}$ son combinación lineal de los vectores columna de X_a.

2) De las ecuaciones normales (4.4.4), deducimos:

$$X_a' \cdot (X_a \hat{\beta}_Y - mY) = 0$$

de donde $(X_a \hat{\beta}_Y - mY)$ es ortogonal a F_r.

5.2. ESTIMACION DE LA MATRIZ Σ

De la misma manera que los indivíduos medios y su extensión, las funciones estimables, son, en la práctica, desconocidas, tambien lo es la matriz Σ de varianzas-covarianzas, común a las poblaciones H_i $(i=1,\dots,k)$.

Sea P una matriz de orden (n,n), de modo que sus vectores columna $(P_1,\dots,P_r,P_{r+1},\dots,P_n)$ sean ortonormales, y los r primeros pertenezcan al subespacio F_r generado por los vectores de componentes las columnas de X_a.

La matriz P define un cambio de base isométrico

$$Y^{(n)} = (Y^1,\dots,Y^n) \xrightarrow{P} (Z^1,\dots,Z^n) = Z^{(n)}$$

$$Z^i = p_{1i} Y^1 + \dots + p_{ni} Y^n \qquad i=1,\dots,n$$

Se verifica:

$$(5.2.1)\qquad \begin{array}{ll} \mathcal{E}(Z^i) = P_i' \mathcal{E}(Y^{(n)}) = 0 \quad \text{si} \quad i > r & \text{(por el teorema 5.1.1).} \\ \mathcal{E}(Z^i) = \xi_i \quad \text{si} \quad i \leq r & \end{array}$$

Sea ahora:

$$\mathcal{L}(\hat{\beta})(Y) = \sum_{i=1}^{n} (y_i - \sum_{j=1}^{m} x_{ji} \beta_{Yj})^2 =$$

$$= \| mY - X_a \hat{\beta}_Y \|^2.$$

de acuerdo con la norma que rigen en E_n.

Según el teorema 5.1.1, $(mY - X_a \hat{\beta}_Y)$ es ortogonal al subespacio F_r, luego,

$$(5.2.2)\qquad \mathcal{L}(\hat{\beta})(Y) = \sum_{i=r\ 1}^{n} z_i^2$$

siendo $mZ = (z_1, \ldots, z_r, z_{r+1}, \ldots, z_n)$ el muestreo sobre $Z^{(n)}$, expresión en la nueva base dual $\{(Z^i)^*\}$ que la matriz ortonormal P define sobre E^{*n}_{Y}.

Puesto que (5.2.2) es cierta para toda muestra, considerando $\mathcal{L}(\hat{\beta})(Y)$ como una variable aleatoria, tendremos

$$\mathcal{L}(\hat{\beta})(Y) = \sum_{i=1}^{n} (Y^i - \sum_{j=1}^{m} x_{ji} \beta_{Yj})^2 = \sum_{i=r\ 1}^{n} (Z^i)^2$$

y, tomando la esperanza,

$$\mathcal{E}(\mathcal{L}(\hat{\beta})(Y)) = \sum_{i=r\ 1}^{n} \mathcal{E}[(Z^i)^2] = (n-r)\sigma^2$$

pues, por (5.2.1), $\mathcal{E}[(Z^i)^2] = \sigma^2$.

Teorema 5.2.1

La v.a.:

$$\mathcal{L}(\hat{\beta}_Y)(Y)/(n-r)$$

es un estimador insesgado de $\sigma^2 = \text{var}(Y)$,

$$\mathcal{E}\left[\mathcal{L}(\hat{\beta}_Y)(Y)/(n-r)\right] = \sigma^2$$

Este teorema nos proporciona, en particular, la estimación de cada una de las varianzas de las v.a. $Y_1,\ldots,Y_p$ que forman la base de E. Indicaremos:

$$R_o(i,i) = \mathcal{L}(\hat{\beta})(Y_i) \qquad \sigma_{ii} = \text{var}(Y_i)$$

Propongámonos ahora, la estimación de $\sigma_{ij} = \text{cov}(Y_i, Y_j)$.

Sea $Y = Y_i + Y_j$, $R_o^2 = \mathcal{L}(\hat{\beta})(Y)$

Entonces, si $R_o(i,j)$ es el producto escalar de las proyecciones de (mY_i) y (mY_j) en F_r, tendremos:

$$R_o^2 = R_o(i,i) + R_o(j,j) + 2\,R_o(i,j)$$

siendo, en expresión matricial,

$$R_o(i,j) = (mY_i - X_a\hat{\beta}_{Y_i})'(mY_j - X_a\hat{\beta}_{Y_j})$$

Entonces,

$$\mathcal{E}(R_o^2) = (n-r)\text{var}(Y) = (n-r)(\sigma_{ii} + \sigma_{jj} + 2\sigma_{ij})$$

$$\Longrightarrow \quad \mathcal{E}(R_o(i,j)) = (n-r)\sigma_{ij}.$$

Teorema 5.2.2

Sea R_o la matriz cuyos elementos son $R_o(i,j)$. Los elementos de la matriz

$$\hat{\Sigma} = \frac{1}{n-r} R_o$$

son estimadores insesgados de Σ.

5.3. HIPOTESIS LINEALES ESTADISTICAMENTE DEMOSTRABLES

Sean $\psi_1^*,\ldots,\psi_s^*$ funciones paramétricas estimables, linealmente independientes, siendo $s \leqslant r = \text{rang}(X) = \dim M$.

El principal objetivo de esta memoria es establecer una forma sencilla de diferentes funciones estimables, para lo cual necesitamos poder aceptar, con un nivel de significación α, que las s f.p.e. son realmente distintas.

Planteemos la <u>hipótesis nula</u>:

$$H_o: \quad \psi_1^* = \psi_2^* = \ldots = \psi_s^*$$

bajo la cual las s f.p.e. coinciden.

Puesto que las $t=s-1$ funciones paramétricas

$$\psi_{si}^* = \psi_s^* - \psi_i^* \qquad i=1,\ldots,t-1$$

son también estimables y linealmente independientes, (teorema 3.2.1), H_o se expresa también así:

$$H_o': \quad \psi_{s1}^* = \ldots = \psi_{st}^* = 0$$

Sea ahora

$$\psi_{si}^* = h_{i1}\beta_1^* + \ldots + h_{im}\beta_m^* \qquad i=1,\ldots,s$$

e indiquemos por H la matriz

$$H = \begin{pmatrix} h_{11} & \cdots & h_{1m} \\ & \cdots\cdots & \\ h_{t1} & \cdots & h_{tm} \end{pmatrix} \qquad \text{rang } H = t$$

La hipótesis nula tiene una tercera expresión:

$$H_o'': \quad H\,\beta = 0 \qquad \mathfrak{F}(H) \subset \mathfrak{F}(X)$$

siendo $\mathfrak{F}(H)$ y $\mathfrak{F}(X)$ los subespacios generados por las filas de H y X respectivamente.

A esta última formulación de la hipótesis nula se la conoce como <u>hipótesis lineales</u> acerca de los parámetros β. Se dice que estas hipótesis son demostrables, (en el sentido de inferencia estadística), si $\mathcal{F}(H) \subset \mathcal{F}(X)$ (ROY, 1957). Equivalentemente, una comparación H_0 de funciones paramétricas es estadísticamente demostrable si las funciones son estimables.

5.4. TRANSFORMACION DE LOS PARAMETROS BAJO LA HIPOTESIS NULA

Si es cierta la hipótesis nula, existirán $t=s-1$ restricciones lineales, independientes, sobre los parámetros que nos permitirá pasar de m parámetros a $r-s+1$ nuevos parámetros.

Sea C' una matriz de orden $(r-s+1,m)$, tal que

(5.4.1) $$H \cdot C = 0 \qquad \text{rang}(C) = r-s+1$$

Para obtener la matriz C', basta completar las filas de H, con $r-s+1$ vectores ortogonales linealmente independientes.

El vector de parámetros $\theta = (\theta_1, \ldots, \theta_{r-s+1})$ representa la solución general de la ecuación $H\beta = 0$, pues

$$H\beta = H \cdot C\theta = 0$$

El diseño factorial (4.4.3), se transformará, en función de los nuevos parámetros, en:

(5.4.2) $$\mathcal{E}(Y^{(n)}) = X_a C\,\theta_Y \qquad \text{rang}(X_a C) = r-s+1$$

y la nueva matriz ampliada del diseño factorial, será $X_a C$. Al igual que en § 5.1, los vectores columna de esta matriz generan un subespacio F_{r-t}, siendo

$$F_{r-t} \subset F_r \subset E_n \qquad (t=s-1)$$

Podemos aplicar toda la teoría de mínimos cuadrados del capítulo 4, para estimar los nuevos parámetros θ. Procediendo como en §5.2, podremos construir una base ortogonal en E_Y^n:

$$< Z^1,\ldots,Z^t,Z^{t+1},\ldots,Z^r,Z^{r+1},\ldots,Z^n >$$

de modo que

$$F_Y^{r-t} = < Z^{t+1},\ldots,Z^r > \subset F_Y^r = < Z^1,\ldots,Z^r >$$

Además, dada una muestra mY,

$$\mathcal{L}(\hat{\theta})(Y) = (mY - X_a C\ )(mY - X_a C\ )'$$

y, considerado como v.a. (§5.2),

(5.4.3) $$\mathcal{L}(\hat{\theta})(Y) = \sum_{i=t+1}^{r} (Z^i)^2$$

Puesto que los parámetros θ provienen de β restringidos a H_o'', pondremos:

$$\mathcal{L}(\hat{\theta})(Y) = \mathcal{L}(\hat{\beta})\ (Y)_{H\beta=0}$$

5.5. UN TEOREMA FUNDAMENTAL PARA EL ANALISIS CANONICO

El siguiente teorema, de álgebra lineal, relativo a dos productos internos en un espacio vectorial, será fundamental en este capítulo y en los siguientes de la memoria. La demostración puede verse en DEMPSTER (1969, pg. 84).

Teorema fundamental

Sea V un espacio vectorial de dimensión p, y sean π_o y π_1 dos productos escalares en V, de matrices asociadas A_o y A_1, con π_o definido positivo y $\mathrm{rang}(A_1)=q\leq p$

Sean $\{\lambda_i\}$ y $\{V_i\}$ $(i=1,\ldots,q)$, los valores propios y vectores propios de A_1 relativos a A_o, es decir, tales que:

$$A_1(V_i) = \lambda_i\, A_o(V_i) \qquad i=1,\ldots,q$$

siendo $\lambda_1 > \lambda_2 > \ldots > \lambda_q$.

Entonces, V_1 es el vector de V que hace máximo al cociente de módulos al cuadrado, $(V,V)_1/(V,V)_o$, y este máximo es λ_1,

$$\lambda_1 = \frac{(V_1,V_1)_1}{(V_1,V_1)_o} = \max \frac{(V,V)_1}{(V,V)_o} \qquad \forall\, V \in V$$

Del subespacio ortocomplementario a V_1, V_2 maximiza el cociente anterior, y el valor máximo es λ_2,

$$\lambda_2 = \frac{(V_2,V_2)_1}{(V_2,V_2)_o} = \max \frac{(V,V)_1}{(V,V)_o} \qquad V \in \{V_1\}^{\perp} \subset V$$

y así sucesivamente con $V_3,\ldots,V_q$.

Aplicaremos en este capítulo, el teorema fundamental a los dos productos escalares que $\mathcal{L}(\hat{\beta})$ y $\mathcal{L}(\hat{\theta})$ definen sobre E.

Sea $m \in \mathfrak{m}$ un muestreo. Definimos:

$$\mathcal{L}m(\hat{\beta})(Y,\bar{\bar{Y}}) = (mY - X_a\hat{\beta})(m\bar{\bar{Y}} - X_a\hat{\beta})'$$

$$\mathcal{L}m(\hat{\theta})(Y,\bar{\bar{Y}}) = (mY - X_aC\hat{\theta})(m\bar{\bar{Y}} - X_aC\hat{\theta})'$$

La matriz asociada al producto escalar $\mathcal{L}m(\hat{\beta})$ es R_o (§5.2). Indicaremos por R_1, la matriz asociada a $\mathcal{L}m(\theta)$.

$$R_1 = (R_1(i,j)) \quad \text{siendo} \quad R_1(i,j) = \mathcal{L}m(\hat{\theta})(Y_i,Y_j).$$

Para un determinado muestreo m, las matrices R_o y R_1, son fijas. Consideradas como funciones de m son matrices aleatorias.

5.6. TEST ESTADISTICO DE COMPARACION DE FUNCIONES ESTIMABLES SUPONIENDO NORMALIDAD

En este apartado supondremos que la distribución conjunta de las v.a. $(Y_1,\ldots,Y_p)$, es normal multivariante en cada celda H_i $(i=1,\ldots,k)$.

$$(Y_1,\ldots,Y_p) \quad \text{es} \quad N((\mathcal{E}(Y_1),\ldots,\mathcal{E}(Y_p))',\Sigma) \text{ en } H_i \quad (i=1,\ldots,k)$$

5.6.1. Caso univariante

Sea $Y=\sum_h a_h Y_h$ una v.a. de E. La distribución de Y en la celda o población H_i es normal (univariante). Planteemos la hipótesis nula relativa a la v.a. Y,

$$H_{oY}: \quad \psi_1^*(Y) = \ldots\ldots = \psi_s^*(Y)$$

Teorema 5.6.1

Sea $\sigma^2 = \mathrm{var}(Y)$

1) La v.a.:

$$\frac{\mathcal{L}(\hat{\beta})(Y)}{\sigma^2} = \frac{\min \mathcal{L}(\beta)(Y)}{\sigma^2}$$

sigue una distribución ji-cuadrado con n-r grados de libertad.

2) Bajo H_{oY}, la v.a.:

$$\frac{1}{\sigma^2}\mathcal{L}(\hat{\theta})(Y) = \frac{1}{\sigma^2}\mathcal{L}(\hat{\beta})(Y)_{H.\beta=0}$$

sigue una distribución ji-cuadrado con n-t=n-s+1 g. de libertad.

3) Bajo H_{oY}, las v.a.:

$$\mathcal{L}(\hat{\beta})(Y) \quad y \quad (\mathcal{L}(\hat{\theta})(Y) - \mathcal{L}(\hat{\beta})(Y))$$

son estocásticamente independientes.

Demostración:

La distribución conjunta de las v.a.:

$$Z^1,\ldots,Z^t,Z^{t+1},\ldots,Z^r,Z^{r+1},\ldots,Z^n \qquad (t=s-1)$$

introducidas en § 5.4, es normal multivariante, y, puesto que son incorrelacionadas, tales v.a. serán estocásticamente independientes. Según § 5.2, tenemos:

$$\mathcal{L}(\hat{\beta})(Y) = \sum_{i=r+1}^{n} (Z^i)^2, \quad \mathcal{E}(Z^i) = 0$$

$$\text{var}(Z^i) = \sigma^2 \qquad i=r+1,\ldots,n.$$

y de aquí deducimos 1) Análogamente, bajo H_0

$$\mathcal{L}(\hat{\theta})(Y) = \sum_{i=t+1}^{n} (Z^i)^2, \quad \mathcal{E}(Z^i) = 0$$

$$\text{var}(Z^i) = \sigma^2 \qquad i=t+1,\ldots,n.$$

Finalmente,

$$\mathcal{L}(\hat{\theta})(Y) - \mathcal{L}(\hat{\beta})(Y) = \sum_{i=t+1}^{n} (Z^i)^2$$

es una v.a. en la que no aparece ninguna Z^i que esté también en $\mathcal{L}(\hat{\beta})(Y)$, luego es independiente de $\mathcal{L}(\hat{\beta})(Y)$.

c. q. d.

Corolario 5.6.1

Bajo la hipótesis H_0, la v.a.:

$$F_Y = \frac{(\mathcal{L}(\hat{\theta})(Y) - \mathcal{L}(\hat{\beta})(Y))/(r-s+1)}{\mathcal{L}(\hat{\beta})(Y)/(n-r)}$$

sigue una distribución F, con $(r-s+1)$ y $(n-r)$ g. de l.

Elegido un nivel de significación α, el test se resuelve hallando F_α tal que

$$P(F > F_\alpha) = \alpha, \quad F \text{ con } (r-s+1) \text{ y } (n-r) \text{ g. de l.}$$

Rechazaremos H_{oY}, si $F_Y > F_\alpha$, con probabilidad α de equivocarnos, y aceptaremos H_{oY}, si $F_Y \leq F_\alpha$.

5.6.2. Caso multivariante

De acuerdo con las métricas de matrices asociadas R_o y R_1 (§5.5), pongamos:

$$\mathcal{L}(\hat{\beta})(Y) = \|Y\|_o^2 \qquad \mathcal{L}(\hat{\theta})(Y) = \|Y\|_1^2$$

Planteemos ahora la hipótesis nula:

$$H_o : \quad \psi_1^* = \ldots = \psi_s^*$$

Se verifica:

$$H_o = \bigcap_{Y \in E} H_{oY}$$

es decir, aceptar H_{oY}, para toda Y, equivale a aceptar H_o (principio de unión-intersección de ROY, 1957).

Para una v.a. $Y \in E$, sea:

$$\lambda(Y) = \frac{\|Y\|_1^2}{\|Y\|_o^2}$$

El valor F para esta v.a. es:

$$F_Y = \frac{(\|Y\|_1^2 - \|Y\|_o^2)/(r-s+1)}{\|Y\|_o^2/(n-r)} = (\lambda(Y)-1)\frac{(n-r)}{(r-s+1)}$$

Aceptaremos H_{oY} si $F_Y \leq F_\alpha$, y aceptaremos H_o si esto es cierto para toda $Y \in E$. Busquemos pues, para toda $Y \in E$,

$$F = \text{máx } F_Y = (\lambda_1 - 1)\frac{n-r}{r-s+1}$$

siendo $\lambda_1 = \text{máx} \dfrac{\|Y\|_1^2}{\|Y\|_o^2}$

Según el teorema fundamental, λ_1 es la mayor raiz de la ecuación matricial

$$(5.6.1) \qquad \det (R_1 - \lambda R_o) = 0$$

En principio, pues, aceptamos H_o si

$$F = \text{máx } F_Y \leq F_\alpha \ .$$

En realidad F, que es función de λ_1, no sigue la distribución F de Fisher, si lo consideramos pendiente del muestreo sobre las p v.a. (excepto si $p \leq 2$). Para resolver correctamente el test, deberíamos conocer la distribución exacta de λ_1.

Se puede resolver también utilizando todas las raices de la ecuación (5.3.1), definiendo el estadístico

$$\Lambda = \frac{\det(R_o)}{\det(R_1)} = (\lambda_1 \cdot \lambda_2 \ \ldots \lambda_p)^{-1}$$

que sigue una distribución $U_{p,r-t,n-r}$ (WILKS, 1932), siendo p el nº de v.a., r-t y n-r el nº de grados de libertad.

En efecto: (véase RAO, 1965) Λ es el cociente de determinantes de dos matrices con distribuciones de Wishart independientes.

$$R_o \text{ es } W(n-r, \Sigma)$$

$$R_1 - R_o \text{ es } W(r-t, \Sigma) \qquad R_1 = (R_1 - R_o) + R_o \text{ es } W(n-t, \Sigma)$$

La distribución de Λ coincide con la distribución del producto de variables aleatorias independientes con distribución beta.

Bajo H_o, y en algunos casos particulares, la distribución de Λ coincide con la distribución F de Fisher:

$(t=s-1)$

$r-t=1 \qquad F = \frac{1-\Lambda}{\Lambda}\,\frac{n-t-p}{p} \qquad p,\ n-t-p$ g. de l.

$r-t=2 \qquad F = \frac{1-\sqrt{\Lambda}}{\sqrt{\Lambda}}\,\frac{n-t-p-1}{p}, \qquad 2p, 2(n-p-t-1)$ g. de l.

$p=1 \qquad F = \frac{1-\Lambda}{\Lambda}\,\frac{n-r}{r-t} \qquad r-t,\ n-r$ g. de l.

$p=2 \qquad F = \frac{1-\sqrt{\Lambda}}{\sqrt{\Lambda}}\,\frac{n-r-1}{r-t} \qquad 2(r-t), 2(n-r-1)$ g. de l.

(véase ANDERSON, 1958).

Para los demás casos, puede utilizarse la propiedad asintótica (RAO, 1951) del estadístico:

$$F = \frac{1-\Lambda^{1/s}}{\Lambda^{1/s}}\,\frac{ms-2\lambda}{p(r-t)}$$

siendo:

$$m = n-t-\frac{p+r-t+1}{2}$$

$$s = \sqrt{\frac{p^2(r-t)^2-4}{p^2+(r-t)^2-5}}$$

$$\lambda = \frac{p(r-t)-2}{4}$$

que, bajo H_o, converge en ley a la distribución F con

$p(r-t),\ ms-2\lambda$ grados de libertad

cuando n tiende ainfinito.

Respecto a la potencia del test, véase, por ejemplo, ROY(1966).

6 ANALISIS CANONICO DE POBLACIONES

Presentamos en este capítulo el análisis canónico de RAO (1952), basándonos en la exposición de COOLEY y LOHNES (1962,1971) y SEAL (1964). Las notaciones se apoyan en el álgebra lineal, siguiendo la tónica de los capítulos anteriores.

Se incluye una aportación que puede ser interesante: la obtención de una región confidencial exacta para representar a cada una de las poblaciones en el espacio canónico. Puede sustituir con ventaja a la región aproximada que se viene utilizando (SEAL,1964,pág.137).

Se ha consultado también: ANDERSON(1958), DEMPSTER(1969), FISHER (1936,1950), GOODMAN(1972), HOPE(1972), KENDALL(1957), MORRISON(1967).

6.1. PARAMETRIZACIONES

Cuando el sistema de funciones estimables son los k individuos medios de las k poblaciones, podemos introducir m=k parámetros tales que,

$$m_1^* = \beta_1^*, \ldots, m_k^* = \beta_k^*$$

La matriz del diseño factorial X es entonces la matriz identidad. Tendremos:

$$(6.1.1) \qquad m_i^*(Y) = 0.\beta_{Y1} + \ldots + 1.\beta_{Yi} + \ldots + 0.\beta_{Ym} \qquad Y \in E$$

En particular, para las v.a. Y_j ,

$$m_i^*(Y_j) = \beta_{ji} \qquad i=1,\ldots,k; \quad j=1,\ldots,p.$$

Con la parametrización (6.1.1), toda función paramétrica es estimable.(Corolario 4.3.2).

Puede utilizarse también otra parametrización:

$$(6.1.2) \qquad m_i^*(Y) = \mu + \alpha_i \qquad i=1,\ldots,k \qquad Y \in E$$

en la que intervienen m=k+1 parámetros, representando μ la <u>media general</u>, y α_i el <u>efecto de grupo</u> de la variable Y, población i.

En esta parametrización, una función estimable será:

$$\psi^* = \sum_i c_i m_i^* = \sum_i c_i \sum_j (\mu_j + \alpha_{ji}) Y_j^* =$$

$$= \sum_j ((\sum_i c_i)\mu_j + \sum_i c_i \alpha_{ji}) Y_j^*$$

es decir, para que una función paramétrica sea estimable es necesario y suficiente que los coeficientes que afectan a cada μ_j, sean suma de los k coeficientes que afectan a los efectos de grupo α_{ji}.

La parametrización (6.1.2) será generalizada en el próximo capítulo. Aquí nos interesa analizar la estructura de los k individuos medios, en el caso dim M=k, que en particular son funciones paramétricas estimables

Utilizaremos, para este fin, la parametrización (6.1.1), en la que rang (X) = m = k = dim M.

6.2. COMPARACION DE INDIVIDUOS MEDIOS

Aplicando la teoría general del capítulo anterior, expondremos, brevemente, la estimación de Σ, y el test estadístico para contrastar la hipótesis nula

$$H_o: \quad m_1^* = m_2^* = \dots = m_k^*$$

La hipótesis nula, planteada a través de los parámetros, es

$$H_o': \quad \beta_{j1} = \dots = \beta_{jk} \qquad (j=1,\dots,p)$$

Para un muestreo de tamaño $\Delta = (n_1,\dots,n_k)$, indiquemos:

$$mY_j = (y_{j11},\dots,y_{j1n_1};\dots;y_{jk1},\dots,y_{jkn_k})$$
$$j=1,\dots,p$$

Cada individuo paramétrico es función estimable, y la estimación de β_{Y_j} es:

$$\hat{\beta}_{Y_j} = (\hat{\beta}_{j1},\dots,\hat{\beta}_{jm})$$

siendo

$$\hat{\beta}_{jt} = \bar{y}_{jt.} = \frac{1}{n_t} \sum_{h=1}^{n_t} y_{jth}$$

La matriz ampliada del diseño factorial es:

$$X_a = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 1 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & 1 \\ \vdots & & \vdots & \vdots \\ 0 & \cdots & 0 & 1 \end{pmatrix} \begin{matrix} \left.\vphantom{\begin{matrix}1\\ \vdots\\ 1\end{matrix}}\right\} n_1 \\ \\ \left.\vphantom{\begin{matrix}1\\ \vdots\\ 1\end{matrix}}\right\} n_k \end{matrix} \qquad \text{rang}X_a = k = m$$

de donde

$$R_o(i,j) = (mY_j - X_a \hat{\beta}_{Y_i})'(mY_j - X_a \hat{\beta}_{Y_j}) =$$

$$= \sum_{t=1}^{k} \sum_{h=1}^{n_t} (y_{ith} - \bar{y}_{it.})(y_{jth} - \bar{y}_{jt.})$$

Si es válida la hipótesis H_o, entonces

$$\beta_{j1} = \cdots = \beta_{jm} = \beta_j$$

y la estimación de β_j es

$$\hat{\beta}_j = \bar{y}_{j..} = \frac{1}{n} \sum_t \sum_h y_{jth}$$

de donde,

$$R_1(i,j) = \min_{H_o}(mY_i - X_a \beta_{Y_i})(mY_j - X_a \beta_{Y_j}) =$$

$$= \sum_{t=1}^{k} \sum_{h=1}^{n_t} (y_{ith} - \bar{y}_{i..})(y_{jth} - \bar{y}_{j..})$$

Finalmente, obtenidos los valores propios de R_1 respecto a R_o, resolviendo,

$$\det(R_1 - \lambda R_0) = 0$$

se calcula el estadístico F tal como se explica en §5.6

6.3. COVARIABILIDAD ENTRE POBLACIONES

Supongamos dim M = k, es decir, los k indivíduos medios son M-linealmente independientes. Con un nivel de significación α, una condición necesaria es haber podido rechazar la hipótesis nula H_o.

Vamos a introducir, de una forma natural, una nueva métrica en el e.v. E, que refleje la covariabilidad de las variables debida a las diferencias entre los indivíduos medios.

Basándonos en que cada celda o población queda definida por el correspondiente indivíduo medio, intriducimos un espacio de probabilidades discreto $(\Omega_M, \mathcal{A}_M, P_M)$, siendo:

$$\Omega_M = \{ m_1, \ldots, m_k \}$$

$$\mathcal{A}_M = \mathcal{P}(\Omega_M)$$

$$P_M(m_i) = P(H_i) = p_i \qquad i=1,\ldots,k.$$

Este espacio es isométrico al subespacio de $(\Omega, \mathcal{A}, P)$ generado por las k celdas o poblaciones H_i $(i=1,\ldots,k)$.

De una forma natural podemos considerar que una v. a. Y de E, alcanza los siguientes valores sobre los átomos de $\mathcal{A}_M$,

$$Y(m_1) = \beta_{Y1}, \ldots, Y(m_k) = \beta_{Yk}$$

El producto escalar que proporciona la covarianza será:

$$Y \circ Y' = \mathrm{cov}_M(Y, Y') = \sum_{h=1}^{k} (\beta_{Yh} - \bar{\beta}_Y)(\beta_{Y'h} - \bar{\beta}_{Y'}) p_h \tag{6.3.1}$$

siendo

$$\bar{\beta}_Y = \sum_{h=1}^{k} p_h \beta_{Yh} = \mathcal{E}_M(Y)$$

la esperanza matemática en Ω_M.

El producto escalar (6.3.1) es un nuevo producto escalar definido sobre el espacio E. Indicaremos por Σ_E a la matriz de este producto escalar referido a la base $<Y_1,\ldots,Y_p>$. Se puede intróducir, también por otro camino.

Sea $\mathcal{F}_M$ el espacio vectorial, de dimensión k, generado por $(m_1,\ldots,m_k)$, considerados como átomos de $\mathcal{A}_M$.

Definamos en $\mathcal{F}_M$ un producto escalar de matriz

$$M_m = \begin{pmatrix} p_1 & & \\ & \ddots & \\ & & p_k \end{pmatrix}$$

según la cual, la base de $\mathcal{F}_M$ es ortogonal.

Sea entonces la aplicación

$$f_m: \quad E \longrightarrow \mathcal{F}_M$$

$$f_m(Y) = (\beta_{Y1}-\bar{\beta}_Y)m_1 + \ldots + (\beta_{Yk}-\bar{\beta}_Y)m_k$$

La matriz Σ_E es la matriz del producto escalar en $\mathcal{F}_M$ de los p vectores $f_m(Y_1),\ldots,f_m(Y_p)$.

Indicando:

$$B = \begin{pmatrix} \beta_{11}-\bar{\beta}_1, \ldots, \beta_{p1}-\bar{\beta}_p \\ \ldots\ldots\ldots \\ \beta_{1k}-\bar{\beta}_1, \ldots, \beta_{pk}-\bar{\beta}_p \end{pmatrix}$$

se tiene:

$$(6.3.2) \qquad \Sigma_E = B'M_m B$$

El producto escalar (6.3.1), de matriz asociada (6.3.2) en la base de E, refleja la covariabilidad de las v.a. de E debida a las diferencias <u>entre</u> los grupos o poblaciones. Por otra parte, la matriz Σ representa la covariabilidad que existe <u>dentro</u> de cada población.

Teorema 6.3.1

Supongamos dim M = k.

1) Si es cierta la hipótesis nula H_o, entonces $\Sigma_E = 0$.

2) Si podemos rechazar la hipótesis nula, y los individuos medios son funciones paramétricas M-linealmente independientes, se verifica:

$$\text{rang}(\Sigma_E) = \text{mín}\ (p, k-1)$$

Demostración:

El rango de Σ_E coincidirá con el rango de B.

Si H_o es cierta, $\beta_{ji} = \bar{\beta}_i$, y entonces B=0, lo que implica $\Sigma_E = 0$.

El rango por columnas de B, es p como máximo. Sumando las filas previamente multiplicadas por $p_1, p_2, \ldots, p_k$, detectamos una combinación lineal entre las mismas, es decir, el máximo rango por filas es (k-1).

c. q. d.

6.4. ESTIMACION DE Σ Y Σ_E

Aplicando directamente la teoría general de §5.2 al modelo paramétrico (6.1.1), tendremos que los elementos de la matriz

$$(6.4.1) \qquad \hat{\Sigma} = \frac{1}{n-k} R_o$$

son estimadores insesgados de los elementos de Σ. La matriz R_o ha sido obtenida en §6.2.

La matriz Σ_E depende, por una parte, de los parámetros desconocidos $\beta_{i1}, \ldots, \beta_{ik}, \bar{\beta}_i$ (i=1,...,p), y por otra de las probabilidades, en la práctica también desconocidas, $p_1, \ldots, p_k$.

Ya hemos visto que $\tilde{y}_{it.}$ es el estimador lineal insesgado de varianza mínima de β_{it}. Es bien conocido que la misma propiedad tiene la frecuencia relativa n_t/n como estimación de la probabilidad p_t $(t=1,\ldots,p)$. Finalmente, para estimar

$$\tilde{\beta}_i = \sum_{t=1}^{k} p_t \beta_{it}$$

podremos utilizar el estimador

$$\sum_{t=1}^{k} \frac{n_t}{n} \bar{y}_{it.} = \bar{y}_{i..} = \frac{1}{n} \sum_t \sum_h y_{ith}$$

Con todo esto, la matriz $\hat{\Sigma}_E$, cuyos elementos son

$$(6.4.2) \qquad \hat{\sigma}_{Eij} = \sum_{t=1}^{k} \frac{n_t}{n} (\bar{y}_{it.} - \bar{y}_{i..})(\bar{y}_{jt.} - \bar{y}_{j..})$$

es una buena estimación de Σ_E.

6.5. EJES CANONICOS

Pasemos a diferenciar los individuos medios. En principio, parece que bastaría un análisis de componentes principales sobre la matriz Σ_E, perodado que la métrica de matriz Σ también influye en las variables, deberemos encontrar unas variables que expliquen al máximo la variabilidad de Σ_E, al tiempo que minimicen la influencia de Σ.

Sea nv= mín(p,k-1) = rang(Σ_E). Llamaremos <u>variables canónicas</u> a las v.a. $V_1,\ldots,V_{nv}$ de E tales que:

1) Son simultáneamente ortogonales respecto a Σ y Σ_E

$$V_i \cdot V_j = 0 \qquad V_i \circ V_j = 0 \qquad i,j=1,\ldots,nv \quad i \neq j$$

2) Maximizar los cocientes de los módulos al cuadrado

$$\frac{V_1 \circ V_1}{V_1 \cdot V_1} = \lambda_1 = \text{máx}, \ldots, \frac{V_{nv} \circ V_{nv}}{V_{nv} \cdot V_{nv}} = \lambda_{nv} = \text{máx}.$$

Según el teorema fundamental, el problema de determinar las variables canónicas se resuelve obteniendo los vectores propios y valores propios de Σ_E respecto a Σ

$$(6.5.1) \qquad \det(\Sigma_E - \lambda\Sigma) = 0, \qquad \Sigma_E v_i = \lambda_i \Sigma v_i \qquad i=1,\ldots,nv$$

Para determinar el módulo de tales vectores, podemos exponer 2) en la forma equivalente

$$V_1 \circ V_1 = \lambda_1 = \text{máx}, \ldots, V_n \circ V_{nv} = \lambda_{nv} = \text{máx}.$$

$$\|V_1\|_\Sigma = \ldots = \|V_{nv}\|_\Sigma = 1$$

es decir, respecto a Σ, el sistema de vectores canónicos es ortonormal.

Las d primeras v.a. canónicas serán las direcciones vectoriales que reflejarán la máxima covariabilidad entre las poblaciones. Llamaremos <u>espacio vectorial canónico</u> EC_d, de dimensión d, al subespacio de E generado por estas d primeras v.a. canónicas,

$$EC_d = \langle V_1, \ldots, V_d \rangle$$

Consideremos ahora el subespacio F^*_{p-d} de E^*

$$F^*_{p-d} = \{ \omega^* \in E^*, \omega^*(V) = 0, \ \forall V \in EC \}$$

Recordando que en E^* hay una métrica, de matriz asociada Σ^{-1} en la base dual (§2.2), llamaremos <u>espacio dual canónico</u> al espacio $EC^*_d = (F^*_{p-d})^{\perp}$, ortocomplementario de F^*_{p-d} en E^* Se tiene:

$$E^* = EC_d^* \oplus F_{p-d}^*$$

A los elementos de la base de EC_d^*,

$$EC_d^* = \langle V_1^*, \ldots, V_d^* \rangle, \qquad \|V_i^*\|_{\Sigma^{-1}} = 1$$

tales que

$$EC_1^* = \langle V_1^* \rangle \quad EC_2^* = \langle V_1^*, V_2^* \rangle \quad \ldots \quad EC_d^*$$

les llamaremos <u>ejes canónicos</u>.

La <u>inyección</u>,

(6.5.2) $$EC_d \xrightarrow{i} E$$

nos definirá una proyección

$$\pi : E^* \longrightarrow EC_d^*$$

(6.5.3) $$\pi(\omega^*) = \omega_1^* \qquad \text{si} \qquad \omega^* = \omega_1^* + \omega_2^*$$
$$\omega_1^* \in EC_d^*, \; \omega_2^* \in F_{p-d}^*$$

que llamaremos <u>proyección canónica</u>.

Sea V la matriz cuyos vectores columna son las componentes en la base $\{Y_i\}$ de las d primeras variables canónicas. La imagen de ω^*, con componentes $W = (w_1, \ldots, w_p)'$ en la base dual $\{Y_i^*\}$, por la proyección canónica, será ω_1^*, y cuyas componentes $\bar{W} = (\bar{w}_1, \ldots, \bar{w}_d)$ en los ejes canónicos se obtendrán de

(6.5.4) $$\bar{W} = V'W$$

6.6. REPRESENTACION CANONICA

Aplicando la proyección canónica (6.5.3) a cada uno de los individuos medios $m_1^*, \ldots, m_k^*$, obtendremos k puntos en el espacio dual canónico EC_d^*

$$m_{c1}^* = \pi(m_1^*), \ldots, m_{ck}^* = \pi(m_k^*)$$

que se pueden representar tomando como origen

$$\overline{m}^*_o = \eta(\overline{m}^*) \quad \text{siendo} \quad \overline{m}^* = \sum_{h=1}^{k} p_i m^*_i$$

el centroide ponderado de los k individuos medios.

Para la representación gráfica se tomará, naturalmente, $d=1,2,3$ y en ella quedará reflejada la parte más importante de las diferencias entre los individuos medios.

Supongamos ahora que la distribución conjunta de las v.a. $(Y_1,\dots,Y_p)$ es normal multivariable,

$$N((\beta_{1j},\dots,\beta_{pj})',\Sigma) \quad \text{en } H_j \quad (j=1,\dots,k)$$

Consideremos el suceso $(S_j, H_j) \in \mathcal{B}^p \circ \mathcal{A}_H$, siendo

$$S_j = \{ i^* \in E^*, \| i^* - m^*_j \| \leq R_\alpha \}$$

La distancia $\| i^* - m^*_j \|$ viene dada por la métrica en E de matriz asociada Σ^{-1} (§ 2.2). S_j es un <u>elipsoide</u> en E^*.

A este suceso le corresponde, por la isometría del teorema 2.3.1, el suceso $\tilde{B}_j \subset H_j/\mathbb{N}$, siendo

$$\tilde{P}_{H_j}(\tilde{B}_j) = P_{F_j}(S_j) \qquad j=1,\dots k$$

Si esta probabilidad es elevada, $\tilde{B}_j$ representa la "mayor parte" de la población H_j, y por la aplicación δ, su representación multivariante es $S_j \subset E^*$(§ 2.3, § 2.5). Ocupémonos ahora de la representación canónica de S_j.

<u>Teorema 6.6.1</u>

Sea $\dim EC_d = \dim EC^*_d = d$. La probabilidad del suceso

$$S_j = \{ i^* \in E^*, \| i^* - m^*_j \| \leq R_\alpha \} \qquad j=1,\dots,k$$

es $P(S_j) = 1-\alpha$, siendo α y R_α tales que $P(\chi^2_p > R^2_\alpha) = \alpha$, y χ^2_p una v.a. con distribución ji-cuadrado, con p g. de l.

La imagen de S_j por la proyección canónica,

$$\pi(S_j) = \{ i_c^* \in EC_d^*, \| i_c^* - m_{cj}^* \| \leq R_\alpha \} \qquad j=1,\ldots,k$$

es una <u>esfera</u> en EC_d^*, de centro m_{cj}^* y radio R_α.

<u>Demostración</u>:

La aplicación:

$$G: E^* \longrightarrow R$$

$$G(i^*) = \| i^* - m_j^* \|^2 \qquad i^* \in E^*$$

define una v.a. sobre E^*, que sigue la distribución ji-cuadrado, con p grados de libertad. Luego,

$$P(S_j) = P(G \leq R_\alpha^2) = 1-\alpha$$

siendo $P(\chi_p^2 > R_\alpha^2) = \alpha$

(χ_p^2, ji-cuadrado con p g. de l.)

La imagen de S_j por la proyección canónica π, es una <u>ésfera</u> en EC_d^*, porque estará referido a la base canónica que es ortonormal.

c. q. d.

Utilizando las estimaciones $\hat{m}_1^*,\ldots,\hat{m}_k^*$ de los indivíduos medios, y la estimación $\hat{\Sigma}$ de la matriz Σ, el teorema anterior nos da una forma sencilla, pero <u>aproximada</u>, de representar los indivíduos de cada una de las poblaciones, mediante esferas de radio común R_α, en el sentido de que un indivíduo de H_j, pertenece a la esfera $\pi(S_j) \subset EC_d^*$, con probabilidad $(1-\alpha)$.

Seguidamente, determinaremos unas regiones confidenciales, esta vez <u>exactas</u>, para los k indivíduos medios.

Necesitaremos el siguiente resultado:

Teorema 6.6.2

Sea γ un vector aleatorio de dimensión p, con distribución $N(0,\Sigma)$, estocásticamente independiente de nS, en donde S es una matriz aleatoria, con la misma distribución que

$$\sum_{\alpha=1}^{n} Z_\alpha Z'_\alpha$$

siendo los Z_α también p vectores aleatorios estocásticamente independientes, con distribución $N(0,\Sigma)$. Dada entonces la v.a.:

$$T^2 = \gamma' S^{-1} \gamma$$

la distribución de

$$\frac{T^2}{n} \frac{n-p+1}{p}$$

sigue una F de Fisher con p y (n-p+1) grados de libertad

La demostración de este teorema, puede verse en ANDERSON (1958, pág. 106).

Teorema 6.6.3

El conjunto de EC^*_d (de dimensión d), referido a los ejes canónicos,

$$R(\hat{m}^*_{cj}) = \{ i^*_c \mid \| i^*_c - \hat{m}^*_{cj} \| \leq \frac{R_\alpha}{\sqrt{n_j}} \}$$

es una esfera de radio $R_\alpha / \sqrt{n_j}$ y centro $\hat{m}^*_{cj}$, que define una región confidencial para m^*_{cj}, con probabilidad $(1-\alpha)$

$$P\left[m^*_{cj} \in R(\hat{m}^*_{cj}) \right] = 1-\alpha$$

siendo R_α tal que

$$R^2_\alpha = F_\alpha \frac{(n-k)p}{(n-k-p+1)}$$

y $P(F > F_\alpha) = \alpha$, en donde F es una v.a. que sigue una distribución de Fisher con p y (n-k-p+1) g. de l.

Demostración:

De una muestra de tamaño $\Lambda : (n_1, \ldots, n_k)$, indiquemos por:

$$\mathcal{Y}_{ih} = (Y_{1ih}, \ldots, Y_{pih})'$$

el vector aleatorio de los p valores muestrales, muestra h, población Π_i, y por:

$$\bar{\mathcal{Y}}_i = (\bar{Y}_{1i}, \ldots, \bar{Y}_{pi})'$$

el vector aleatorio de medias muestrales en la población Π_i. Entonces:

$$(n-k)\,\hat{\Sigma} = \sum_{i=1}^{k} \sum_{h=1}^{n_i} (\mathcal{Y}_{ih} - \bar{\mathcal{Y}}_i)(\mathcal{Y}_{ih} - \bar{\mathcal{Y}}_i)'$$

y se verifica la identidad:

$$\sum_{i=1}^{k} \sum_{h=1}^{n_i} \mathcal{Y}_{ih}\mathcal{Y}'_{ih} = (n-k)\left(\mathcal{Y}_1 \cdot \mathcal{Y}'_1 + \ldots + \mathcal{Y}_k \cdot \mathcal{Y}'_k + \hat{\Sigma} \right)$$

Supongamos ahora que la esperanza de cada una de las v.a. $(Y_1, \ldots, Y_p)$ es cero. Esto no altera los elementos de la matriz Σ, que son formas cuadráticas de rango $(n-k)$ Asimismo, el rango de cada $\mathcal{Y}_i \cdot \mathcal{Y}'_i$ es uno. Puesto que:

$$n_1 + \ldots + n_k = (n-k) + (\overset{k}{1 + \ldots + 1})$$

por el teorema de Cochran, generalizado al análisis multivariante, (v. ANDERSON, 1958, pág. 165), $(n-k)\hat{\Sigma}$ es independiente de cada $\mathcal{Y}_i \cdot \mathcal{Y}'_i$, es decir, de $\mathcal{Y}_i$, y sigue la distribución

$$\sum_{\alpha=1}^{n-k} Z_\alpha \cdot Z'_\alpha$$

en donde los Z_α son p vectores aleatorios independientes, y cada uno de ellos con distribución $N(0,\Sigma)$.

Entonces, el vector aleatorio