

ANÁLISIS DISCRIMINANTE DE MEDIDAS REPETIDAS

Emma Cambillo Moyano 
Universidad Nacional Mayor de San Marcos

RESUMEN

El análisis discriminante y el análisis de medidas repetidas han sido estudiados desde muchos años atrás. Este trabajo trata del estudio de las reglas de clasificación de máxima verosimilitud suponiendo distribución multinormal con matrices de covarianza iguales dentro de los grupos y en medidas repetidas.

La reglas de clasificación son derivadas para el caso de 'g' grupos y particularizadas para dos grupos en los datos originales y en transformaciones a coeficientes polinomiales.

Los resultados obtenidos ayudan a un mejor entendimiento del análisis discriminante en este contexto.

I. INTRODUCCIÓN

El análisis discriminante es una técnica dedicada a investigar que tanto es posible distinguir entre individuos de varios grupos, con observaciones hechas sobre ellos. En la fase de clasificación nos lleva a reglas bien definidas las cuales son utilizadas para clasificación de nuevos individuos que proceden de uno de los grupos.

El interés principal de este trabajo es la fase de clasificación investigando las reglas de máxima verosimilitud en medidas repetidas sobre hipótesis de multinormalidad y matrices de covarianza común dentro de los grupos, esto porque cuando los datos tienen estructura de medidas repetidas las reglas de clasificación pueden revelar donde está el poder de separación de los grupos.

Una estructura de medidas repetidas se refiere a datos constituidos de valores de una misma variable aleatoria, observadas sobre el mismo individuo en ocasiones distintas. Estos individuos provienen de "g" grupos o fueron colocados en los "g"

grupos según el carácter experimental u observacional del problema; en estos casos los objetivos y las interpretaciones son diferentes sin embargo para el desarrollo de la metodología no serán distinguidos.

En el caso general de vectores multivariantes las interrelaciones entre las variables componentes del vector son tratadas a través de su correspondiente matriz de covarianza, en este caso específico en que los datos provienen de medidas repetidas, tenemos una clase también específica de vector multivariante, lo que nos sugiere estructuras para la matriz de covarianza y dentro de ellas son investigadas las estructuras de la matriz uniforme y seriada.

II. REGLAS DE CLASIFICACIÓN EN LOS DATOS ORIGINALES

En el análisis multivariante de perfiles para "g" grupos, Singer y Andrade (1986), tenemos:

$$E(X) = AE \quad (\text{II-1})$$

$X_{n \times p}$ = Matriz de los p discriminadores en los n individuos colocados en los g grupos.

$A_{n \times g}$ = Matriz de diseño.

$E_{g \times p}$ = matriz de parámetros.

$x = (x_{j\alpha 1}, x_{j\alpha 2}, \dots, x_{j\alpha p})'$ el vector fila de X que corresponde a los discriminadores del α -ésimo individuo en el j-ésimo grupo, según el modelo de medias de Bock (1975), puede ser modelados por:

$$x_{ja} = \tau + \theta_j + \varepsilon_{ja}, \quad j = 1, 2, \dots, g; \alpha = 1, 2, \dots, n_j \quad (\text{II-2})$$

τ , vector de medias a lo largo de las ocasiones.

θ_j = vector de efecto del j-ésimo grupo del cual proviene el individuo

$\varepsilon_{j\alpha}$ = vector de los errores y por hipótesis

$$\varepsilon_{j\alpha} \sim N_p(0, \Sigma) \quad (\text{II-3})$$

y

$$\mu_j = \tau + \theta_j$$

es el vector de medias el j-ésimo grupo y μ es la media general.

La regla de clasificación de máxima verosimilitud, Mardia et al (1979), para dos grupos:

Asignar x (nueva observación) al primer grupo si

$$d' \Sigma^{-1}(x - \mu) > 0 \quad (\text{II-4})$$

$$d = \theta_1 - \theta_2$$

$$\mu = \frac{\mu_1 + \mu_2}{2}$$

y asigna x al segundo grupo en caso contrario.

dependiendo por un lado de la magnitud de la diferencia de los efectos de grupo y por otro lado de las desviaciones con relación a la media general de los grupos a través de la matriz de covarianza.

Debido a la existencia de algunas estructuras frecuentes para la matriz de covarianza en este contexto entre ellas la uniforme e.g. Geisser (1963) y la de correlación serial, e.g. Cole y Grizzle (1963) la regla de clasificación será estudiada en estos dos casos

MATRIZ UNIFORME

$$\Sigma = \sigma^2 \mathbf{I} + \sigma_\alpha^2 \mathbf{1}\mathbf{1}' \quad (\text{II-5})$$

σ_α^2 , es la variación dentro del individuo a lo largo de las ocasiones

σ^2 , es la variación entre individuos dentro de cada ocasión.

- a) Cuando las diferencias de los efectos de los grupos varían de ocasión a ocasión es decir los perfiles no son paralelos, la regla de clasificación (II.4) es dada por,

Asigna x al primer grupo si,

$$\sigma^2 \left(\sum_{k=1}^p \left[d_k - \frac{p\sigma_\alpha^2}{\sigma^2 + p\sigma_\alpha^2} \bar{d} \right] [x_k - \mu_k] \right) > 0 \quad (\text{II-6})$$

y asigna x al segundo grupo en caso contrario.

donde μ_k es la media general en la k -ésima ocasión. En este caso la regla de clasificación es el producto escalar usual del vector de desviaciones de la observación con relación a la media general, por el vector de desviaciones de la diferencia de medias con relación a su vector medio corregido por el factor

$$\frac{p\sigma_\alpha^2}{\sigma^2 + p\sigma_\alpha^2}$$

donde la relación

$$\sigma_\alpha^2 = \frac{\rho}{1-\rho} \sigma^2, \sigma^2 = \frac{1-\rho}{\rho} \sigma_\alpha^2$$

y cuando el coeficiente de correlación ρ tiende a uno. Consideremos dos situaciones, primero σ^2 fijo y segundo σ_α^2 fijo. Notamos que en el primer caso σ_α^2 tiende a explotar y en segundo σ^2 tiende a valores pequeños.

En el primer caso (σ^2 fijo) la variación dentro grupo es grande relativamente a la variación entre y dificulta la clasificación, en el segundo caso (σ_α^2 fijo) la variación entre es pequeña relativamente a la variación dentro y facilita la clasificación.

- b) Cuando las diferencias de los efectos de los grupos permanecen constantes de ocasión a ocasión, es decir cuando hay paralelismo, la reglas de clasificación es:

Asigna x al primer grupo si,

$$\frac{\mathbf{d}}{\sigma^2 + p\sigma_\alpha^2} \sum_{k=1}^p (\mathbf{x}_k - \mu_k) \gg 0 \quad (\text{II-7})$$

y asigna x al segundo grupo en caso contrario.

en este caso la regla de clasificación se reduce al producto escalar del vector de diferencias medias (\mathbf{d}) por el de desviaciones de la nueva observación con relación a la media general, el papel de la magnitud de las diferencias está implícito en la media general mediante las componentes del vector de desviaciones de la nueva observación en relación a la media general.

Consideremos dos situaciones de las fuentes de variación entre y dentro fijas, el poder de separación entre los grupos es menor cuando es fijado la variación entre pues la variación dentro aumenta y es mayor cuando la variación dentro es fijada pues la variación entre disminuye.

MATRIZ SERIADA

$$\Sigma = \frac{\sigma_u^2}{1-\rho^2} \begin{bmatrix} 1 & \rho & \rho^2 & \rho^{p-1} \\ \rho & 1 & \rho & \rho^{p-2} \\ \cdot & \cdot & \cdot & \cdot \\ \rho^{p-1} & \rho^{p-2} & \cdot & 1 \end{bmatrix}, |\rho| < 1, \sigma_u^2 > \neq 0 \quad (\text{II-8})$$

donde ρ es la correlación entre dos ocasiones sucesivas, σ_u^2 es la variación entre individuos corregida por los factores que envuelven la estructura de correlación dentro del individuo.

a) Cuando los perfiles no son paralelos, la regla de clasificación es dada por,

Asigna \mathbf{x} al primer grupo si,

$$\sigma_u^2 \left[\left(\sum_{k=2}^{p-1} \mathbf{d}_k - \rho(\mathbf{d}_{k-1} - \rho \mathbf{d}_k + \mathbf{d}_{k+1}) \right) (\mathbf{x}_k - \mu_k) + (\mathbf{d}_1 - \rho \mathbf{d}_2) (\mathbf{x}_1 - \mu_1) + (\mathbf{d}_p - \rho \mathbf{d}_{p-1}) (\mathbf{x}_p - \mu_p) \right] > 0 \quad (\text{II-9})$$

y asigna \mathbf{x} al segundo grupo en caso contrario.

en este caso el producto escalar tiene el vector de desviaciones de las diferencias vecinas corregidas por el factor ρ y en el centro con relación a la expresión,

$$\rho(\mathbf{d}_{k-1} - \rho \mathbf{d}_k + \mathbf{d}_{k+1})$$

b) Cuando los perfiles son paralelos, la regla de clasificación es,

Asigna \mathbf{x} al primer grupo si,

$$\sigma_u^2 \left[(1 - \rho)^2 \sum_{k=2}^{p-1} (\mathbf{x}_k - \mu_k) + (1 - \rho)(\mathbf{x}_1 - \mu_1) + (\mathbf{x}_p - \mu_p) \right] > 0 \quad (\text{II-10})$$

y asigna \mathbf{x} al segundo grupo en caso contrario.

Para valores altos de ρ disminuye el peso relativo del vector de desviaciones de la nueva observación a la media general, además es de notar, la mayor magnitud del factor de ponderaciones en las componentes extremas del vector de desviaciones.

3. REGLAS DE CLASIFICACIÓN EN LOS COEFICIENTES POLINOMIALES

Las reglas de clasificación son investigadas después de realizado el ajuste de curvas polinomiales esto es por medio de transformaciones de variables en modelos saturados o sea que el grado del polinomio ajustado es el máximo.

En el análisis de curvas de crecimiento, en el caso de "g" grupos Potthoff y Roy (1964), introducen el siguiente modelo,

$$E(\mathbf{X}) = \mathbf{A}\mathbf{E} \quad (\text{III-1})$$

\mathbf{X} y \mathbf{A} definidos como la sección anterior

\mathbf{E}_{gxp} , matriz de los coeficientes asociados a los polinomios naturales

\mathbf{T}_{pxp} , matriz de las potencias de los tiempos.

Ahora

$$\mathbf{x}_{j\alpha} = \mathbf{T}\boldsymbol{\xi}_j + \boldsymbol{\varepsilon}_{j\alpha} \quad (\text{III-2})$$

y

$$x_{j\alpha k} = \xi_{j\alpha} t_k^0 + \xi_{jk} t_k^1 + \dots + \xi_{j(p-1)} t_k^{p-1} + \varepsilon_{j\alpha k}$$

$\boldsymbol{\xi}_j$, vector p-variado de los coeficientes naturales del j-ésimo grupo

$\boldsymbol{\varepsilon}_{j\alpha}$, vector p-variado de los errores y por hipótesis

$$\boldsymbol{\varepsilon}_j \sim \mathcal{N}(0, \Sigma) \quad (\text{III-3})$$

$$\mathbf{x}_{j\alpha} \sim \mathcal{N}(\mathbf{T}\boldsymbol{\xi}_j, \Sigma) \quad (\text{III-4})$$

Sea,

$$\mathbf{T} = \mathbf{P}\mathbf{S}$$

la descomposición de Gramm Schmidt para la matriz de las potencias de los tiempos donde \mathbf{P} es la matriz de los polinomios ortogonales Fisher-Yates (1963), dados en la tabla de Bock (1975) y \mathbf{S} es la matriz triangular superior que establece las relaciones entre coeficientes naturales y ortogonales.

La reparametrización polinomial es dada por,

$$E(\mathbf{x}_{j\alpha}) = \mathbf{P}\mathbf{S}\xi_j = \beta_j; \quad \beta_j = \mathbf{S}\xi_j \quad (\text{III-5})$$

donde β_j es el vector de los coeficientes asociados a los polinomios ortonormalizados al j -ésimo grupo.

Consideremos,

$$\mathbf{y}_{j\alpha} = \mathbf{P}'\mathbf{x}_{j\alpha} \quad (\text{III-6})$$

donde $\mathbf{y}_{j\alpha}$ es el vector de los coeficientes ortonormalizados asociados a las curvas ajustadas al α -ésimo individuos en el j -ésimo grupo.

De (III.5) tenemos,

$$\mathbf{y}_{j\alpha} \sim N(\beta_j, \mathbf{P}'\Sigma\mathbf{P}) \quad (\text{III-7})$$

la regla de clasificación para dos grupos es dada por,

Asigna \mathbf{y} (coeficientes de la curva ajustada al nuevo individuo al primer grupo si,

$$(\beta_1 - \beta_2)'(\mathbf{P}'\Sigma\mathbf{P})^{-1}(\mathbf{y} - \beta^*) > 0, \quad \beta^* = \frac{\beta_1 + \beta_2}{2} \quad (\text{III-8})$$

en caso contrario al segundo grupo.

dependiendo por un lado de la magnitud de la diferencia de los coeficientes ortonormales de los grupos y por el otro lado de las desviaciones de los coeficientes de la curva ajustada al nuevo individuo con relación a la media de los coeficientes ortonormales a través de la matriz de covarianza de las variables transformadas.

En los coeficientes naturales la interpretación de los términos de mayor grado y la estructura de las interrelaciones entre los coeficientes naturales son complicados,

los coeficientes ortonormales además de que las relaciones con los naturales pueden establecerse (III.5), puede haber una gran simplificación de las estructuras de interrelaciones entre los coeficientes ortonormales y las reglas de clasificación como la representación gráfica de los coeficientes ganan en simplicidad.

En el caso uniforme los coeficientes ortonormales son no correlacionados y en el caso seriado los coeficientes de orden par son no correlacionados con los de orden impar.

MATRIZ UNIFORME

a) Cuando los perfiles no son paralelos la reglas de clasificación.

Asigna y (coeficientes de la curva ajustada al nuevo individuo) al primer grupo si,

$$\frac{1}{\sigma^2 + \sigma_\alpha^2} [\beta_{10} - \beta_{20}] (y_0 - \beta_0^*) + \frac{1}{\sigma^2} \sum_{k=1}^{p-1} (\beta_{1k} - \beta_{2k}) (y_k - \beta_k^*) > 0 \quad (\text{III-9})$$

en caso contrario al segundo grupo .

β_{jk} es el coeficiente de los polinomios ortonormales del k-ésimo grado en el j-ésimo grupo y β^* es el vector de medias de los coeficientes,

$$\beta^* = (\beta_1^*, \dots, \beta_k^*)'$$

La regla de clasificación en este caso es el producto escalar común de dos vectores, el de diferencias entre los coeficientes del mismo grado y corregidas por los factores multiplicativos correspondientes y el de las desviaciones entre los coeficientes del nuevo individuo en relación a la media de los coeficientes (mantenida la correspondencia de los grados), los factores de las diferencias entre los coeficientes del mismo grado son los inversos de las varianzas correspondientes a cada uno de ellos.

Si los perfiles no son paralelos la ponderación de la componente relativa al coeficiente del menor grado es menor que las componentes relativas a los coeficientes de mayor grado; comparada esta regla con la obtenida en los datos

originales (III.6) destacamos la posibilidad de un mejor diagnóstico de la separación entre grupos. La reglas de clasificación revela los coeficientes responsables por la clasificación, por tanto hay ganancia en cuanto a interpretación pues a los términos de cada grado pueden corresponder significados diferentes.

b) Cuando los perfiles son paralelos, la regla de clasificación para el caso de dos grupos,

Asigna y al primer grupo si,

$$\frac{1}{\sigma^2 + \sigma_\alpha^2} (\beta_{10} + \beta_{20}) [y_0 - \beta_0^*] \geq 0 \quad (\text{III-10})$$

en caso contrario al segundo grupo .

El poder de separación de los grupos depende del signo de la diferencia en los coeficientes de grado cero e implícitamente de la varianza del término constante que aparece en la variación de y_0 alrededor de β_0^* ,

MATRIZ SERIADA

a) Cuando los perfiles son paralelos la regla de clasificación es ,

Asigna y al primer grupo si,

$$[\beta_{1par} - \beta_{2par}] \Sigma_{\beta par}^{-1} [y_{\beta par} - \beta_{par}^*] + [\beta_{1impar} - \beta_{2impar}] \Sigma_{\beta impar}^{-1} [y_{\beta impar} - \beta_{impar}^*] \geq 0 \quad (\text{III-11})$$

en caso contrario al segundo grupo.

La regla de clasificación es la suma de los productos escalares a través de las correspondientes inversas de las matrices de covarianzas y un examen de (III.11) muestra la dificultad de interpretación, pero casos específicos de polinomios de grado reducido pueden ser investigados.

b) Cuando los perfiles son paralelos, la reglas de clasificación es,

Asigna y al primer grupo si,

$$[\beta_{10} - \beta_{20}] a_{11} [y_0 - \beta_0^*] \geq 0 \quad (\text{III-12})$$

y al segundo grupo en caso contrario, donde,

$$a_{11} = \frac{1}{\sigma_u^2} \left[\sum_{k=1}^p P_0^2(t_k) - 2\rho \sum_{k=1}^{p-1} P_0(t_k)P_0(t_{k+1}) + \rho^2 \sum_{k=2}^{p-1} P_0^2(t_k) \right]$$

La diferencia de los grupos estará evidentemente en el coeficiente de grado cero (término constante) pero es obvio que la correlación con los demás coeficientes pares aparece en la medida que la matriz $\sum \beta_{par}$. Se aleja del patrón original o sea en la medida en que ρ aumenta.

IV. CONCLUSIONES

La ventaja de utilizar los coeficientes en lugar de los datos originales en la clasificación es la reducción de dimensión, pues al ajustar curvas polinomiales lo que se trata es de explicar la variación de los perfiles en función de un menor número de parámetros, esto es llevar en consideración lo que está sucediendo en todas las ocasiones con menor número de parámetros, permitiendo establecer diferencias entre grupos, tanto como cuando son utilizados los datos originales.

Permite una mayor interpretación de los coeficientes que tiene el poder de separación, por ejemplo si los grupos se diferencian por el coeficiente lineal es que tasa de crecimiento medio en los dos grupos son diferentes; si la diferencia está en el coeficiente del término cuadrático la tasa de cambio de la tasa de crecimiento en los dos grupos son diferentes y así sucesivamente.

Una potente aplicación de los resultados en cuanto a la clasificación de un nuevo individuo, puede ser en estudios prospectivos en los cuales los individuos son

agrupados por la ocurrencia o no de un determinado evento al final de período de observación.

BIBLIOGRAFÍA

1. BOCK, R.D. *Multivariate Statistical Methods in Behavioral Research*. Mc. Graw Hill, New York. (1975)
2. COLE, J., GRIZZLE, J. *Applications of multivariate analysis to repeated measurements experiments*. *Biometrics*, 22. 810-828. (1966)
3. FISHER, R.A.; YATES, F. *Statistical tables for Biological Agricultural and Medical Research*. Sixth ED. Hafner New York. (1963)
4. GEISSER, S. *Multivariate Analysis of variance for a special covariance case*. *Journal of the American Statistical*. (1963)
5. MARDIA, K; KENT, J. ; BIBBY, J. *Multivariate Analysis* . Academic Press. London. (1979)
6. POTTHOF, R; ROY, S. *A generalized multivariate analysis of variance model useful especially for growth curve problems*. *Biometrika* 51, 313-326. (1964)
7. SINGER, J.M.; ANDRADE, D.F. *Análisis de datos longitudinales*. VII SINAPE. Simposio Nacional de Probabilidade e Estatística. Campinas, Sao Paulo. (1986)