INSTITUTO DE INVESTIGACIONES EN MATEMATICAS APLICADAS Y EN SISTEMAS UNIDAD ACADEMICA DE LOS CICLOS PROFESIONAL Y DE POSTGRADO DEL C. C. H.
U. N. A. M.

RELACION ENTRE TRES TECNICAS DE ANALISIS MULTIVARIADO

T E S I S

QUE PARA OBTENER EL GRADO DE:

MAESTRA EN ESTADISTICA E INVESTIGACION DE OPERACIONES

PRESENTA LA MATEMATICA:

Martha de Garay y Gómez del Villar

DESEO EXPRESAR MI AGRADECIMIENTO Y ADMIRACION AL DR. ALBERTO CASTILLO POR SU VALIOSA DIRECCION DURANTE EL DESARROLLO DEL PRESENTE TRABAJO, Y NO MENOS POR SU AMIGABLE PACIENCIA PARA CONMIGO. ASI MISMO QUEDO AGRADECIDA CON LOS DRS. IGNACIO MENDEZ Y ALFONSO HERNANDEZ POR SUS ACERTADOS COMENTARIOS Y SUGERENCIAS.

A MIS PADRES

I N D I C E

Introducción.

Las técnicas de análisis multivariado describen datos que dependen no solamente del tamaño de la muestra, sino también del número de variables que se consideran simultáneamente. La descripción se hace en términos de un número relativamente pequeño de parámetros, facilitándose así la interpretación de los datos.

En la presente tesis se hace una revisión de las relaciones que existen entre algunas técnicas de la Estadística Multivariada a saber: regresión, correlación canónica y análisis discriminante.

La regresión multivariada es el método que ayuda a predecir y comprender un fenómeno, expresando un vector como una combinación lineal de otro vector más un error. En la primera parte del trabajo se verá que cuando la distribución conjunta de los dos vectores es normal, el problema de regresión conduce al modelo lineal.

Una forma de medir la correlación entre dos vectores consiste en encontrar combinaciones lineales de ambos de tal manera que la correlación entre ellas sea máxima. En la segunda parte de este trabajo se tratará el método de análisis multivariado que determina estas combinaciones lineales y que se le conoce con el nombre de correlación canónica.

El análisis discriminante es el tercer método que será tratado en este trabajo. Dada la existencia de R poblaciones y dadas R muestras aleatorias de individuos de las cuales se sabe que la R-ésima proviene de la R-ésima población, el análisis discriminante proporciona una regla para asignar de manera óptima otros individuos a una de las R poblaciones de las cuales se sabe proviene pero no exactamente de cuál de ellas.

Usualmente el planteamiento de las técnicas arriba descritas se hace en forma independiente. Aquí se pretende mostrar la parte común a estos métodos, así como algunas semejanzas. Se expondrá la forma específica en la que en cada uno de ellos aparece la matriz de coeficientes de regresión β.

En la primera parte del trabajo se presenta la ecuación de regresión, con el objeto de mostrar que la regresión múltiple no es sino un caso particular de la multivariada.

También se explica la similitud entre regresión múltiple y correlación canónica, al exhibir que la correlación canónica entre un escalar y un vector, coincide con la correlación entre el escalar y la combinación lineal del vector determinada por los coeficientes de regresión. Se muestra además que tal correlación es igual al coeficiente de correlación múltiple. Este hecho permite considerar a la correlación canónica como una forma de generalización de la correlación múltiple.

Con el objeto de vincular el análisis discriminante con los métodos anteriores, se presenta la relación que existe entre la función discriminante de dos poblaciones y el coeficiente de regresión.

Las relaciones antes expuestas pueden parecer teóricas, sin embargo, se muestra que la concordancia tiene sentido práctico en base a varios ejemplos.

Finalmente se muestra que los valores característicos necesarios para determinar la función discriminante, están relacionados con la correlación canónica tanto para el caso de dos poblaciones como para el caso de R poblaciones ($R>2$)

El objeto de presentar las relaciones que existen entre las técnicas del análisis multivariado antes descritas, es el de facilitar al estudiante la comprensión y manejo de éstas.

Se omite la descripción detallada de cada técnica, haciéndose énfasis únicamente en las propiedades más relevantes y en ejemplos que permitan al lector captar los conceptos en forma intuitiva.

1. Regresión y modelo lineal en la normal.

El estudio de relaciones de dependencia entre varias variables se presenta frecuentemente en estudios con fines de predicción. El modelo de regresión lineal múltiple es utilizado para este propósito, por ejemplo, en la predicción de la producción a partir de las exportaciones, la inversión y el gasto público.

En este capítulo se obtendrá la ecuación de regresión y se mostrará que en el caso de que la distribución conjunta de y y x sea normal, el modelo de regresión lineal

$$Y'(1xq) = X(1xm)B(mxq)+\varepsilon(1xq) \qquad (1)$$

está representa do adecuadamente por esta ecuación.

Es importante notar que la ecuación de regresión, definida como la media de la distribución condicional de y dado x, y el modelo de regresión lineal coinciden en el caso de que la distribución conjunta sea normal, pero pueden ser diferentes en otros casos como sucede en el siguiente

Sea $f(x,y) = R\dfrac{\lambda^x\mu^y a^{xy}}{x!y!}$ con $x,y=0,1,\ldots$ y donde $\lambda,\mu,a>0$ y R es una constante definida de tal forma que f(x,y) es una densidad. La densidad marginal de x es entonces

$f(x) = \sum_{y=0}^{\infty} \dfrac{R\lambda^x\mu^y a^{xy}}{x!y!} = \dfrac{R\lambda^x}{x!}\, e^{\mu a^x}$ y la densidad condicional de

Y dado X=x es

$$f(Y|X=x) = \frac{f(x,y)}{f(x)} = \frac{\frac{R\lambda^x\mu^y a^{xy}}{x!y!}}{\frac{R\lambda^x}{x!}e^{\mu a^x}} = \frac{\mu^y a^{xy} e^{-\mu a^x}}{y!} = \frac{e^{-\mu a^x}(\mu a^x)^y}{y!}$$

una Poisson con parámetro μa^x, por lo tanto la ecuación de regresión es $E(Y|X=x) = \mu a^x$ que no es lineal en x y por lo tanto no coincide con el modelo de regresión lineal (1).

Sea $U(p\times 1) \sim N(\mu, A)$ y sean las particiones

$$U = \begin{pmatrix} Y \\ X \end{pmatrix} \begin{matrix} (q\times 1) \\ (m\times 1) \end{matrix}, \quad \mu = \begin{pmatrix} \mu_1^{(1)} \\ \vdots \\ \mu_q^{(1)} \\ \mu_1^{(2)} \\ \vdots \\ \mu_m^{(2)} \end{pmatrix} = \begin{pmatrix} \mu^{(1)} \\ \mu^{(2)} \end{pmatrix} \begin{matrix} (q\times 1) \\ (m\times 1) \end{matrix}, \quad A = \begin{pmatrix} A_{11}(q\times q) & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

donde q+m=p, $\mu^{(1)}$ es la media de Y, $\mu^{(2)}$ la de X, A_{11} la matriz de covarianza de Y, A_{22} la de X y A_{12} la de Y y X. A continuación se ilustra con un ejemplo las particiones de un vector de dimensión 3 y de su correspondiente vector de medias y matriz de covarianza .

Ejemplo.- El maíz mejorado híbrido se distingue del común porque aquél es el resultado de un cruzamiento entre dos tipos de maíz. Dependiendo de la forma en que se suministra agua a la tierra en la que se siembra un cultivo, se distinguen tres tipos de ésta: tierra de riego, de temporal y de jugo o de humedad. La primera es la que normalmente recibe agua suministrada por medios, obras o mecanismos ideados por el hombre, la segunda es aquella que depende para su cultivo del agua de lluvia que cae directamente sobre ella, y la tercera es la que en forma natural y permanente recibe y conserva de fuentes subterráneas humedad suficiente adicional a la de lluvia.

Considérese el vector $U=(u_1,u_2,u_3)'$ cuyas componentes son:

u_1- precio rural máximo por Kg. de maíz mejorado,

u_2- rendimiento máximo en Kg. por hectárea de maíz mejorado en tierra de riego y

u_3- rendimiento máximo en Kg. por hectárea de maíz mejorado en tierra de temporal.

El vector U se distribuye normal con media μ y matriz de covarianza A. A partir de 22 observaciones del vector U tomadas del 5° censo agrícola-ganadero y ejidal, se calculan estimadores

$$\hat{\mu} = \frac{1}{n}\sum_{h=1}^{n} U_h \quad \text{y} \quad S = \frac{1}{n-1}\sum_{h=1}^{n}(U_h-\hat{\mu})(U_h-\hat{\mu})'$$

de la media y la matriz de covarianza, donde n=22. De aquí se obtiene

$$\hat{\mu} = \begin{pmatrix} 1.08 \\ 3660.09 \\ 2114.23 \end{pmatrix} \quad \text{y} \quad S = \begin{pmatrix} .007 & 3.24 & 1.52 \\ & 319065.8 & 67110.55 \\ & & 87907.04 \end{pmatrix}$$

Sea $U = \begin{pmatrix} u_1 \\ \hline u_2 \\ u_3 \end{pmatrix} = \begin{pmatrix} Y \\ X \end{pmatrix} \begin{matrix} (1x1) \\ (2x1) \end{matrix}$ una partición del vector U, entonces

las correspondientes particiones de $\hat{\mu}$ y S son

$$\hat{\mu} = \begin{pmatrix} 1.08 \\ 3660.09 \\ 2114.23 \end{pmatrix} = \begin{pmatrix} u_1 \\ \hline u_2 \\ u_3 \end{pmatrix} = \begin{pmatrix} u^{(1)} \\ u^{(2)} \end{pmatrix} \qquad \text{y}$$

$$S = \begin{pmatrix} .007 & 3.24 & 1.52 \\ & 319065.8 & 67110.55 \\ & & 87907.04 \end{pmatrix} = \left(\begin{array}{c|cc} s_{11} & s_{12} & s_{13} \\ \hline & s_{22} & s_{23} \\ & & s_{33} \end{array}\right) = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}$$

donde $\bar{u}_i = \frac{1}{n} \sum_{h=1}^{n} u_{hi}$ para i=1,2,3, esto es, $u^{(1)}=1.08$ es la media del precio rural máximo por Kg. de maíz mejorado y $u^{(2)} = \begin{pmatrix} 3660.09 \\ 2114.23 \end{pmatrix}$ está formado por las medias de rendimiento máximo en Kg. por hectárea de maíz mejorado en tierra de riego y de temporal.

Recuérdese que $s_{ij} = \frac{1}{n-1} \sum_{h=1}^{n} (u_{hi}-\bar{u}_i)(u_{hj}-\bar{u}_j)$ para i,j=1,2,3, por tanto $S_{11} = \frac{1}{n-1} \sum_{h=1}^{n} (u_{h1}-\bar{u}_1)^2 = .01$ representa la varianza de la variable Y,

$$S_{22} = \frac{1}{n-1} \begin{pmatrix} \sum_{h=1}^{n} (u_{h2}-\bar{u}_2)^2 & \sum_{h=1}^{n} (u_{h_2}-\bar{u}_2)(u_{h3}-\bar{u}_3) \\ & \sum_{h=1}^{n} (u_{h3}-\bar{u}_3)^2 \end{pmatrix}$$

$$= \begin{pmatrix} 319065.8 & 67110.55 \\ & 87907.04 \end{pmatrix}$$ la matriz de varianzas y covarianzas de X y

$$S_{12} = \frac{1}{n-1}\left(\sum_{h=1}^{n} (u_{h1}-\bar{u}_1)(u_{h2}-\bar{u}_2) \quad \sum_{h=1}^{n} (u_{h1}-\bar{u}_1)(u_{h3}-\bar{u}_3)\right) = (3.24 \;\; 1.52)$$

la matriz de covarianza de las variables Y y X.

Sea $U(p\times 1) \sim N(\mu, A)$ y considérense las particiones inicialmente dadas.

Sean $f((y,x'))$ la densidad de U y $g(x')$ la densidad marginal de X. Entonces la densidad condicional de Y dado X=x está dada por

$$h(y|X=x) = \frac{f((y,x'))}{g(x')} \tag{2}$$

$$= \frac{\frac{1}{(2\pi)^{1/2p}} \frac{1}{|A|^{1/2}} \exp[-\frac{1}{2}(u-\mu)'A^{-1}(u-\mu)]}{\frac{1}{(2\pi)^{1/2p}} \frac{1}{|A_{22}|^{1/2}} \exp[-\frac{1}{2}(x-\mu^{(2)})' A_{22}^{-1}(x-\mu^{(2)})]}$$

donde $u = \begin{pmatrix} y \\ x \end{pmatrix}$

La expresión reducida de (2) se encuentra en libros de análisis multivariado. Debido a que son importantes las operaciones algebráicas para llegar a dicha expresión, a continuación se presenta el desarrollo paso a paso.

Primeramente, se escribe de manera adecuada la forma cuadrática del numerador. Para lograrlo, se busca una forma explícita para el inverso de A.

Sea $R = A^{-1} = \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix}$, multiplicando A por R e igualando a I(pxp), ya que $AA^{-1} = I$, se obtiene

$$R_{11} = (A_{11}-A_{12}\,A_{22}^{-1}\,A_{21})^{-1}$$
$$R_{22} = (A_{22}-A_{21}\,A_{11}^{-1}\,A_{12})^{-1}$$
$$R_{12} = -A_{11}^{-1}\,A_{12}(A_{22}-A_{21}\,A_{11}^{-1}\,A_{12})^{-1}$$
$$R_{21} = -A_{22}^{-1}\,A_{21}(A_{11}-A_{12}\,A_{22}^{-1}\,A_{21})^{-1}$$

por ser A^{-1} simétrica, $R_{12} = R_{21}$.

Sea $B = A_{12}\, A_{22}^{-1}$ y denótese $A_{11}-A_{12}\, A_{22}^{-1}\, A_{21} = A_{11.2}$ y $A_{22}-A_{21}\, A_{11}^{-1}\, A_{12} = A_{22.1}$, entonces

$$A^{-1} = \begin{pmatrix} A_{11.2}^{-1} & -A_{11.2}^{-1}B \\ -B'A_{11.2}^{-1} & A_{22.1}^{-1} \end{pmatrix}$$

Por otro lado, sea

$$M = \begin{pmatrix} A_{11.2}^{-1} & -A_{11.2}^{-1}B \\ -B'A_{11.2}^{-1} & A_{22}^{-1}+B'A_{11.2}^{-1}B \end{pmatrix}$$

como

$$\begin{pmatrix} A_{11.2} & 0 \\ 0 & A_{22} \end{pmatrix} = \begin{pmatrix} I & -B \\ 0 & I \end{pmatrix} \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} I & 0 \\ -B' & I \end{pmatrix} ,$$

de aquí se tiene

$$A = \begin{pmatrix} I & -B \\ 0 & I \end{pmatrix}^{-1} \begin{pmatrix} A_{11.2} & 0 \\ 0 & A_{22} \end{pmatrix} \begin{pmatrix} I & 0 \\ -B' & I \end{pmatrix}^{-1} \qquad \text{y además}$$

$$M = \begin{pmatrix} I & 0 \\ -B' & I \end{pmatrix} \begin{pmatrix} A_{11.2}^{-1} & 0 \\ 0 & A_{22}^{-1} \end{pmatrix} \begin{pmatrix} I & -B \\ 0 & I \end{pmatrix}$$

como $M = A^{-1}$ y $A_{22.1}^{-1} = A_{22}^{-1} + B'A_{11.2}^{-1}B$

ahora se escribe $(u-\mu)'\, A^{-1}(u-\mu)$ como

$$(u-\mu)' \quad A^{-1}(u-\mu) = \begin{bmatrix} y-\mu^{(1)} \\ x-\mu^{(2)} \end{bmatrix}' \begin{bmatrix} A_{11.2}^{-1} & -A_{11.2}^{-1}B \\ -B'A_{11.2}^{-1} & A_{22}^{-1}+B'A_{11.2}^{-1}B \end{bmatrix}$$

$$\begin{bmatrix} y-\mu^{(1)} \\ x-\mu^{(2)} \end{bmatrix} = [\,(y-\mu^{(1)})A_{11.2}^{-1} \; -(x-\mu^{(2)})B'A_{11.2}^{-1}$$

$$(x-\mu^{(2)})(A_{22}^{-1}+B'A_{11.2}^{-1}B)-(y-\mu^{(1)})A_{11.2}^{-1}B] \begin{bmatrix} y-\mu^{(1)} \\ x-\mu^{(2)} \end{bmatrix}$$

$$= (y-\mu^{(1)})' \; A_{11.2}^{-1}(y-\mu^{(1)})-(x-\mu^{(2)})B' \; A_{11.2}^{-1}(y-\mu^{(1)}) +$$

$$(x-\mu^{(2)})'(A_{22}^{-1} + B' \; A_{11.2}^{-1}B)(x-\mu^{(2)})-(y-\mu^{(1)})A_{11.2}^{-1}B(x-\mu^{(2)})$$

$$= [\,y-\mu^{(1)}-B(x-\mu^{(2)})]' \; A_{11.2}^{-1}[\,y-\mu^{(1)}-B(x-\mu^{(2)})]+(x-\mu^{(2)})' \; A_{22}^{-1}(x-\mu^{(2)})$$

Nótese que el primer término incluye el vector compuesto por el transpuesto de la diferencia de la variable y respecto a su media, menos una combinación lineal determinada por B de las diferencias de x menos su media.

Sustituyendo en la densidad (2) se obtiene

$h(y|X=x) =$

$$\frac{\dfrac{1}{|A|^{1/2}}\exp\left\{-\frac{1}{2}[y-\mu^{(1)}-B(x-\mu^{(2)})]' A_{11.2}^{-1}[y-\mu^{(1)}-B(x-\mu^{(2)})]-\frac{1}{2}(x-\mu^{(2)})'A_{22}^{-1}(x-\mu^{(2)})\right\}}{\dfrac{1}{|A_{22}|^{1/2}}\exp[-\frac{1}{2}(x-\mu^{(2)})' \; A_{22}^{-1} \; (x-\mu^{(2)})]}$$

$$= \frac{1}{|A_{11.2}|^{1/2}} \exp\left\{-\frac{1}{2}[y-\mu^{(1)}-B(x-\mu^{(2)})]'A_{11.2}^{-1}[y-\mu^{(1)}-B(x-\mu^{(2)})]\right\}$$

Puesto que ésta es una densidad normal, la variable $Y|X=x$ es normal con media

$$E(Y|X=x) = \mu^{(1)}+B(x-\mu^{(2)}) \qquad (3)$$

y matriz de covarianza

$$V(Y|X=x) = A_{11.2}$$

$E(Y|X=x)$ es lo que se conoce como la ecuación de regresión de Y en $X=x$, con $B=A_{12}\,A_{22}^{-1}$ la matriz de coeficientes de regresión.

Denotando a $E(Y|X=x)$ por h y tomando la traspuesta de la ecuación (3), se tiene la expresión más usual

$$h' = (x-\mu^{(2)})'\;B'+\mu^{(1)'} \qquad (4)$$

que desglosada en sus elementos se escribe

$$(h_1,\ h_2,\ \ldots,\ h_q) =$$

$$(\mu_1^{(1)}+\sum_{i=1}^{m}(x_i-\mu_i^{(2)})B_{i1},\mu_2^{(1)}+\sum_{i=1}^{m}(x_i-\mu_i^{(2)})B_{i2},\ldots,\mu_q^{(1)}+\sum_{i=1}^{m}(x_i-\mu_i^{(2)})B_{iq})$$

Mejía (1976) presenta algunas aplicaciones de esta ecuación.

A partir de (4), el modelo de regresión múltiple se obtiene cuando $q=1$. En este caso (4) es un número real con $(x-\mu^{(2)})'$ de dimensión $1xm$, $B'mx1$ y $\mu^{(1)'}$ es también un número real. Volviendo a la ecuación de regresión multivariada, nótese que x' es un vector de constantes conocidas de dimensión $1xm$, B' es una matriz de parámetros desconocidos mxq, $\mu^{(1)'}$ y $\mu^{(2)'}$

vectores de parámetros desconocidos 1xq y 1xm, respectivamente.

De la ecuación (4) se obtiene

$$h' = x'B' - \mu^{(2)'}B' + \mu^{(1)'}$$

Agrupando todos los parámetros desconocidos, agregando a la matriz B' un renglón (el primero) con los elementos de $\mu^{(1)'}-\mu^{(2)'}B'$, se tiene ahora $B^* = \begin{bmatrix} \mu^{(1)'}-\mu^{(2)'}B' \\ B' \end{bmatrix}$ de dimensión (m+1)xq.

Agregando a x' una columna de unos (la primera), se tiene $v=(1x')$ vector de dimensión 1x(m+1) y (4) se puede expresar como

$$h' = vB^*$$

denotando a la variable aleatoria $Y|X=x$ como w se tiene

$$w = vB^* + \varepsilon \qquad (5)$$

donde ε es un vector aleatorio 1xq con $E(\varepsilon)=0$ y $V(\varepsilon)=A_{11.2}$

La ecuación (5) desglosada se escribe

$$(w_1, w_2,\ldots,w_q) =$$
$$\left(\sum_{j=1}^{m+1} v_j B^*_{j1}, \sum_{j=1}^{m+1} v_j B^*_{j2},\ldots, \sum_{j=1}^{m+1} V_j B^*_{jq}\right)+(\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_q)$$

Nótese que $\sum_{j=1}^{m+1} V_j B^*_{jK} = \mu^{(1)}_K + \sum_{i=1}^{m}(x_i-\mu^{(2)}_i)B_{iK}$ con $K=1,2,\ldots,q$

Ahora bien si se toman n observaciones de w, i.e., una observación del vector Y para cada uno de los valores $x_{(1)},\ldots,x_{(n)}$ - del vector X, se obtiene

$$W = VB^* + e$$

donde $W = \begin{pmatrix} y'_{(1)} \\ \vdots \\ y'_{(n)} \end{pmatrix}$ es una matriz nxq, $V = \begin{pmatrix} 1 & x'_{(1)} \\ 1 & x'_{(2)} \\ \vdots & \vdots \\ 1 & x'_{(n)} \end{pmatrix}$ es una matriz nx(m+1) y e es la matriz de residuales nxq.

Hemos llegado entonces al modelo lineal multivariado (5) a partir de la ecuación de regresión (4).

Ejemplo.- Siguiendo con el ejemplo anterior, sea $U=(u_1,u_2,u_3)' \sim N(\mu,A)$ y considérense las particiones de U, $\hat{\mu}$ y S como anteriormente se definieron.

Considérense 22 observaciones del vector X (rendimiento máximo en Kg. por hectárea de maíz mejorado en tierra de riego y de - temporal) y sean $y_{(1)}, y_{(2)}, \ldots, y_{(22)}$ las 22 muestras correspondientes del vector Y (precio rural máximo por Kg. de maíz mejorado) dados aquellos 22 valores del vector X, $x_{(1)},x_{(2)},\ldots,x_{(22)}$. Entonces

$$W = \begin{pmatrix} .95 \\ 1.09 \\ 1.08 \\ 1.09 \\ 1.22 \\ 1.10 \\ 1.05 \\ 1.22 \\ 1.04 \\ 1.20 \\ .98 \\ .96 \\ 1.16 \\ 1.15 \\ 1.16 \\ 1.08 \\ 1.12 \\ 1.02 \\ .95 \\ 1.10 \\ 1.09 \\ 1.00 \end{pmatrix}$$ es de dimensión 22x1 ,

$$V = \begin{pmatrix} 1 & 4000 & 2555 \\ 1 & 3140 & 1835 \\ 1 & 3423 & 1700 \\ 1 & 4085 & 1755 \\ 1 & 3175 & 2423 \\ 1 & 3826 & 1837 \\ 1 & 3650 & 2418 \\ 1 & 4927 & 2683 \\ 1 & 3916 & 2720 \\ 1 & 3639 & 2232 \\ 1 & 3687 & 2015 \\ 1 & 2480 & 1990 \\ 1 & 2934 & 1708 \\ 1 & 2936 & 1964 \\ 1 & 3753 & 2000 \\ 1 & 3750 & 2195 \\ 1 & 4533 & 2214 \\ 1 & 4476 & 2000 \\ 1 & 3714 & 2110 \\ 1 & 3167 & 2210 \\ 1 & 3720 & 1993 \\ 1 & 3591 & 1956 \end{pmatrix}$$, de dimensión 22x3

Sean $\hat{\mu} = \left(\begin{array}{c} 1.08 \\ \hline 3660.09 \\ 2114.23 \end{array}\right) = \left(\begin{array}{c} \bar{u}_1 \\ \hline \bar{u}_2 \\ \bar{u}_3 \end{array}\right) = \begin{pmatrix} u^{(1)} \\ u^{(2)} \end{pmatrix}$,

$$S = \left(\begin{array}{c|cc} .007 & 3.24 & 1.52 \\ \hline & 319065.8 & 67110.55 \\ & & 87907.04 \end{array}\right) = \left(\begin{array}{c|cc} s_{11} & s_{12} & s_{13} \\ \hline & s_{22} & s_{23} \\ & & s_{33} \end{array}\right) = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} \quad y$$

$$S_{22}^{-1} = \begin{pmatrix} .000003734 & -.000002850 \\ & .000013552 \end{pmatrix} = \begin{pmatrix} s^{22} & s^{23} \\ s^{32} & s^{33} \end{pmatrix} \text{, entonces}$$

$$\hat{B} = S_{12}\, S_{22}^{-1} = (3.24 \quad 1.52) \begin{pmatrix} .000003734 & -.000002850 \\ & .000013552 \end{pmatrix}$$

$$= (.00000776 \qquad .0000113)$$

$$\hat{B}^* = \begin{pmatrix} u^{(1)'} - \hat{B}' u^{(2)'} & \hat{B}' \end{pmatrix}$$

$$= \begin{pmatrix} \bar{u}_1 - \bar{u}_2(s_{12}s^{22}+s_{13}s^{32}) + \bar{u}_3(s_{12}s^{23}+s_{13}s^{33}) \\ s_{12}s^{22} + s_{13}s^{32} \\ s_{12}s^{23} + s_{13}s^{33} \end{pmatrix}$$

$$= \begin{pmatrix} 1.029 \\ .00000776 \\ .0000113 \end{pmatrix} \qquad y$$

$$S_{11.2} = S_{11} - S_{12}\, S_{22}^{-1}\, S_{21} = \hat{A}_{11.2}$$

$$= .007-(3.24 \quad 1.52) \begin{pmatrix} .000003734 & -.000002850 \\ & .000013552 \end{pmatrix} \begin{pmatrix} 3.24 \\ 1.52 \end{pmatrix}$$

$$= .0069$$

Nótese que el primer elemento de $\hat{B}^*$ se obtiene de sus otros dos elementos, y a su vez éstos son combinaciones lineales del primer renglón de la matriz S_{12} con la matriz S_{22}^{-1}.

Con los datos anteriores se tiene el modelo lineal

$$W = V\hat{B}^* + e \tag{6}$$

donde e es el vector de residuales $e = W - V\hat{B}^*$

En este caso $\hat{A}_{11.2}$ es un escalar, coincidiendo con el modelo de regresión usual. Sustituyendo los valores de W, V y $\hat{B}^*$ en (6) se obtiene

$$\begin{pmatrix} .95 \\ 1.09 \\ 1.08 \\ 1.09 \\ 1.22 \\ 1.10 \\ 1.05 \\ 1.22 \\ 1.04 \\ 1.20 \\ .98 \\ .96 \\ 1.16 \\ 1.15 \\ 1.16 \\ 1.08 \\ 1.12 \\ 1.02 \\ .95 \\ 1.10 \\ 1.09 \\ 1.00 \end{pmatrix} = \begin{pmatrix} 1.08 \\ 1.07 \\ 1.07 \\ 1.08 \\ 1.08 \\ 1.07 \\ 1.08 \\ 1.09 \\ 1.09 \\ 1.08 \\ 1.08 \\ 1.07 \\ 1.07 \\ 1.07 \\ 1.08 \\ 1.08 \\ 1.08 \\ 1.08 \\ 1.08 \\ 1.07 \\ 1.08 \\ 1.07 \end{pmatrix} + \begin{pmatrix} e_{11} \\ e_{21} \\ e_{31} \\ e_{41} \\ e_{51} \\ e_{61} \\ e_{71} \\ e_{81} \\ e_{91} \\ e_{101} \\ e_{111} \\ e_{121} \\ e_{131} \\ e_{141} \\ e_{151} \\ e_{161} \\ e_{171} \\ e_{181} \\ e_{191} \\ e_{201} \\ e_{211} \\ e_{221} \end{pmatrix} \tag{7}$$

mientras que la ecuación de regresión es

$$\hat{h}' = 1.08 + (x_1 - 3660.09\ ,\ x_2 - 2114.23) \begin{pmatrix} .00000776 \\ .00001133 \end{pmatrix}$$

$$= 1.08 + .00000776x_1 - .028 + .00001133x_2 - .023$$

$$= 1.029 + .00000776x_1 + .00001133x_2 \tag{8}$$

Nótese que al sustituir valores de $x'=(x_1,x_2)$ en (8), se obtiene el renglón correspondiente de la matriz $V\hat{B}*$ en (7). Por ejemplo, tómese el séptimo renglón de la matriz V sin considerar la primera columna, i.e. $x'_{(7)} = (3650 \quad 2418)$, al sustituir en (8) se obtiene

$$\hat{h}' = 1.029+.0283+.0274 = 1.085$$

que corresponde al séptimo renglón de la matriz $V\hat{B}*$.

2. Correlación Canónica.

Supóngase que se quiere conocer la correlación entre dos variables vectoriales, por ejemplo, el consumo de, leche y huevo con relación al precio de los mismos. Para resolver esto, se propone encontrar una suma ponderada de cada variable en forma tal que la correlación entre ambas sumas sea máxima. Al método que determina estas combinaciones se le llama correlación canónica.

Dado $U=\begin{bmatrix} Y \\ X \end{bmatrix} \begin{matrix} (qx1) \\ (mx1) \end{matrix}$, el problema de correlación canónica consiste

en encontrar una combinación lienal de Y, $a'Y=\Upsilon$ $(a'(1xq))$, y otra de X, $b'X=W$ en forma tal que la correlación entre ellas sea máxima. El coeficiente de correlación canónica ρ_c entre las dos variables resultantes se define entonces como

$\rho_c = \frac{cov(\Upsilon,w)}{\sqrt{var(\Upsilon)var(w)}}$. Al máximo coeficiente de correlación canónica se le denota como $\rho_c^{(1)}$ y se le llama primer coeficiente de correlación canónica entre Y y X. A las combinaciones lineales Υ_1 y w_1 asociadas a $\rho_c^{(1)}$ se les denomina primer par de variables canónicas. La segunda correlación canónica se determina por medio de combinaciones lineales Υ_2 y w_2 tales que de todas las combinaciones lineales no correlacionadas con Υ_1 y w_1, estas segundas variables canónicas proporcionen la segunda máxima correlación canónica $\rho_c^{(2)}$. La tercera correlación canónica se determina por medio de combinaciones lineales Υ_3 y w_3 tales que entre todas las combinaciones lineales no correlacionadas con

γ_1 y w_1, ni γ_2 y w_2, estas terceras variables canónicas proporcionan la tercera máxima correlación canónica. Cada pareja sucesiva de variables canónicas se determina en forma análoga, y si $q \leq m$, habrá q correlaciones canónicas y q parejas de variables canónicas. En este capítulo se formaliza la forma de obtener todas las posibles correlaciones y parejas de variables canónicas.

Sea $U=\begin{pmatrix} Y \\ X \end{pmatrix} \begin{matrix} (q\times 1) \\ (m\times 1) \end{matrix}$ considérense las particiones

$\mu=\begin{pmatrix} \mu^{(1)} \\ \mu^{(2)} \end{pmatrix} \begin{matrix} (q\times 1) \\ (m\times 1) \end{matrix}$ y $A=\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$ de la media y la matriz de covarianza de U.

Sea $p=q+m$, para obtener la primera correlación canónica entre Y y X se plantea el problema

$$\max_{a,b} \ \operatorname{corr}(a'Y,\ b'X)= \max_{a,b} \frac{\operatorname{cov}(a'Y,\ b'X)}{[\operatorname{var}(a'Y)\operatorname{var}(b'X)]^{1/2}}$$

Sin pérdida de generalidad, ya que la correlación es invariante a la multiplicación por constantes, se toman las condiciones

$$\operatorname{var}(a'Y)=1 \quad y \quad \operatorname{var}(b'X)=1$$

nótese que $\operatorname{var}(a'Y)=a'A_{11}a$, $\operatorname{var}(b'X)=b'A_{22}b$ y

$\operatorname{cov}(a'Y,\ b'X)=a'A_{12}b$, entonces se busca

$$\max_{a,b} \frac{a'A_{12}b}{[(a'A_{11}a)(b'A_{22}b)]^{1/2}}$$

sujeto a $a'A_{11}a=1$ y $b'A_{22}b=1$

Introduciendo multiplicadores de Lagrange, esto es lo mismo que maximizar

$$L=a'A_{12}b+\lambda_1(1-b'A_{22}b)+\lambda_2(1-a'A_{11}a) \ ,$$

Las primeras derivadas parciales igualadas a cero son

$$\frac{\partial L}{\partial a}=A_{12}b-\lambda_2A_{11}a=0 \qquad (1)$$

$$\frac{\partial L}{\partial b}=A_{21}a-\lambda_1A_{22}b=0 \qquad (2)$$

$$\frac{\partial L}{\partial \lambda_1}=1-b'A_{22}b=0 \qquad (3)$$

$$\frac{\partial L}{\partial \lambda_2}=1-a'A_{11}a=0 \qquad (4)$$

Premultiplicando la primera ecuación por a' y la segunda por b' se tiene $b'A_{12}a=a'A_{12}b=\lambda_1=\lambda_2=\lambda$. De aquí se ve que si A_{12} tiene elementos distintos de cero, $\lambda\neq0$, y que la primera correlación canónica coincide con λ. De esta manera se construye con las ecuaciones (1) y (2) el sistema

$$\begin{pmatrix} -\lambda A_{11} & A_{12} \\ A_{21} & -\lambda A_{22} \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = 0 \tag{5}$$

de aquí se observa que la primera correlación canónica es la máxima solución λ_1 de

$$\begin{vmatrix} -\lambda A_{11} & A_{12} \\ A_{21} & -\lambda\ A_{22} \end{vmatrix} = 0 \tag{6}$$

Como A_{22} es no singular, se tiene

$$\begin{aligned}
\begin{vmatrix} -\lambda A_{11} & A_{12} \\ A_{21} & -\lambda A_{22} \end{vmatrix} &= |-\lambda A_{22}|\,|-\lambda A_{11}-A_{12}(-\lambda)^{-1}A_{22}^{-1}A_{21}| \\
&= (-1)^m\lambda^m|\ A_{22}|\,|-\lambda A_{11}-A_{12}(-\lambda)^{-1}A_{22}^{-1}A_{21}| \\
&= (-1)^m\lambda^m(-1)^q\lambda^{-q}|A_{22}|\,|(-\lambda A_{11}-A_{12}(-\lambda)^{-1}A_{22}^{-1}A_{21})(-\lambda)| \\
&= (-1)^{m+q}\lambda^{m-q}\,|A_{22}|\,|\lambda^2 A_{11}-A_{12}A_{22}^{-1}A_{21}|
\end{aligned}$$

Por lo tanto, el determinante en (6) será cero si y solo si

$$|\lambda^2\ A_{11}-A_{12}A_{22}^{-1}A_{21}| = 0 \tag{7}$$

Nótese que (7) representa una ecuación polinomial de grado q y por lo tanto tiene q raíces. Sean estas raíces

$\lambda_1{}^2 \geqslant \lambda_2{}^2 \geqslant \ldots \geqslant \lambda^2 q$.

Ya se viō que $b'A_{12}a=\lambda$, por lo tanto maximizar $b'A_{12}a$ significa tomar $\lambda=\lambda_1$. Sean a_1 y b_1 las soluciones de (5) para $\lambda=\lambda_1$, entonces $\gamma_1=a_1{}'Y$ y $w_1=b_1{}'X$ tienen māxima correlaciōn y cumplen con var $(\gamma_1)=$ var $(w_1)=1$.

Supongamos que ahora se desea obtener una segunda combinaciōn lineal de Y, $\gamma=a'Y$ y una segunda comibnaciōn de X, $w=b'X$, con varianza 1, tal que de todas las combinaciones lineales no correlacionadas con γ_1 y w_1 tenga māxima correlaciōn i.e. se quiere maximizar

$$\frac{a'A_{12}b}{[(a'A_{11}a)(b'A_{22}b)]^{1/2}} \text{sujeto a} \quad a'A_{11}a=1 \quad , \; b'A_{22}b=1 \; ,$$

$a'A_{11}a_1=0$ y $b'A_{22}b_1=0$. Introduciendo nuevamente multiplicadores de Lagrange se tiene ahora

$$L=a'A_{12}b+\lambda_1 \; (1-b'A_{22}b)+\lambda_2(1-a'A_{11}a)+2\delta_1 a'A_{11}a_1+2\Gamma_1 b'A_{22}b_1$$

Los derivadas parciales igualadas a cero son

$$\frac{\partial L}{\partial a}= A_{12}b-\lambda_2 A_{11}a+\delta_1 A_{11}a_1 =0 \qquad (8)$$

$$\frac{\partial L}{\partial b}= A_{21}a-\lambda_1 A_{22}b+ \Gamma_1 A_{22}b_1= 0 \qquad (9)$$

Premultiplicando la primera ecuaciōn por a'_1 y la segunda por b'_1

se obtiene $\delta_1 a_1' A_{11} a_1 = \delta_1 = 0$ y $\Gamma_1 b_1' A_{22} b_1 = \Gamma_1 = 0$, ya que $a_1' A_{12} b = b_1' A_{21} a = 0$. Por lo tanto (8) y (9) es lo mismo que (1) y (2). La solución al problema es entonces $\lambda = \lambda_2$, sean a_2 y b_2 las soluciones de (5) para estevalor de λ, entonces $\gamma_2 = a_2' Y$ y $w_2 = b_2' X$ tienen máxima correlación y cumplen con las condiciones antes mencionadas.

Este procedimiento se continúa en forma tal que es posible descomponer Y y X en combinaciones lineales de sus elementos, mutuamente independientes y con correlación máxima entre parejas de combinaciones. Dicho de otra manera, se puede descomponer Y en $\gamma_1, \gamma_2, \ldots, \gamma_q$ y X en $w_1, w_2, \ldots w_q$ tales que $\mathrm{cov}(\gamma_i, \gamma_j) = \mathrm{cov}(w_i, w_j) = \mathrm{cov}(\gamma_i, w_j) = 0$ $i \neq j$ y las cantidades

$$\lambda_i = \frac{\mathrm{cov}(\gamma_i, w_i)}{[\mathrm{var}(\gamma_i)\mathrm{var}(w_i)]^{1/2}} \quad i = 1, 2, \ldots, q$$ se maximizan sujetas a la

restricción $\mathrm{var}(\gamma_i) = \mathrm{var}(w_i) = 1$.

Como los coeficientes de correlación canónica son invariantes bajo cambios en las unidades de medición, éstos pueden ser estimados indistintamente a partir de S, la matriz de covarianza muestral, o R, la matriz de correlación muestral. En cuanto a los coeficientes $\hat{a}_i$ y $\hat{b}_i$ no se puede decir lo mismo. Si $\hat{a}_i$ y $\hat{b}_i$ se calculan a partir de S, éstas se aplican a las variables originales; si se calculan a partir de R, se aplican a las variables ~~originales~~ estandarizadas.

La relación que existe entre unos coeficientes y otros es la siguiente

Sea

$$S=\begin{pmatrix} s_{11} & s_{12}\cdots s_{1p} \\ s_{21} & s_{22}\cdots s_{2p} \\ \vdots & \vdots \quad\quad \vdots \\ s_{p1} & s_{p2}\cdots s_{pp} \end{pmatrix} \quad y \quad R=\begin{pmatrix} 1 & r_{12}\cdots r_{1p} \\ r_{21} & 1 \quad \cdots r_{2p} \\ \vdots & \vdots \quad\quad \vdots \\ r_{p1} & r_{p2} \quad\quad 1 \end{pmatrix}$$

entonces

$$\mathrm{diag}\left(\frac{1}{\sqrt{s_{11}}},\frac{1}{\sqrt{s_{22}}},\dots,\frac{1}{\sqrt{s_{pp}}}\right) S \ \mathrm{diag}\left(\frac{1}{\sqrt{s_{11}}},\frac{1}{\sqrt{s_{22}}},\dots,\frac{1}{\sqrt{s_{pp}}}\right) = R$$

Por lo tanto, si $\hat{a}_i$, $\hat{b}_i$ $i=1,2,\dots q$ se calcularon a partir de S, y $\hat{a}_i^*$, $\hat{b}_i^*$ se calcularon a partir de R, se tiene que

$$\mathrm{diag}\left(\frac{1}{\sqrt{s_{11}}},\frac{1}{\sqrt{s_{22}}},\dots \frac{1}{\sqrt{s_{pp}}}\right)\begin{pmatrix}\hat{a}_i\\ \hat{b}_i\end{pmatrix} = \begin{pmatrix}\hat{a}_i^*\\ \hat{b}_i^*\end{pmatrix}$$

Para ilustrar considérese el siguiente ejemplo. En base a dos mediciones dadas por Rao, sobre el primero y segundo hijos adultos en una muestra de 25 familias, Anderson investiga las relaciones entre las mediciones del primero y segundo hijos como sigue. Sea $X'_\alpha=(x_{1\alpha},x_{2\alpha},x_{3\alpha},x_{4\alpha})$ el vector formado por las mediciones $x_{1\alpha}$ longitud de la cabeza del primer hijo en la familia α, $x_{2\alpha}$ el ancho de la cabeza del primer hijo, $x_{3\alpha}$ longitud de la cabeza del segundo hijo y $x_{4\alpha}$ el ancho de la cabeza del segundo hijo.

Se calculan estimadores $\hat{\mu}=\frac{1}{n}\sum_{h=1}^{n} X_h$ y $S=\frac{1}{n-1}\sum_{h=1}^{n}(X_h-\hat{\mu})(X_h-\hat{\mu})'$ de la media y la matriz de covarianza, donde n=25. De aquí se obtiene $\hat{\mu}'=(185.72,\ 151.12,\ 183.84,\ 149.24)$ y

$$S = \left(\begin{array}{cc:cc} 95.2933 & 52.8683 & 69.6617 & 46.1117 \\ & 54.3600 & 51.3117 & 35.0533 \\ \hdashline & & 100.8067 & 56.5400 \\ & & & 45.0233 \end{array}\right)$$

$$= \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}$$

Sea $R = \begin{pmatrix} 1.0000 & 0.7346 & 0.7108 & 0.7040 \\ & 1.0000 & 0.6932 & 0.7086 \\ & & 1.0000 & 0.8392 \\ & & & 1.0000 \end{pmatrix}$

$$= \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix}$$

la matriz de correlaciones. Dado que las soluciones de $|\lambda^2 S_{11} - S_{12} S_{22}^{-1} S_{21}| = 0$ son las mismas que las de $|\lambda^2 R_{11} - R_{12} R_{22}^{-1} R_{21}| = 0$, y como resulta más sencillo trabajar con esta última se tiene

$$R_{22}^{-1} R_{21} = \begin{pmatrix} 0.405769 & 0.333205 \\ 0.363480 & 0.428976 \end{pmatrix} ,$$

$$R_{12} R_{22}^{-1} R_{21} = \begin{pmatrix} 0.544311 & 0.538841 \\ 0.538841 & 0.534950 \end{pmatrix} , \text{ sea } \nu = \lambda^2, \text{ entonces}$$

$$0 = \begin{vmatrix} 0.544311 - \nu & 0.538841 - 0.7346\nu \\ 0.538841 - 0.7346\nu & 0.534950 - \nu \end{vmatrix}$$

$$= 0.460363\nu^2 - 0.287596\nu + 0.000830$$

Las raíces son 0.621816 y 0.002900, por lo tanto λ_1=0.788553 y λ_2= 0.053852

La solución de $\begin{pmatrix} -\lambda R_{11} & R_{12} \\ R_{21} & -\lambda\ R_{22} \end{pmatrix} \begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} = 0$ correspondiente a

λ_1= 0.788553 es $\hat{a}_1 = \begin{pmatrix} 0.0566 \\ 0.0707 \end{pmatrix}$ y $\hat{b}_1 = \begin{pmatrix} 0.0502 \\ 0.0802 \end{pmatrix}$

y la correspondiente a λ_2= 0.053852 es $\hat{a}_2 = \begin{pmatrix} 0.1400 \\ -0.1870 \end{pmatrix}$ y

$\hat{b}_2 = \begin{pmatrix} 0.1760 \\ -0.2619 \end{pmatrix}$.

3. Análisis Discriminante.

Se tienen definidos dos grupos de lectores, dependiendo del períodico al que están suscritos, y se conocen vectores muestra de características socio-económicas de cada uno de ellos. Supóngase que se tiene un vector observado de características socio-económicas de un individuo que se sabe lee uno de los dos periódicos, pero no exactamente, cuál de ellos. Se plantea entonces el siguiente problema. Dadas las muestras de los grupos, se desea construir una regla que permita decidir a cuál de los grupos pertenece el individuo de quien se conoce un vector de observaciones. Por lo general, esto se resolverá a través de funciones discriminantes, y es el caso que se tratará aquí.

En general, si se tienen definidas R poblaciones, y se conocen muestras de cada una de ellas, el análisis discriminante proporciona reglas basadas en dichas muestras que asignan una nueva observación dada, a una de estas poblaciones.

En este capítulo se obtiene la función discriminante entre dos poblaciones y en el capítulo que se relaciona el análisis discriminante con correlación canónica, se generaliza la obtención para R poblaciones. Para obtener la función discriminante entre dos poblaciones y el criterio para clasificar una observación determinada en una de estas poblaciones, se seguirán dos métodos distintos y se mostrará que se llega al mismo resultado.

El primer método consiste en definir dos regiones, bajo ciertas restricciones, en forma tal de que si una observación está en la región uno, se dice que tal observación pertenece a la población uno, si esto no sucede, entonces la observación pertenece a la población dos.

A continuación se desarrolla este método.

Supóngase que se tienen dos poblaciones π_1 y π_2 las cuales - están totalmente caracterizadas por sus densidades $f_1(x)$ y $f_2(x)$, respectivamente. Esto es, si un individuo al que se le asocia un vector de mediciones $x=(x_1,x_2,...x_m)'$, pertenece a π_i, entonces $x \sim f_i(x)$, $i=1,2$.

Supóngase que se conocen las probabilidades de pertenencia a las poblaciones. Sean p_1 la probabilidad de que un individuo x pertenezca a π_1 y p_2 la probabilidad de que x pertenezca a π_2.

Dado un individuo $x \in R^m$, al ser asignado a una de las dos poblaciones π_1 o π_2, puede cometerse uno de los dos siguientes errores. O bien x se asigna a π_2 y pertenecía a π_1, o x se asigna a π_1 y pertenecía a π_2. Sean $c(2|1)$ y $c(1|2)$, respectivamente, los costos imputados a estos errores.

La pérdida esperada de la clasificación errónea es entonces,

$$c(2|1)\ p_1 \int_{R_2} f_1(x)dx + c(1|2)p_2 \int_{R_1} f_2(x)dx \qquad (1)$$

donde R_1 es una región de $\mathbb{R}^m$ definida de tal forma que si $x \in R_1$, entonces x se asigna en π_1 y análogamente para R_2. - Además $R_1 \cup R_2 = \mathbb{R}^m$ y $R_1 \cap R_2 = \phi$

Dado un individuo x, el problema de asignarlo a π_1 o π_2 se resolvería si R_1 estuviese definida.

Una forma de definirla sería minimizando la pérdida esperada (1). Como $\int_{R_1} f_1(x)dx + \int_{R_2} f_1(x)dx = 1$, (1) se puede escribir de la siguiente forma

$$\int_{R_1} c(1|2)p_2 f_2(x)dx + c(2|1)p_1[1 - \int_{R_1} f_1(x)dx]$$

$$= \int_{R_1} [c(1|2)p_2 f_2(x) - c(2|1)p_1 f_1(x)]dx + c(2|1)p_1$$

De aquí se ve que este valor será mínimo cuando

$$\int_{R_1} [c(1|2)p_2 f_2(x) - c(2|1)p_1 f_1(x)]dx \qquad (2)$$

sea mínimo.

Sea $R_1^c = \{x \in \mathbb{R}^m | f_2(x) \geq \frac{c(2|1)}{c(1|2)} \frac{p_1}{p_2} f_1(x)\}$, se va a demostrar que $\int_{R_1^c} [c(1|2)p_2 f_2(x) - c(2|1)p_1 f_1(x)]dx$ es máxima, esto es, que

$$\int_{R_1^c} [c(1|2)p_2f_2(x)-c(2|1)p_1f_1(x)]\,dx$$

$$-\int_R [c(1|2)p_2f_2(x)-c(2|1)p_1f_1(x)]\,dx \geq 0 \quad \forall\ R\subset\mathbb{R}^m \qquad (3)$$

Dado que R_1^c se puede expresar como $(R_1^c\cap R)\cup(R_1^c\cap R^c)$ y R se puede expresar como $(R\cap R_1^c)\cup(R\cap R_1)$, el lado izquierdo de la desigualdad en (3) es mayor o igual a

$$\int_{R_1^c\cap R} [c(1|2)p_2f_2(x)-c(2|1)p_1f_1(x)]\,dx$$
$$-\int_{R\cap R_1} [c(1|2)p_2f_2(x)-c(2|1)p_1f_1(x)]\,dx \qquad (4)$$

Por la definición de R_1^c, la primera integral es ≥ 0 y la segunda ≤ 0, por lo tanto, la expresión en (4) es ≥ 0 $\forall$ R y la expresión en (2) es mínima.

Nótese que la definición de R_1^c coincide con la de la mejor región crítica dada en el lema de Neyman-Pearson para realizar la prueba $\quad H_1 : f(x) = f_1(x)$

$\quad H_2 : f(x) = f_2(x)$

Suponiendo que $f_1(x)$ y $f_2(x)$ son densidades normales con medias μ_1 y μ_2, respectivamente, matriz de covarianza común A y $c(1|2)= c(2|1)$, R_1 se puede expresar como sigue

$$R_1 = \{x\in\mathbb{R}^m \mid p_1 e^{-\frac{1}{2}(x-\mu_1)'A^{-1}(x-\mu_1)} > p_2 e^{-\frac{1}{2}(x-\mu_2)'A^{-1}(x-\mu_2)}\}$$

$$= \{x \in R^m | -\tfrac{1}{2}(x-\mu_1)'A^{-1}(x-\mu_1)+\ln p_1 > -\tfrac{1}{2}(x-\mu_2)'A^{-1}(x-\mu_2)+\ln p_2\}$$

$$= \{x \in R^m | (x-\mu_1)'A^{-1}(x-\mu_1)-2\ln p_1 < (x-\mu_2)'A^{-1}(x-\mu_2)-2\ln p_2\}$$

$$= \{x \in R^m | -2x'A^{-1}\mu_1+\mu_1'A^{-1}\mu_1-2\ln p_1 < -2x'A^{-1}\mu_2+\mu_2'A^{-1}\mu_2-2\ln p_2\}$$

$$= \{x \in R^m | x'A^{-1}\mu_1-x'A^{-1}\mu_2-\tfrac{1}{2}\mu_1'A^{-1}\mu_1+\tfrac{1}{2}\mu_2'A^{-1}\mu_2>\ln(p_2/p_1)\}$$

$$= \{x \in R^m | x'A^{-1}(\mu_1-\mu_2)-\tfrac{1}{2}(\mu_1+\mu_2)'A^{-1}(\mu_1-\mu_2) > \ln(p_2/p_1)\}$$

$$= \{x \in R^m | (\mu_1-\mu_2)'A^{-1}x > \ln(p_2/p_1)+\tfrac{1}{2}(\mu_1+\mu_2)'A^{-1}(\mu_1-\mu_2)\}$$

Por tanto, la regla de decisión será, asígnese x a la población π_1 si

$$(\mu_1-\mu_2)'A^{-1}x > \ln(p_2/p_1)+\tfrac{1}{2}(\mu_1+\mu_2)'A^{-1}(\mu_1-\mu_2) \qquad (5)$$

y asígnese x a π_2 si esta desigualdad no se cumple.

El segundo método consiste en maximizar la variación entre las poblaciones y al mismo tiempo minimizar la variación dentro de cada una de ellas.

Supóngase nuevamente, que las poblaciones π_1 y π_2 se distribuyen normalmente con medias μ_1 y μ_2, respectivamente, y matriz de covarianza común A. Se quiere determinar una combinación lineal de $x \in R^m$ de forma tal que las transformaciones así obtenidas tengan distribuciones con traslape mínimo, i.e. que las dos poblaciones estén lo más diferenciadas que sea posible.

Dicho de otra manera, supóngase que la función discriminante tiene la forma

$$y_{it} = a_1x_{i1t}+a_2x_{i2t}+\ldots+a_mx_{imt} \qquad (6)$$

donde i=1,2 representa la población, $t=1,2,\ldots,n_i$ es el número de observaciones de cada población y a_j son coeficientes desconocidos $\forall j$ con $j=1,2,\ldots m$.

Sean $a=(a_1,a_2,\ldots,a_m)'$ y $x_{it}=(x_{i1t},x_{i2t},\ldots x_{imt})'$, la ecuación (6) se expresa entonces en forma matricial como

$$y_{it} = a'x_{it} \qquad (7)$$

Defínase $\bar{y}_i = a_1\bar{x}_{i1}+a_2\bar{x}_{i2}+\ldots a_m\bar{x}_{im}$, i=1,2 donde $\bar{x}_{ij}$ es la media de la variable j-ésima en el i-ésimo grupo. Para encontrar la combinación lineal de x antes mencionada, lo que se quiere por un lado, es que el cuadrado de la distancia $(\bar{y}_1-\bar{y}_2)^2$, la cual representa la separación de los dos grupos, sea lo más grande posible. Por otro lado, se quiere que la variación dentro de cada grupo que está dada por $\sum_{t=1}^{n_1}(y_{1t}-\bar{y}_1)^2$ y $\sum_{t=1}^{n_2}(y_{2t}-\bar{y}_2)^2$ para los grupos 1 y 2, respectivamente, sea lo más pequeña posible. Esto es, se quiere que los elementos dentro de un mismo grupo sean lo más homogéneos posibles. La forma de determinar los coeficientes de la función discriminante (6), y de acuerdo a la explicación anterior, es maximizando la razón

$$R = \frac{(\bar{y}_1-\bar{y}_2)^2}{\sum_{i=1}^{2}\sum_{t=1}^{n_i}(y_{it}-\bar{y}_i)^2} \quad , \qquad (8)$$

ya que ésta representa la distancia estandarizada entre los dos grupos.

Nótese que R es una función monótona de F en un análisis de varianza para probar la diferencia entre grupos con la variable y. Esto se ve porque $(\bar{y}_1-\bar{y}_2)^2 = \frac{n_1+n_2}{n_1 n_2}[n_1(\bar{y}_1-\bar{y})^2+n_2(\bar{y}_2-\bar{y})^2]$ es la suma de cuadrados entre grupos y el denominador de (8) es la suma de cuadrados dentro de grupos.

Como $y_{it}-\bar{y}_i = a_1(x_{i1t}-\bar{x}_{i1})+a_2(x_{i2t}-\bar{x}_{i2})+\ldots+a_m(x_{imt}-\bar{x}_{im})$ y $\bar{x}_i = (\bar{x}_{i1},\bar{x}_{i2},\ldots,\bar{x}_{im})'$ con $i=1,2$

R se puede expresar como

$$R = \frac{a'(\bar{x}_1-\bar{x}_2)(\bar{x}_1-\bar{x}_2)'a}{\sum_{i=1}^{2}\sum_{t=1}^{n_i} a'(x_{it}-\bar{x}_i)(x_{it}-\bar{x}_i)'a} \qquad (9)$$

Con el objeto de tener una notación que se usará más adelante, sean

$$x_1 = [x_{11}-\bar{x}_1, x_{12}-\bar{x}_1, \ldots, x_{1n_1}-\bar{x}_1]$$

y

$$x_2 = [x_{21}-\bar{x}_2, x_{22}-\bar{x}_2, \ldots, x_{2n_2}-\bar{x}_2]$$

las matrices de desviaciones de dimensiones mxn_1 y mxn_2, respectivamente. De acuerdo a esta relación se tiene

$$\begin{aligned}\sum_{i=1}^{2}\sum_{t=1}^{n_i}(y_{it}-\bar{y}_i)^2 &= a'x_1x_1'a+a'x_2x_2'a\\ &= a'(x_1x_1'+x_2x_2')a\\ &= a'D_d a\end{aligned}$$

donde $D_d = x_1x_1' + x_2x_2'$ es la matriz de sumas de cuadrados y productos dentro de grupos. Por lo tanto la ecuación (9) queda

$$R = \frac{a'(\bar{x}_1-\bar{x}_2)(\bar{x}_1-\bar{x}_2)'a}{a'D_d a} \tag{10}$$

Si a_o maximiza a R, entonces ca_o también maximiza a R para cualquier c. Para eliminar esta indeterminación, se toma la restricción $a'D_d a = 1$.

El problema es entonces

$$\max_{a'D_d a=1} \frac{a'(\bar{x}_1-\bar{x}_2)(\bar{x}_1-\bar{x}_2)'a}{a'D_d a} \tag{11}$$

Esto es, se quiere maximizar

$$L = a'(\bar{x}_1-\bar{x}_2)(\bar{x}_1-\bar{x}_2)'a + \gamma(1-a'D_d a)$$

Las condiciones de primer orden son

$$\frac{\partial L}{\partial a} = (\bar{x}_1-\bar{x}_2)(\bar{x}_1-\bar{x}_2)'a-\gamma D_d a = 0 \tag{12}$$

$$\frac{\partial L}{\partial \gamma} = 1-a'D_d a = 0 \tag{13}$$

De la condición (12) se tiene premultiplicando por a'

$$\gamma = a'(\bar{x}_1-\bar{x}_2)(\bar{x}_1-\bar{x}_2)'a \tag{14}$$

Maximizar (11) es, por tanto, lo mismo que maximizar γ.

De (12) se ve que para resolver el problema, hay que encontrar solución al sistema de ecuaciones

$$[D_d^{-1}(\bar{x}_1-\bar{x}_2)(\bar{x}_1-\bar{x}_2)'-\gamma I]\ a=0 \qquad (15)$$

Nótese que D_d y $D_p = \frac{n_1 n_2}{n_1+n_2}(\bar{x}_1-\bar{x}_2)(\bar{x}_1-\bar{x}_2)'$ son las matrices de sumas de cuadrados dentro y entre grupos, respectivamente, para las observaciones x.

El sistema en (15) tiene solución no trivial si y sólo si

$$|D_d^{-1}(\bar{x}_1-\bar{x}_2)(\bar{x}_1-\bar{x}_2)'-\gamma I| = 0$$

Por tanto, el vector de coeficientes a de la función discriminante es el vector característico determinado por el máximo - valor característico γ del sistema de ecuaciones en (15).

Para encontrar el máximo valor característico γ, nótese que $(\bar{x}_1-\bar{x}_2)(\bar{x}_1-\bar{x}_2)'$ tiene rango 1, y por lo tanto $D_d^{-1}(\bar{x}_1-\bar{x}_2(\bar{x}_1-\bar{x}_2)'$ tiene también rango 1, de aquí se tiene que γ es igual a la traza de $D_d^{-1}(\bar{x}_1-\bar{x}_2)(\bar{x}_1-\bar{x}_2)'$.

Esto es

$$\gamma = (\bar{x}_1-\bar{x}_2)'D_d^{-1}(\bar{x}_1-\bar{x}_2) \qquad (16)$$

y el correspondiente vector característico es

$$a = D_d^{-1}(\bar{x}_1-\bar{x}_2) \qquad (17)$$

Sea $S_d = \frac{D_d}{n_1+n_2-2}$ un estimador de la matriz de covarianza dentro de grupos. Multiplicando (16) por (n_1+n_2-2), se obtiene que

$(n_1+n_2-2)\gamma=\gamma'$ es la que se conoce como distancia de Mahalanobis entre $\bar{x}_1$ y $\bar{x}_2$, i.e.

$$\gamma' = (\bar{x}_1-\bar{x}_2)' \ S_d^{-1}(\bar{x}_1-\bar{x}_2)$$

con su correspondiente vector característico

$$a = S_d^{-1}(\bar{x}_1-\bar{x}_2)$$

Así, la función discriminante de la ecuación (6) está dada, en términos de S_d por

$$y = (\bar{x}_1-\bar{x}_2)' \ S_d^{-1}x \qquad (18)$$

Cabe señalar que con este procedimiento se llegó a la función (18), pero no a una regla de decisión como en la sección anterior. La regla se daría determinando un punto c*, de tal forma que este punto se encontrara por el mínimo traslape entre valores de y. c* está dado por $c^* = \frac{1}{2}(\bar{x}_1-\bar{x}_2)'S_d^{-1}(\bar{x}_1+\bar{x}_2)$ que es el punto medio entre $\bar{x}_1$ y $\bar{x}_2$.

La regla sería enconces

clasifíquese en el grupo 1 si $(\bar{x}_1-\bar{x}_2)'S_d^{-1}x \geq \frac{1}{2}(\bar{x}_1-\bar{x}_2)'S_d^{-1}(\bar{x}_1+\bar{x}_2)$

y clasifíquese en el grupo 2 si $(\bar{x}_1-\bar{x}_2)'S_d^{-1}x < \frac{1}{2}(\bar{x}_1-\bar{x}_2)'S_d^{-1}(\bar{x}_1+\bar{x}_2)$

Nótese que c* es el valor de y en $\frac{1}{2}(\bar{x}_1+\bar{x}_2)$, coincidiendo este criterio con el establecido en (5) para el caso en que $p_1=p_2$, y con la diferencia de que en el que se acaba de encontrar se tienen estimadores de los parámetros que aparecen en (5).

4. Relaciones entre las tres técnicas.

Se han obtenido hasta ahora: la ecuación de regresión en el caso de una distribución normal y la construcción del modelo lineal a partir de esta ecuación, las correlaciones canónicas de dos variables vectoriales Y y X, y por último la función discriminante y el criterio de asignación para dos poblaciones multivariadas.

Habiendo presentado el desarrollo general de cada una de estas técnicas, es posible ahora mostrar las relaciones que existen entre ellas.

4.1 Regresión y Correlación Canónica.

En el capítulo 2 se encontraron las correlaciones canónicas asociadas a dos variables vectoriales Y y X. En vista de que el problema de regresión múltiple consiste en encontrar la máxima correlación entre un escalar y una variable vectorial, la correlación canónica es una generalización de la regresión. Se va a comprobar esta afirmación de dos maneras equivalentes. La primera es demostrando que para el caso en que la variable Y es real, la única correlación canónica que existe coincide con el coeficiente de correlación múltiple. La segunda es demostrando que el coeficiente de correlación múltiple es máximo con respecto a la correlación de Y y cualquier otra combinación lineal de X. A continuación se da la primera.

Sea $A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$ la matriz de covarianza del vector

$\begin{pmatrix} Y \\ X \end{pmatrix} \begin{matrix} (q\times 1) \\ (m\times 1) \end{matrix}$. En el capítulo 2 se encontró la primera correlación canónica entre Y y X planteando el problema

$$\max_{a,b} \frac{a'A_{12}b}{[(a'A_{11}a)(b'A_{22}b)]^{1/2}} \quad \text{sujeto a} \quad a'A_{11}a=1 \text{ y } b'A_{22}b=1$$

Esto es, se desea maximizar $L=a'A_{12}b+\lambda_1(1-b'A_{22}b)+\lambda_2(1-a'A_{11}a)$

con sus respectivas condiciones de primer orden dadas en (1), (2), (3) y (4) del capítulo 2. De estas condiciones se observó que la primera correlación canónica es la máxima solución λ_1 de

$$\begin{vmatrix} -\lambda A_{11} & A_{12} \\ A_{21} & -\lambda A_{22} \end{vmatrix} = 0$$ y que este determinante es cero si y solo si

$$\left|\lambda^2 A_{11}-A_{12}A_{22}^{-1}A_{21}\right| = 0$$

Considérese ahora el caso en que Y es una variable real, esto es $q=1$. Se tiene entonces que $\lambda^2 A_{11}-A_{12}A_{22}^{-1}A_{21}$ es un escalar, por lo tanto la única solución al determinante $|\lambda^2 A_{11}-A_{12}A_{22}^{-1}A_{21}| = 0$ es $\lambda=(A_{11})^{-1/2}(A_{12}A_{22}^{-1}A_{21})^{1/2}$, que es precisamente el coeficiente de correlación múltiple.

Ahora, de la condición (4), como $q=1$, se tiene $1-a^2A_{11}=0$, de aquí $a=A_{11}^{-1/2}$. Para encontrar el valor de b, se sustituyen los valores de a y λ en (2) obteniéndose

$$A_{21}A_{11}^{-1/2}-A_{11}^{-1/2}(A_{12}A_{22}^{-1}A_{21})^{1/2}A_{22}b = 0$$

de donde

$$b= \frac{A_{22}^{-1}A_{21}}{(A_{12}A_{22}^{-1}A_{21})^{1/2}} = B'(A_{12}A_{22}^{-1}A_{21})^{-1/2}$$

y como $A_{12}A_{22}^{-1}A_{21}$ es un escalar, B el coficiente de regresión de Y en X proporciona la combinación lineal de X que maximiza la correlación con $a'Y$ cuando $a=A_{11}^{-1/2}$.

En vista de que la correlación entre dos variables es invariante bajo la multiplicación por un escalar, la demostración anterior es equivalente a la que a continuación se presenta, de que el coeficiente de correlación múltiple es máximo respecto a la correlación de Y y cualquier otra combinación lineal de X.

Sea $b'X$ una combinación lineal de X, se va a demostrar que cuando $b=B=A_{12}A_{22}^{-1}$, $Y-B'X$ tiene varianza mínima. Para cualquier b se tiene

$$\text{var }(Y-b'X)=\text{var }[(Y-B'X)+(B'X-b'X)]$$

$$= \text{var }(Y-B'X)+\text{var}(B'X-b'X)+2 \text{ cov }(Y-B'X,\ B'X-b'X)$$

$$= A_{11\cdot 2}+(B-b)'\ A_{22}(B-b)$$

ya que

$$\text{cov }(Y-B'X,\ B'X-b'X)=$$

$$\text{cov }(Y,\ B'X)-\text{cov}(Y,b'X)-\text{cov}(B'X,B'X)+\text{cov}(B'X,b'X)$$

$$= B'A_{21}-b'A_{21}-B'A_{22}B+b'A_{22}B$$

$$= (B-b)'\ (A_{21}-A_{22}B)$$

$$= (B-b)'\ (A_{21}-A_{22}A_{22}^{-1}A_{21}) = 0$$

por tanto, cuando $b=B$

$\text{var }(Y-b'X)=A_{11\cdot 2}$ es el valor mínimo.

Ahora se va a demostrar que el coeficiente de correlación múltiple es máximo.

Por el resultado anterior se tiene que

$$\mathrm{var}\,(Y - B'X) \leqslant \mathrm{var}(Y - cb'X)$$

para cualquier c y b.

De aquí

$$A_{11} + \mathrm{var}(B'X) - 2\ \mathrm{cov}(Y, B'X) \leqslant A_{11} + c^2 \mathrm{var}(b'X) - 2c\ \mathrm{cov}(Y\ , b'X)$$

como c es un número arbitrario, esta desigualdad se cumple en particular para $c^2 = \dfrac{B'A_{22}B}{b'A_{22}b}$. Sustituyendo este valor y multiplicando por $A_{11}^{-1/2}$ se tiene

$$\frac{B'A_{22}B}{A_{11}^{-1/2}} - 2\,\frac{A_{12}B}{A_{11}^{-1/2}} \leqslant \frac{B'A_{22}B}{b'A_{22}b}\,\frac{b'A_{22}b}{A_{11}^{-1/2}} - 2\,\frac{(B'A_{22}B)^{1/2}}{(b'A_{22}b)^{1/2}}\,\frac{A_{12}b}{A_{11}^{-1/2}}$$

simplificando y multiplicando por -1

$$\frac{A_{12}B}{A_{11}^{-1/2}} \geqslant \frac{(B'A_{22}B)^{1/2}}{(b'A_{22}b)^{1/2}}\,\frac{A_{12}b}{A_{11}^{-1/2}}$$

por lo tanto

$$\frac{A_{12}B}{A_{11}^{-1/2}(B'A_{22}B)^{1/2}} \geq \frac{A_{12}b}{A_{11}^{-1/2}(b'A_{22}b)^{1/2}} \quad \text{y}$$

$$\text{corr}(Y,B'X) \geq \text{corr}(Y,b'X)$$

Ejemplo- Se va a calcular la primera correlación canónica entre Y, precio rural máximo por Kg. de maíz mejorado y X, rendimientos máximos en Kg. por Hectárea de Maíz mejorado en tierra de riego y de temporal. Esto es, se van a determinar combinaciones lineales de Y y X, de tal forma que tengan entre sí máxima correlación, en otras palabras el problema es

$$\max_{a,b} \frac{\text{cov}(aY,b'X)}{[\text{var}(aY)\text{var}(b'X)]^{1/2}} \quad \text{sujeto a} \quad \text{var}(aY)=1$$

y $\text{var}(b'X)=1$

Por lo anteriormente demostrado se sabe que

$$a=A_{11}^{-1/2} \quad \text{y} \quad b'=B'(A_{12}A_{22}^{-1}A_{21})^{-1/2}$$

A partir de $S=\begin{pmatrix} .007 & 3.24 & 1.52 \\ & 319065.8 & 67110.55 \\ & & 87907.04 \end{pmatrix}$ se obtiene

$\hat{a}=(.007)^{-1/2}=11.90$

$\hat{b}'=(.0000423)^{-1/2}\ (.00000776 \quad .00001133)$

$=153.72\ (.00000776 \quad .00001133)$

$=(.001192 \quad .001737)$

Finalmente, la correlación entre $aY=11.90Y$ y

$$\hat{b}'X=(.001192 \quad .001737)\begin{pmatrix}u_2\\u_3\end{pmatrix} = .001192u_2+.001737u_3$$

está dada por

$$\hat{\lambda}=\left(\frac{S_{12}S_{22}^{-1}S_{21}}{S_{11}}\right)^{1/2} = \left(\frac{.0000423}{.007}\right)^{1/2} = .077$$

Este resultado significa que los rendimientos máximos de maíz en tierras de riego y temporal presentan una correlación de .077 con el precio rural máximo del mismo.

Este resultado se comprueba calculando la primera correlación canónica entre Y y X como se vió en el capítulo 2.

$\lambda=a'A_{12}b$, de aquí

$$\hat{\lambda}=11.9(3.24 \quad 1.52)\begin{pmatrix}.001192\\.001737\end{pmatrix} = (38.556 \quad 18.088)\begin{pmatrix}.001192\\.001737\end{pmatrix}$$

$$= .077$$

4.2 Regresión y Análisis Discriminante.

Siguiendo la notación del capítulo 3, considérese el caso de discriminación para 2 poblaciones y sea

$x_{it} = (x_{i1t}, x_{i2t}, \ldots, x_{imt})$ la t-ésima observación de la

i-ésima población i=1,2 y $t=1,2,\ldots,n_i$. Sea Y_t una variable indicadora definida por

$$Y_t = \begin{cases} n_2/n_1+n_2 & \text{si } t=1,2,\ldots,n_1 \\ -n_1/n_1+n_2 & \text{si } t=n_1+1,\ldots,n_1+n_2 \end{cases}$$

Se va a demostrar que la ecuación de regresión de Y en X conduce al mismo procedimiento de discriminación que la función discriminante en (18) capítulo 3, excepto por una constante de proporcionalidad.

Sean $Y=(y_1,y_2,\ldots,y_{n_1+n_2})'$ de dimensión $(n_1+n_2)x1$,

$X=(x_{11},x_{12},\ldots,x_{1n_1},x_{2n_1+1},\ldots,x_{2n_1+n_2})'$ de dimensión

$(n_1+n_2)xm$, $e=(1,1,\ldots,1)'$ es un vector columna con u_1+n_2

elementos y $X^*=(e,X)$ de dimensión$(n_1+n_2)x(m+1)$.

Para encontrar $b^*=(X^{*\prime}X^*)^{-1}X^{*\prime}Y$, el coeficiente de regresión

de Y en X*, se calcula primero

$$X^{*\prime}Y = \begin{pmatrix} \sum_{t=1}^{n_1+n_2} y_t \\ \\ \sum_{i=1}^{?} \sum_{t=1}^{n_1+n_2} x'_{it} y_t \end{pmatrix} = \begin{pmatrix} 0 \\ \\ \frac{1}{n_1+n_2}(n_2 \sum_{t=1}^{n_1} x'_{1t} - n_1 \sum_{s=1}^{n_2} x'_{2n_1+s} \end{pmatrix}$$

$$X^{*\prime}Y = \begin{pmatrix} 0 \\ \\ \frac{n_1 n_2}{n_1+n_2} \; (\bar{x}_1 - \bar{x}_2) \end{pmatrix} \tag{1}$$

Por otro lado

$$X^{*\prime}X^{*}b^{*} = \begin{pmatrix} e'e \; b_o + e'Xb \\ X'eb_o \; + X'Xb \end{pmatrix} \tag{2}$$

donde $b^{*} = \begin{pmatrix} b_o \\ b \end{pmatrix} \begin{matrix} (1x1) \\ (mx1) \end{matrix}$ de (1) y (2) y de las ecuaciones normales

$X^{*\prime}X^{*}b^{*} = X^{*}Y$ se obtiene

$$e'eb_o + e'Xb = 0$$

$$X'eb_o + X'Xb = \frac{n_1 n_2}{n_1+n_2} \; (\bar{x}_1 - \bar{x}_2) \tag{3}$$

por tanto, $\quad b_o = \frac{-e'X}{e'e} \; b$

sustituyendo esta expresión en (3) se tiene

$$\left[\; X'X - X'e \; \frac{(e'X)}{e'e} \; \right] b = \frac{n_1 n_2}{n_1+n_2} \quad (\bar{x}_1 - \bar{x}_2) \qquad ;$$

$$\left[\frac{X'X}{n_1+n_2} - \frac{(X'e)(e'X)}{(n_1+n_2)^2} \right] b = \frac{n_1 n_2}{(n_1+n_2)^2} (\bar{x}_1-\bar{x}_2) \qquad ;$$

de aquí se obtiene $\quad b=n_1n_2/(n_1+n_2)^2\ G^{-1}(\bar{x}_1-\bar{x}_2) \qquad (4)$

donde $\quad G= \dfrac{X'X}{n_1+n_2} - \dfrac{(X'e)(e'X)}{(n_1+n_2)^2}$

Sea $\bar{x}$ la media total de las observaciones x_{it}, entonces

$$G = \frac{1}{n_1+n_2} \sum_{i=1}^{2} \sum_{t=1}^{n_i} (x_{it}-\bar{x})(x_{it}-\bar{x})'$$

$$= \frac{x_1x'_1+x_2x'_2}{n_1+n_2} + \frac{n_1n_2}{(n_1+n_2)^2} (\bar{x}_1-\bar{x}_2) \ (\bar{x}_1-\bar{x}_2)' \qquad (5)$$

es un estimador de la matriz de covarianza A

Bajo la hipótesis $\mu_1=\mu_2$, un estimador de A está dado por

$S_d=\dfrac{x_1x'_1+x_2x'_2}{n_1+n_2-2}$, dentro de este contexto la expresión en (4)

queda $\quad \mathring{b}= \dfrac{n_1n_2}{(n_1+n_2)(n_1+n_2-2)}\ S_d^{-1}\ (\bar{x}_1-\bar{x}_2)$

Por otro lado

$$\mathring{b}'X^{*\prime}X^{*}\mathring{b} = \mathring{b}\ X^{*\prime}\ Y$$

$$= \frac{n_1n_2}{n_1+n_2}\ \mathring{b}'\ (\bar{x}_1-\bar{x}_2)$$

$$= \frac{(n_1 n_2)^2}{(n_1+n_2)^2 (n_1+n_2-2)} (\bar{x}_1-\bar{x}_2)' S_d^{-1} (\bar{x}_1-\bar{x}_2) \qquad (8)$$

esto es, la suma de cuadrados de la regresión es un múltiplo de la función discriminante calculada en $\bar{x}_1-\bar{x}_2$, y por lo tanto puede ser utilizada para probar la significancia de esta función.

Bajo $H_o: \mu_1=\mu_2$

$$\frac{(n_1+n_2-m-1)(n_1+n_2)}{n_1 n_2 m} \mathring{b}' X^{*\prime} X^* \mathring{b} \sim F_{m, n_1+n_2-m-1} \quad ,$$

ya que

$$T^2 = \frac{n_1 n_2}{n_1+n_2} (\bar{x}_1-\bar{x}_2)' S_d^{-1} (\bar{x}_1-\bar{x}_2)$$

y $$\frac{n_1+n_2-m-1}{(n_1+n_2-2)m} T_2 \sim F_{m, n_1+n_2-m-1}$$

Por lo tanto, para probar $H_o: \mu_1=\mu_2$

$H_a: \mu_1 \neq \mu_2$

a un nivel de significancia α, se rechaza H_o si

$$\frac{(n_1+n_2-m-1)(n_1+n_2)}{n_1 n_2 m} \; b^{*\prime} X^{*\prime} X^{*} b^{*} \begin{cases} > F_{1-\alpha/2,m,n_1+n_2-m-1} \\ \text{o} \\ < F_{\alpha/2,m,n_1+n_2-m-1} \end{cases}$$

Si se rechaza H_o, entonces tiene sentido continuar con el análisis discriminante, ya que significaría que efectivamente existe una diferencia entre las dos poblaciones.

A continuación se describe un trabajo realizado por Smith (1947). Ejemplo - Se aplicaron ciertas pruebas a un grupo de 25 individuos normales y 25 psicóticos, y se determinaron para cada individuo dos variables x_1 y x_2, la primera mide grado de enfermedad y la segunda tipo de enfermedad. Los resultados fueron

NORMALES		PSICOTICOS	
x_1	x_2	x_1	x_2
22	6	24	38
20	14	19	36
23	9	11	43
23	1	6	60
17	8	9	32
24	9	10	17
23	13	3	17
18	18	15	56
22	16	14	43
19	18	20	8

##

	NORMALES x_1	NORMALES x_2	PSICOTICOS x_1	PSICOTICOS x_2
	20	17	8	46
	20	31	20	62
	21	9	14	36
	13	13	3	12
	20	14	10	51
	19	15	22	22
	20	11	11	30
	18	17	6	30
	20	7	20	61
	23	6	20	43
	23	23	15	43
	25	4	5	53
	23	5	10	43
	21	12	13	19
	23	7	12	4
Totales	520	303	320	905
Media	20.8	12.12	12.8	36.2

$$S_d = \begin{pmatrix} 21.83 & 3.66 \\ & 163.68 \end{pmatrix}$$

la inversa de S_d es

$$S_d^{-1} = \begin{pmatrix} .04598 & -.001028 \\ & .006132 \end{pmatrix}$$

para la diferencia de medias (normales-psicóticas) se tiene

$$\bar{x}_1 - \bar{x}_2 = \begin{pmatrix} 8.0 \\ -24.08 \end{pmatrix}$$

Entonces, los coeficientes de la función discriminante están dados por

$$(8.0 \quad -24.08) \begin{pmatrix} .04598 & -.001028 \\ & .006132 \end{pmatrix}$$

$$= (.39 \quad -.16)$$

Supóngase que las probabilidades de ser un individuo normal o un individuo psicótico son iguales, entonces, dada una observación (x,y), la regla de decisión es:

asígnese (x,y) al grupo de individuos normales si

$$.39x - .16y \geq \frac{1}{2} (8.0 \quad -24.08) \begin{pmatrix} .04598 & -.001028 \\ & .006132 \end{pmatrix} \begin{pmatrix} 33.6 \\ 48.32 \end{pmatrix}$$

$$= 2.825$$

de otra manera asígnese al grupo de individuos psicóticos.

Ahora se va a probar la significancia de la función discriminante

$$.39x-.16y$$

Por (25)

$$\frac{(n_1+n_2-m-1)(n_1+n_2)}{n_1n_2m}\ \mathring{b}'X^{*\prime}X^*\mathring{b} =$$

$$= \frac{(n_1+n_2-m-1)n_1n_2}{(n_1+n_2-2)(n_1+n_2)m}\ (\bar{x}_1-\bar{x}_2)'S_d^{-1}(\bar{x}_1-\bar{x}_2)$$

$$= 6.12\ (8.0\quad -24.08)\begin{pmatrix} .04598 & -.001028 \\ & .006132 \end{pmatrix}\begin{pmatrix} 8.0 \\ -24.08 \end{pmatrix}$$

$$= 6.12\ (.39\quad -.16)\begin{pmatrix} 8.0 \\ -24.08 \end{pmatrix}$$

$$= 6.12\ (6.97)$$

$$= 42.66$$

Sea $\alpha=.05$, entonces $F_{.975,2,47}=39.5$ y

$$F_{.025,2,47}=4.18$$

Por lo tanto, como $42.66>39.5$, se rechaza la hipótesis de que las medias de las dos poblaciones sean iguales. Esto a su vez implica que la función discriminante sí es significativa.

4.3 Correlación Canónica y Análisis Discriminante.

Caso para 2 poblaciones.

En el capítulo 2 se vió que las correlaciones canónicas de dos variables Y (qx1) y X(mx1) se estiman calculando los valores característicos de la matriz $S_{11}^{-1}S_{12}S_{22}^{-1}S_{21}$, donde $S= \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}$

es el estimador corregido por la media, de la matriz de covarianza del vector $\begin{pmatrix} Y \\ X \end{pmatrix}$. Esto es, las correlaciones se estiman resolviendo

$$|S_{11}^{-1}S_{12}S_{22}^{-1}S_{21}-\lambda^2 I| = 0 \qquad (1)$$

Siguiendo la notación del capítulo 3, sean x_{1t}(mx1) observaciones de un grupo π_1, con $t=1,\dots,n_1$ y x_{2t}(mx1) observaciones de un grupo π_2, con $t=n_1+1,\dots,n_1+n_2$. Sean D_d,$\bar{x}_1$ y $\bar{x}_2$ como se definieron anteriormente. En el capítulo 3 se mostró que la máxima raíz de la ecuación

$$|D_d^{-1}(\bar{x}_1-\bar{x}_2)\ (\bar{x}_1-\bar{x}_2)'-\gamma I| = 0 \qquad (2)$$

determina los coeficientes de la función discriminante.

Sea $D_p= \frac{n_1n_2}{n_1+n_2}\ (\bar{x}_1-\bar{x}_2)(\bar{x}_1-\bar{x}_2)'$ la matriz de sumas de cuadrados y productos entre poblaciones. La ecuación (2) se puede escribir entonces como

$$\left| \frac{n_1+n_2}{n_1 n_2} \quad D_d^{-1} D_p - \gamma I \right| = 0 \qquad (3)$$

Ahora, en el contexto de correlación canónica sea $q=1$ y defínase $Y_t=0$ si la t-ésima observación de la variable X pertenece al grupo π_1 y $Y_t=1$ si esto no sucede.

Debe notarse que por simplicidad, aquí conviene una definición de Y_t diferente de la utilizada en análisis discriminante.

Se va a demostrar que bajo esta definición, las soluciones de (1) y (3) guardan la relación

$$\gamma = \frac{\lambda^2}{1-\lambda^2} \qquad (4)$$

Por un lado

$$D_{11} = \sum_{t=1}^{n_1+n_2} (y_t-\bar{y})(y_t-\bar{y})'$$

$$= \sum_{t=1}^{n_1} (y_t-\bar{y})(y_t-\bar{y})' + \sum_{t=n_1+1}^{n_1+n_2} (y_t-\bar{y})(y_t-\bar{y})'$$

$$= \left[\left(\frac{-n_1 n_2}{n_1+n_2}\right)\left(\frac{-n_2}{n_1+n_2}\right) + n_2\left(1-\frac{n_2}{n_1+n_2}\right)\left(1-\frac{n_2}{n_1+n_2}\right)\right]$$

$$= \frac{n_1 n_2^2}{(n_1+n_2)^2} + \frac{n_1^2 n_2}{(n_1+n_2)^2}$$

$$= \frac{n_1 n_2}{n_1+n_2}$$

y

$$D_{12} = [\sum_{t=1}^{n_1} (x_{1t}-\bar{x})'(y_{1t}-\bar{y}) + \sum_{t=n_1+1}^{n_1+n_2} (x_{2t}-\bar{x})'(y_{2t}-\bar{y})]$$

$$= [\frac{-n_2}{n_1+n_2} \sum_{t=1}^{n_1} (x_{1t}-\bar{x})' + (1-\frac{n_2}{n_1+n_2}) \sum_{t=n_1+1}^{n_1+n_2} (x_{2t}-\bar{x})']$$

$$= [\frac{-n_2 n_1}{n_1+n_2} (\bar{x}_1-\bar{x})' + \frac{n_1 n_2}{n_1+n_2} (\bar{x}_2-\bar{x})']$$

$$= \frac{-n_2 n_1}{n_1+n_2} (\bar{x}_1-\bar{x})' + \frac{n_1 n_2}{n_1+n_2} (\bar{x}_2-\bar{x}_1+\bar{x}_1-\bar{x})'$$

$$= \frac{n_1 n_2}{n_1+n_2} (\bar{x}_2-\bar{x}_1)'$$

De aquí se tiene

$$D_{11}^{-1} D_{12} D_{22}^{-1} D_{21} = \frac{n_1+n_2}{n_1 n_2} \frac{n_1 n_2}{n_1+n_2} \frac{n_1 n_2}{n_1+n_2} (\bar{x}_2-\bar{x}_1)' D_{22}^{-1} (\bar{x}_2-\bar{x}_1)$$

$$= \frac{n_1 n_2}{n_1+n_2}(\bar{x}_2-\bar{x}_1)' D_{22}^{-1}(\bar{x}_2-\bar{x}_1) \qquad (5)$$

Como $S_{11} = \frac{1}{n_1+n_2-2} D_{11}$, $S_{12} = \frac{1}{n_1+n_2-2} D_{12}$ y $S_{22} = \frac{1}{n_1+n_2-2} D_{22}$,

$$S_{11}^{-1} S_{12} S_{22}^{-1} S_{21} = \frac{n_1 n_2}{n_1+n_2}(\bar{x}_2-\bar{x}_1)' D_{22}^{-1}(\bar{x}_2-\bar{x}_1)$$

Por otro lado, $D_{22} = \sum_{it} (x_{it}-\bar{x})(x_{it}-\bar{x})'$ es igual a la matriz de sumas de cuadrados total, la cual se denota por $D = Dp + D_d$. A partir de (5), la ecuación (1) se escribe como

$$|\frac{n_1 n_2}{n_1+n_2} (\bar{x}_2-\bar{x}_1)' D^{-1}(\bar{x}_2-\bar{x}_1) - \lambda^2 I| = 0 \qquad (6)$$

Ahora bien, las raíces de la ecuación (6) son las mismas que las de la ecuación

$$\left|\frac{n_1 n_2}{n_1+n_2}(\bar{x}_2-\bar{x}_1)(\bar{x}_2-\bar{x}_1)'\ D^{-1} - \lambda^2 I\right| = 0$$

de aquí se obtiene

$$\begin{aligned}
&|Dp\ D^{-1} - \lambda^2 I| = 0\\
&|Dp - \lambda^2 D| = 0\\
&|Dp - \lambda^2(Dp+D_d)| = 0\\
&|(1-\lambda^2)Dp - \lambda^2 D_d| = 0\\
&|(1-\lambda^2)D_d^{-1}Dp - \lambda^2 I| = 0\\
&\left|D_d^{-1}\ Dp - \frac{\lambda^2}{1-\lambda^2}\ I\right| = 0 \qquad (7)
\end{aligned}$$

Por lo tanto, de (3) y (7) se tiene

$$\gamma = \frac{\lambda^2}{1-\lambda^2}$$

Para ilustrar el resultado establecido, considérese el siguiente

Ejemplo.- En la sección anterior se encontró que la función discriminante para discriminar entre los individuos normales y psicóticos es

$$.39x-.16y$$

El valor de γ es, según la ecuación (16) del capítulo 3

$$\begin{aligned}
\gamma &= .39(8.0)-.16(-24.08)\\
&= 3.12 + 3.85\\
&= 6.97
\end{aligned}$$

Por lo tanto, a partir de la ecuación (4) se tiene

$$\lambda^2 = \frac{6.97}{1 + 6.97} = .874$$

Entonces, la máxima correlación canónica entre

$$Y_t = \begin{cases} 0 & \text{si} \quad x_t \in \pi_1 \\ 1 & \text{si} \quad x_t \notin \pi_1 \end{cases} \quad \text{y} \quad X \quad \text{es igual a } .874$$

A continuación se muestra que se llega al mismo resultado, calculando la correlación canónica entre Y_t y X.

En el capítulo 2 se vió que cuando Y es un real, la correlación canónica está dada por $(A_{11})^{-1/2}(A_{12}\, A_{22}^{-1}\, A_{21})^{-1/2}$. Estimadores de estas matrices son

$$S_{11} = \frac{1}{n_1+n_2} \sum_{t=1}^{n_1+n_2} (y_t-\bar{y})(y_t-\bar{y})'$$

$$= \frac{n_1 n_2}{(n_1+n_2)^2}$$

$$S_{12} = \frac{1}{n_1+n_2} \sum_{t=1}^{n_i} (x_{it}-\bar{x})'(y_{it}-\bar{y})$$

$$= \frac{n_1 n_2}{(n_1+n_2)^2} (\bar{x}_2-\bar{x}_1)'$$

$$S_{22} = \frac{1}{n_1+n_2} \sum_{i=1}^{2} \sum_{t=1}^{n_i} (x_{it}-\bar{x})(x_{it}-\bar{x})'$$

De aquí se tiene que la primera correlación canónica es

$$\lambda^2 = (S_{11})^{-1/2}(S_{12}\, S_{22}^{-1}\, S_{21})^{-1/2}$$

$$= (.25)^{-1/2}(-2.0 \quad 6.02) \begin{pmatrix} 36.96 & 51.67 \\ 51.67 & 302.09 \end{pmatrix}^{-1} \begin{pmatrix} -2.0 \\ 6.02 \end{pmatrix}$$

$$= 2(-2.0 \quad 6.02) \begin{pmatrix} .0355 & -.006 \\ -.006 & .0043 \end{pmatrix} \begin{pmatrix} -2.0 \\ 6.02 \end{pmatrix}$$

$$= 2(-.107 \quad .037) \begin{pmatrix} -2.0 \\ 6.02 \end{pmatrix}$$

$$= 2(.436)$$

$$= .872$$

La relación entre correlación canónica y análisis discriminante se utiliza como un método para discriminar entre R poblaciones.

A continuación se presenta este método.

Caso para R poblaciones (R>2)

Sea $x \in \mathcal{R}^m$ una observación que pertenece a una de $R=q+1$ poblaciones, y sea $y \in \mathcal{R}^q$ un vector que tiene valor uno en la α-ésima componente y cero en las demás, si la observación x pertenece a la población π_α, $\alpha=1,2,\ldots,q$. Si la observación proviene de π_R, entonces el vector y es el vector cero.

Sean $x_{\alpha t}$ la t-ésima observación de la población π_α, $t=1,2,\ldots,n_\alpha$ y $y_{\alpha t}$ la correspondiente observación de y,

$$\bar{x} = \frac{1}{N} \sum_{\alpha=1}^{R} \sum_{t=1}^{n_\alpha} x_{\alpha t} \qquad y \qquad \bar{y} = \frac{1}{N} \sum_{\alpha=1}^{R} \sum_{t=1}^{n_\alpha} y_{\alpha t}$$

Un estimador de la matriz de covarianza del vector $\begin{pmatrix} x \\ y \end{pmatrix}$ en base a las $N = \sum_{\alpha=1}^{R} n_\alpha$ observaciones está dado por $S = \frac{1}{N-R}\begin{pmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{pmatrix}$,

donde $D_{11} = \sum_{\alpha=1}^{R} \sum_{t=1}^{n_\alpha} (x_{\alpha t}-\bar{x})(x_{\alpha t}-\bar{x})'$,

$$D_{22} = \sum_{\alpha=1}^{R} \sum_{t=1}^{n_\alpha} (y_{\alpha t}-\bar{y})(y_{\alpha t}-\bar{y})'$$

$$= \sum_{\alpha=1}^{R} \sum_{t=1}^{n_\alpha} y_{\alpha t}\, y'_{\alpha t} - N\,\bar{y}\,\bar{y}'$$

$$= \mathrm{diag}(n_1,n_2,\ldots,n_q) - \frac{1}{N}(n_1,n_2,\ldots n_q)'(n_1,n_2\ldots,n_q)$$

$$D_{12} = \sum_{\alpha=1}^{R} \sum_{t=1}^{n_\alpha} (x_{\alpha t}-\bar{x})(y_{\alpha t}-\bar{y})'$$

$$= \sum_{\alpha=1}^{R} \sum_{t=1}^{n_\alpha} x_{\alpha t}\, y'_{\alpha t} - N\bar{x}\bar{y}'$$

$$= (n_1\bar{x}_1, n_2\bar{x}_2,\ldots,n_q\bar{x}_q) - (n_1\bar{x}, n_2\bar{x},\ldots,n_q\bar{x})$$

además $D_{22}^{-1} = \mathrm{diag}(\frac{1}{n_1}, \frac{1}{n_2}, \ldots, \frac{1}{n_q}) + \frac{1}{n_R} E_{qxq}$, donde E es una matriz de unos de dimensión qxq.

A partir de estas igualdades se ve que $D_{11}=D$ es la matriz de sumas de cuadrados para las observaciones x, además

$$D_{12}D_{22}^{-1}D_{21} = [n_1(\bar{x}_1-\bar{x}), n_2(\bar{x}_2-\bar{x}),\ldots,n_q(\bar{x}_q-\bar{x})]$$

$$[\mathrm{diag}(\frac{1}{n_1},\frac{1}{n_2},\ldots,\frac{1}{n_q}) + \frac{1}{n_R} E_{qxq}][n_1(\bar{x}_1-\bar{x}),\ldots,n_q(\bar{x}_q-\bar{x})]'$$

$$= \sum_{\alpha=1}^{q} n_\alpha(\bar{x}_\alpha-\bar{x})(\bar{x}_\alpha-\bar{x})' + [n_1(\bar{x}_1-\bar{x}),\ldots,n_q(\bar{x}_q-\bar{x})]$$

$$\frac{1}{n_R} E_{qxq}[n_1(\bar{x}_1-\bar{x}),\ldots,n_q(\bar{x}_q-\bar{x})]'$$

$$= \sum_{\alpha=1}^{q} n_\alpha(\bar{x}_\alpha-\bar{x})(\bar{x}_\alpha-\bar{x})' + \sum_{\alpha=1}^{q} n_\alpha(\bar{x}_\alpha-\bar{x}) \sum_{\alpha=1}^{q} \frac{n_\alpha}{n_R}(\bar{x}_\alpha-\bar{x})'$$

desarrollando el segundo miembro en (8) se tiene

$$\sum_{\alpha=1}^{q} n_\alpha(\bar{x}_\alpha-\bar{x}) \sum_{\alpha=1}^{q} \frac{n_\alpha}{n_R}(\bar{x}_\alpha-\bar{x})' = (\sum_{\alpha=1}^{q} n_\alpha\bar{x}_\alpha - \sum_{\alpha=1}^{q} n_\alpha\bar{x})(\sum_{\alpha=1}^{q} \frac{n_\alpha}{n_R}\bar{x}_\alpha - \sum_{\alpha=1}^{q} \frac{n_\alpha}{n_R}\bar{x})'$$

$$= (N\bar{x}-n_R\bar{x}_R-(N-n_R)\bar{x})(\frac{N}{n_R}\bar{x}-\bar{x}_R-\frac{(N-n_R)}{n_R}\bar{x})'$$

$$= (n_R\bar{x}-n_R\bar{x}_R)(\bar{x}-\bar{x}_R)' = \frac{n_R}{n_R}(n_R\bar{x}-n_R\bar{x}_R)(\bar{x}-\bar{x}_R)'$$

$$= n_R(\bar{x}-\bar{x}_R)(\bar{x}-\bar{x}_R)'$$

por lo tanto

$$D_{12}D_{22}^{-1}D_{21} = \sum_{\alpha=1}^{q} n_\alpha(\bar{x}_\alpha-\bar{x})(\bar{x}_\alpha-\bar{x})' + n_R(\bar{x}_R-\bar{x})(\bar{x}_R-\bar{x})'$$

$$= \sum_{\alpha=1}^{R} n_\alpha(\bar{x}_\alpha-\bar{x})(\bar{x}_\alpha-\bar{x})'$$

$$= Dp \qquad (8)$$

En el capítulo 3 se argumentó que una forma de obtener la función discriminante entre dos poblaciones es maximizando el cociente de la variación entre grupos, dividida por la variación dentro de los grupos. Esta afirmación es también válida para el caso de R poblaciones (R>2)

Se tiene entonces que la forma de determinar los coeficientes de la función discriminante es maximizando la razón $\frac{a'Dpa}{a'D_d a}$ donde

$D_d = \sum_{\alpha=1}^{R} \sum_{t=1}^{n_\alpha} (x_{\alpha t}-\bar{x}_\alpha)(x_{\alpha t}-\bar{x}_\alpha)'$. Para evitar tener una infinidad de soluciones, se impone la restricción $a'(Dp+D_d)a=1$. Se va a demostrar que si v es el vector que maximiza la razón anterior, entonces también maximiza $\frac{a'Dpa}{a'(Dp+D_d)a}$ sujeto a $a'(Dp+D_d)a=1$. Sea w el máximo de esta expresión entonces

$$w'Dpw = \frac{w'Dpw}{w'(Dp+D_d)w} \geq \frac{v'Dpv}{v'(Dp+D_d)v} = v'Dpv$$

como $v'D_dv = 1 - v'Dpv$ y

$$w'D_dw = 1 - w'Dpw$$

de la desigualdad anterior se tiene que

$$w'D_dw \leq v'D_dv$$

de aquí se obtiene

$$\frac{w'Dpw}{w'D_dw} \geq \frac{v'Dpv}{v'D_dv}$$

como v es el máximo de $\frac{a'Dpa}{a'D_da}$

$$\frac{w'Dpw}{w'D_dw} = \frac{v'Dpv}{v'D_dv}$$

por lo tanto v=w.

En base a este resultado el problema es maximizar

$$L = a'Dpa + \gamma[1-a'(Dp+D_d)a]$$

sea $D = Dp+D_d$

Las primeras derivadas parciales igualadas a cero son

$$\frac{\partial L}{\partial a} = Dpa - \gamma Da = 0 \qquad (9)$$

$$\frac{\partial L}{\partial \gamma} = 1-a'Da = 0 \qquad (10)$$

de la condición (9) se tiene premultiplicando por a'

$$\gamma = a'\ Dp\ a$$

por lo tanto maximizar $\frac{a'Dpa}{a'Da}$ sujeta a $a'Da=1$ es lo mismo que maximizar γ. De (9) se ve que para resolver el problema hay que encontrar solución al sistema de ecuaciones

$$[D^{-1}Dp - \gamma I]a = 0 \qquad (11)$$

este sistema tiene solución no trivial si y sólo si

$$|D^{-1}Dp - \gamma I| = 0 \qquad (12)$$

De la igualdad (8) se tiene que

$$D^{-1}Dp = D_{11}^{-1}\ D_{12}\ D_{22}^{-1}\ D_{21} = S_{11}^{-1}\ S_{12}\ S_{22}^{-1}\ S_{21}$$

por lo tanto (12) se escribe como

$$|S_{11}^{-1}\ S_{12}\ S_{22}^{-1}\ S_{21} - \gamma I| = 0 \qquad (13)$$

de (1) se ve que $\gamma=\lambda^2$, esto es, la función discriminante se encuentra obteniendo las correlaciones canónicas entre x y y. Como (13) tiene $f=\min(m,q)$ raices $\lambda_1^2>\lambda_2^2>\ldots>\lambda_f^2$, correspondiente a cada raíz λ_i^2 existe un vector a_i que satisface (11), se tienen así f funciones discriminantes $a_i'x$ $i=1,2,\ldots,f$.

Ya que λ_1^2 es la máxima raíz de (13), la correspondiente función $a_1'x$ proporciona la máxima separación entre las medias de las pobla ciones siendo así útil como una función discriminante. Sin embargo, esto no es suficiente ya que la siguiente función discriminante $a_2'x$ también proporciona separación entre estos grupos en una dirección diferente y es también necesaria para la discriminación. Análogamente con $a_3'x$, y así sucesivamente. La pregunta es ahora si todas estas funciones discriminantes son necesarias, o si son suficientes algunas de ellas para el análisis. La respuesta a esta pregunta depende de la habilidad para discriminar de estas funciones. En el caso de R poblaciones λ_i^2 mide la habilidad para discriminar de $a_i'x$.

Para determinar el número de raices significativas y por tanto el número de funciones discriminantes a ser utilizadas, se hacen pruebas de significancia sobre las correlaciones canónicas que a continuación se presentan.

Ya que las correlaciones canónicas miden el grado de correlación entre x_{mx1} y y_{qx1}, si las dos variables son independientes, la matriz de covarianza D_{12} y la matriz de correlación R_{12} son matrices cero y por lo tanto las correlaciones canónicas serán cero. Para probar la significancia de las q correlaciones canónicas Bartlett define Λ_o como $\Lambda_o = \prod_{i=1}^{q} (1-\lambda_i^2)$. Para muestras grandes, la prueba de la hipótesis nula de que x no está correlacionada con y, se hace en base a $\chi^2 = -[n-1- \frac{1}{2}(p+q+1)] \ln \Lambda_o$ que se distribuye como chi-cuadrada con pq grados de libertad. Si $\chi^2 > \chi^2_{\alpha,pq}$, entonces por lo menos el primer coeficiente de correlación canónica es significativo a un nivel α. Para probar la significancia de los restantes q-1 coeficientes, se forma $\Lambda_1 = \prod_{i=2}^{q} (1-\lambda_i^2)$, que

Para muestras grandes se distribuye como $\chi^2 = -[n-2-\frac{1}{2}(p+q+1)]\ln \Lambda_1$ con $(p-1)(q-1)$ grados de libertad. En general, si se ha rechazado que K correlaciones canónicas sean igual a cero, el criterio para probar que los otros q-K coeficientes son cero es $\Lambda_K = \prod_{i=K+1}^{q} (1-\lambda_i^2)$ que se distribuye para grandes muestras como $\chi^2 = -[n-1-K-\frac{1}{2}(p+q+1)]\ln \Lambda_K$ con $(p-K)(q-K)$ grados de libertad.

Otro tipo de prueba de significancia es la que prueba únicamente la correlación más grande. Para este caso, Marriot (1952) demostró que

$$[-n-1+\frac{1}{2}(p+q+1)]\ln(1-\lambda_1)$$

se distribuye aproximadamente como una χ^2 con $p+q-1+\frac{1}{2}[(p-1)(q-1)]^{2/3}$ grados de libertad.

Esta prueba puede ser utilizada para probar la K+1-ésima correlación canónica, cuando se ha demostrado que las primeras K son significativas, reemplazando p y q por p-K y q-K.

Cabe señalar que el número de funciones significativas s es igual a la dimensión del hiperplano sobre el cual están las medias de las R poblaciones. Esto se puede ver de la siguiente manera.

Los R puntos correspondientes a las R medias de las poblaciones, en general no están sobre una recta, y por lo tanto una sola función discriminante, en general, no será adecuada. Si los puntos están sobre un hiperplano s-dimensional, se requerirán s funciones discriminantes que separen los R puntos en s direcciones ortogonales distintas.

Ahora se considerará el método con el cual se asigna una nueva observación x_o a uno de los R grupos, en base a las s funciones discriminantes $a_i'x$. Sustituyendo x_o por x en las funciones $a_i'x$ se tiene

$$u_{o1} = a_1'\, x_o,\ u_{o2} = a_2'x_o, \ldots, u_{os} = a_s'x_o$$

que son las "nuevas" coordenadas del punto x_o con respecto al plano s-dimensional sobre el cual están las medias de los R grupos. Análogamente, las "nuevas" coordenadas de las medias muestrales $\bar{x}_\alpha$ son

$$u_{\alpha 1} = a_1'\, \bar{x}_\alpha,\ u_{\alpha 2} = a_2'\, \bar{x}_\alpha, \ldots, u_{\alpha s} = a_s'\, \bar{x}_\alpha$$

Sea $\Delta_{\alpha o}$ la distancia entre el punto x_o y la media $\bar{x}_\alpha$ en base a estas nuevas coordenadas, entonces

$$\Delta_{\alpha o}^2 = \sum_{i=1}^{s} (u_{\alpha i} - u_{oi})^2 \quad \alpha=1,2,\ldots,R$$

La regla de decisión es entonces asígnese x_o a π_a si el punto x_o está más cercano a la media de π_a que a la media de cualquier otra π_α. Esto es, asígnese x_o a π_a si

$$\Delta_{ao}^2 = \min(\Delta_{10}^2,\ \Delta_{20}^2, \ldots, \Delta_{R0}^2)$$

Conclusiones.

El modelo de regresión lineal múltiple es ampliamente utilizado para fines de predicción de fenómenos de diversa índole. A lo largo de este trabajo se ha mostrado la relación que el modelo tiene con la explicación de técnicas tales como: regresión multivariada, análisis discriminante, correlación canónica y correlación múltiple.

Se ha presentado a la regresión multivariada como una generalización de la múltiple en el sentido de que la primera predice una variable vectorial en términos de otra vectorial, mientras que la segunda se reduce a la anterior en el caso de que la variable dependiente sea de dimensión uno.

La correlación canónica puede interpretarse como una extensión natural de la correlación múltiple. En regresión se vió que la variable a ser predecida es unidimensional, pero pueden existir situaciones en que la variable a ser predecida sea un vector. El problema consiste entonces en encontrar una combinación lineal de la variable vectorial dependiente, que tenga correlación máxima con una combinación lineal de la variable vectorial independiente, esto es, determinar la primera correlación canónica entre las dos variables.

Por otra parte, se consideró la matriz X formada por todas las observaciones provenientes de dos poblaciones π_1 y π_2, y se de finió una variable indicadora Y que toma valores n_1/n_1+n_2 y $-n_2/n_1+n_2$, dependiendo de si la observación pertenece a π_1 o π_2. Se hizo la regresión de Y en X y se mostró que el vector de coeficientes de regresión es un múltiplo de la función discriminante, y que la suma de cuadrados $b^{\circ\prime} X^{*\prime} X^{*} b^{\circ}$ permite probar la hipótesis de igualdad de medias de las dos poblaciones, y por lo tanto la posibilidad de discriminar en base a la función discriminante.

Para el caso de R poblaciones, se definió una variable indicadora $y \in \mathbb{R}^q$, que tiene valor uno en la α-ésima componente y cero en las demás, si la observación $x \in \mathbb{R}^m$ pertenece a la población π_α, $\alpha=1,2,\ldots q$. Se demostró que existen f=min(m,q) funciones discriminantes que se encuentran obteniendo las correlaciones canónicas entre x y y. Además se vió que gracias a esta relación, para determinar el número de funciones significativas, se hacen pruebas de significancia sobre las correlaciones canónicas.

Finalmente, se mostró que en la regresión, la importancia de B y $\hat{B}$ es obvia, ya que son las que determinan la matriz de coeficientes de regresión y su estimador, respectivamente. En correlación canónica se mostró que las correlaciones están determinadas por las soluciones de la ecuación $|S_{11}^{-1}S_{12}S_{22}^{-1}S_{21}-\lambda^2 I|=0$, donde la matriz $S_{11}^{-1}S_{12}S_{22}^{-1}S_{21}=S_{11}^{-1}S_{12}\hat{B}'$ es una combinación lineal de $\hat{B}'$, y en el caso particular $q=1, \lambda^2=S_{11}^{-1}S_{12}\hat{B}'$.

BIBLIOGRAFIA CONSULTADA

-Anderson, T.W., 1958, An Introduction to Multivariate Statistical Analysis, John Wiley & Sons, Inc.

-Bolch & Huang, 1975, Multivariate Statistical Methods for Business and Economics, Prentice Hall.

-Bryant & Atchley, 1977, Multivariate Statistical Methods: Within-groups Covariation, Dowden, Hutchinson & Ross.

-Cooley, William, Lohnes, Paul, 1971, Multivariate Data Analysis, John Wiley & Sons, Inc.

-Dhrymes, Phoebus J., 1970, Econometrics, Harper & Row.

-De Wall, D.J., 1975, Parametric Multivariate Analysis, , Part 1, Department of Statistics University of North Carolina at Chapel Hill and the University of the Orange Free State Bloemfontein, South Africa.

-Hérnandez Avila, 1976, The Use of Canonical Correlation to Investigate the Relations Between the Soil and its Environment.

-Hilary Seal, 1964, Multivariate Statistical Analysis for Biologists, Methuen and Co. Ltd.

-Kendall & Stuart, 1976, The Advanced Theory of Statistics, Vols. 1 y 2, Hafner Publishing Company.

-Kshirsagar, A., 1972, Multivariate Analysis. Statistics: Textbooks and Monographs, Vol. 2., Marcel Dekker Inc.

-Marriot, F.H.C., 1952, "Test of Significance in Canonical Analysis", Biometrika, 39, 58-64.

-Marriot, F.H.C., 1974, The Interpretations of Multiple Observations, Academic Press.

-Maxwell, A.E., 1977, Multivariate Analysis in Behavioural Research, Chapman and Hall.

-Mejía, A.C., 1976, Presentación de Regresión Multivariada como una Extensión de Regresión Múltiple, Centro de Estadística y Cálculo del Colegio de Postgraduados, SARH, Chapingo, Mex.

-Morrison, F. Donald, 1967, Multivariate Statistical Methods, McGraw Hill Book Company.

-Press, S. James, 1972, Applied Multivariate Analysis, Holt Rinehart & Wiston Inc.

-Smith, C.A.B., 1947, Some Examples of Discrimination, Ann. Eugen. Lond., 13, 272.

-Spencer, Bennett & David Bowers, 1976, An Introduction to Multivariate Techniques for Social and Behavioural Sciences, The MacMillan Press Ltd.