

UNA APLICACIÓN DE DISCRIMINACION CUADRATICA EN LA AGRICULTURA

Doris Gómez Ticerán

Pilar Depaz Apestegui

RESUMEN.- *Se presenta una aplicación del método de discriminación cuadrática en la clasificación de 2 tipos de plantas: Mochero y Tambo Pateño en sus respectivas especies.*

1. INTRODUCCION

Uno de los tópicos más usados de la Estadística Multivariante se presenta con el nombre de Análisis Discriminante o Clasificación, cuyo objetivo principal es ubicar un individuo, objeto o grupos de objetos en una de las poblaciones concurrentes. Los criterios de discriminación también son utilizados para la reducción de la dimensión del problema original.

La clasificación consiste en la identificación del grupo al cual pertenece un nuevo individuo llevando en cuenta sus características observadas. Cuando esas características son mediciones numéricas, la designación a los grupos recibe el nombre de Discriminación y las combinaciones de esas mediciones se denomina Función Discriminante. Mas específicamente hablando, en la discriminación se trata de describir de manera gráfica o algebraica con el uso de funciones lineales o cuadráticas, llamadas funciones discriminantes, los aspectos que sirven para diferenciar o discriminar a los individuos u objetos de las poblaciones.

En problemas prácticos de discriminación y clasificación, los métodos más usados son el de Fisher (1936), de Razón de Verosimilitud (Mardia, 1976), y de Bayes (Anderson, 1984). Estos criterios proporcionan funciones discriminantes lineales para el caso homocedástico y funciones discriminantes cuadráticas para el caso heterocedástico, en base a las mediciones de las características del individuo a ser clasificado. En todos los casos, cuando el número de variables consideradas al construir la función discriminante es grande, no es posible la visualización gráfica de los datos y se requiere utilizar proyecciones en dos o tres dimensiones.

El método basado en estructuras de covarianzas diferentes (Gómez, 1997) no tiene esa limitación para problemas con más de dos variables, pues para cualquier dimensión mayor de 2, siempre proporciona dos combinaciones lineales de las componentes del individuo que se desea clasificar, las cuales pueden ser usadas en representaciones gráficas y en la reducción de la dimensión, toda vez que las dispersiones entre las dos poblaciones concurrentes sean diferentes.

En el contexto descrito, el objetivo del presente trabajo es presentar una aplicación del método de Discriminación Cuadrática, en el campo de la Agricultura.

2. ALGUNOS RESULTADOS PRELIMINARES

Supóngase que se desea clasificar un nuevo individuo de coordenadas $\bar{x} = (x_1, x_2, \dots, x_p)$ en una de dos poblaciones, π_1 o π_2 .

Las poblaciones π_1 y π_2 son normales p-variantes con vectores de medias $\bar{\mu}^{(g)}$ y matrices de covarianzas Σ_g , $g = 1, 2$, respectivamente. De dichas poblaciones se toman muestras aleatorias de tamaños n_1 y n_2 , obteniéndose los vectores de medias \bar{X}_1 y \bar{X}_2 y las matrices de covarianzas muestrales S_1 y S_2 .

Si las matrices de covarianzas de las dos poblaciones son diferentes, entonces, las discrepancias entre los dos grupos pueden ser analizadas a través de las combinaciones lineales de los vectores $\bar{X}^{(g)}$; $g = 1, 2$, esto es, de las combinaciones, $Y^{(g)} = \bar{a}' \bar{X}^{(g)}$ que maximizan o minimizan el cociente de varianzas,

$$\frac{\bar{a}' \Sigma_2 \bar{a}}{\bar{a}' \Sigma_1 \bar{a}}. \quad (1)$$

Los coeficientes $\bar{a} = (a_1, a_2, \dots, a_p)$ de estas combinaciones lineales son los autovectores asociados a los autovalores de la matriz $\Sigma_1^{-1} \Sigma_2$, denominados Componentes Principales (Flury, 1983).

De las p combinaciones lineales, se escogen aquellas que dan el mayor y menor cociente de varianzas o las combinaciones lineales que exacerban el cociente de varianzas en (1), lo que nos permite construir una transformación lineal de R^p en un espacio R^2 , donde pueden ser utilizados los métodos de Bayes o el de razón de verosimilitud para efectuar el análisis discriminante. A continuación se hace una descripción sucinta del método.

- Se calculan los p autovalores y los p autovectores de la matriz $\Sigma_1^{-1} \Sigma_2$ normalizados por $\bar{a}' \Sigma_1 \bar{a} = 1$, donde \bar{a}_i es el autovector asociado al autovalor λ_i de la matriz $\Sigma_1^{-1} \Sigma_2$.
- Se escogen los autovalores máximo (λ_{\max}) y mínimo (λ_{\min}), asociados a los autovectores de la matriz $\Sigma_1^{-1} \Sigma_2$. Los autovectores son tales que $\bar{a}'_{\max} \Sigma_g \bar{a}_{\min} = 0$ $g = 1, 2$.
- Con estos autovectores se definen las combinaciones lineales para cada una de las dos poblaciones π_1 y π_2 :

$$\begin{aligned} \pi_1 : Y_1^{(1)} &= \bar{a}'_{\max} \bar{X}^{(1)} & Y_p^{(1)} &= \bar{a}'_{\min} \bar{X}^{(1)} \\ \pi_2 : Y_1^{(2)} &= \bar{a}'_{\max} \bar{X}^{(2)} & Y_p^{(2)} &= \bar{a}'_{\min} \bar{X}^{(2)} \end{aligned}$$

cuyos parámetros son:

$$\text{En } \pi_1 : E \begin{pmatrix} Y_1^{(1)} \\ Y_p^{(1)} \end{pmatrix} = \begin{bmatrix} \bar{a}'_{\max} \bar{\mu}^{(1)} \\ \bar{a}'_{\min} \bar{\mu}^{(1)} \end{bmatrix}$$

$$\text{Var} \begin{pmatrix} Y_1^{(1)} \\ Y_p^{(1)} \end{pmatrix} = \begin{bmatrix} \bar{a}'_{\max} \sum_1 \bar{a}_{\max} & 0 \\ 0 & \bar{a}'_{\min} \sum_1 \bar{a}_{\min} \end{bmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$\text{En } \pi_2 : E \begin{pmatrix} Y_1^{(2)} \\ Y_p^{(2)} \end{pmatrix} = \begin{bmatrix} \bar{a}'_{\max} \bar{\mu}^{(2)} \\ \bar{a}'_{\min} \bar{\mu}^{(2)} \end{bmatrix}$$

$$\text{Var} \begin{pmatrix} Y_1^{(2)} \\ Y_p^{(2)} \end{pmatrix} = \begin{bmatrix} \bar{a}'_{\max} \sum_2 \bar{a}_{\max} & 0 \\ 0 & \bar{a}'_{\min} \sum_2 \bar{a}_{\min} \end{bmatrix} = \begin{pmatrix} \lambda_{\max} & 0 \\ 0 & \lambda_{\min} \end{pmatrix}$$

- d) Se define $\bar{Y}^{(g)} = A' \bar{X}^{(g)}$ $g = 1, 2$; donde A de orden $2 \times p$, es la matriz de la transformación de R^p en R^2 . Si $\bar{X}^{(g)}$ tiene distribución normal p-variante, entonces $\bar{Y}^{(g)}$ tiene distribución normal bivariante con vectores de medias: $\psi_g = A' \bar{\mu}^{(g)}$ y matriz de covarianzas $\phi_g = A' \Sigma_g A$; donde: $\phi_1 = I_{22}$ para $g = 1$ y $\phi_2 = \text{diag}(\lambda_{\max}, \lambda_{\min})$ para $g = 2$.
- e) Al transformar pasamos de dos poblaciones normales multivariantes a dos normales bivariantes con funciones de probabilidad denotadas por $f_1(\bar{y})$ y $f_2(\bar{y})$, respectivamente.
- f) Gómez (1997) demuestra que el vector \bar{x} es clasificado en π_1 cuando el vector $\bar{y} = \bar{a}' \bar{x}$ es clasificado en π_1 , es decir cuando:

$$\frac{f_1(\bar{y})}{f_2(\bar{y})} \geq \frac{q_2 C(1/2)}{q_1 C(2/1)}$$

donde q_1 y q_2 son las probabilidades a priori y $C(1/2)$ y $C(2/1)$ son los costos de mala clasificación.

- g) Finalmente, usando el supuesto de multinormalidad, se establece la siguiente regla de clasificación óptima :

Clasificar el individuo con vector de observaciones \bar{x} en π_1 si:

$$-\frac{1}{2} \bar{y}' (\phi_1^{-1} - \phi_2^{-1}) \bar{y} + \bar{y}' (\bar{\psi}_1' \phi_1^{-1} - \bar{\psi}_2' \phi_2^{-1}) \geq \ln \frac{q_2 C(1/2)}{q_1 C(2/1)} - \ln \frac{|\phi_2|}{|\phi_1|} + \frac{1}{2} (\bar{\psi}_1' \phi_1^{-1} \bar{\psi}_1 - \bar{\psi}_2' \phi_2^{-1} \bar{\psi}_2) \quad (2)$$

donde:

$\bar{\psi}_1$ y ψ_2 : valores de medios de π_1 y π_2 .

$\hat{\phi}_1$ y $\hat{\phi}_2$: matrices de covarianzas

q_1 y q_2 : son las probabilidades a priori de π_1 y π_2

$C(i/j)$: es el costo de clasificar equivocadamente un individuo de la población π_i en la población π_j .

En las aplicaciones prácticas los parámetros poblacionales son desconocidos y tienen que ser estimados en base a resultados muestrales, de modo que la ecuación (2) se expresa como:

$$-\frac{1}{2} \bar{Y}^T [\hat{\phi}_1^{-1} - \hat{\phi}_2^{-1}] \bar{Y} + \bar{Y}^T [\hat{\phi}_1^{-1} \bar{Y}_1 - \hat{\phi}_2^{-1} \bar{Y}_2] \geq \frac{1}{2} \ln |\hat{\phi}_2^{-1}| - \frac{1}{2} [\bar{Y}_1^T \hat{\phi}_1^{-1} \bar{Y}_1 - \bar{Y}_2^T \hat{\phi}_2^{-1} \bar{Y}_2]$$

donde $\hat{\psi}_1 = \bar{Y}_1$ y $\hat{\psi}_2 = \bar{Y}_2$ son vectores de medias muestrales y $\hat{\phi}_1$ y $\hat{\phi}_2$ son las matrices de covarianzas muestrales.

Para la aplicación se consideran costos de mala clasificación y probabilidades a priori iguales.

3. APLICACIÓN

Un problema que tiene mucha importancia en la actualidad es la escasez de los alimentos, por lo tanto, es medular la preservación de los recursos naturales genéticamente óptimos. Sin embargo, la preservación de dichos recursos requiere de una buena labor de clasificación de los mismos; razón por la cual el uso de una serie de metodologías estadísticas que permitan la clasificación y posterior identificación es necesaria en tales circunstancias.

En esta aplicación centramos nuestra atención en el campo de la agricultura donde se presenta con frecuencia el problema de clasificar los diferentes tipos de plantas en sus respectivas especies. En muchos casos, las diferencias pueden ser difícilmente detectables a simple vista, por lo tanto, el uso de métodos estadísticos tales como el Análisis Discriminante resultan de gran utilidad.

Los datos que utilizamos en la presente aplicación corresponden a un estudio realizado por investigadores de la Universidad Agraria La Molina (Laynes 1996), con dos variedades de maíz.

El primer grupo de observaciones corresponde a las características de 430 plantas de variedad Mochero en una localidad de Tumbes. En el segundo grupo se tiene 450 plantas de la variedad Tambo-Pateño, cultivados en una localidad de Piura. En ambos casos las plantas se distribuyeron en parcelas de 10 unidades cada una.

Se midieron las variables siguientes:

Altura de la planta (X_1): Promedio de 10 plantas competitivas medidas en centímetros desde el nivel del suelo, hasta el nudo donde nace la hoja superior.

Altura de la Inserción de la mazorca superior (X_2): Promedio de las 10 plantas, medidas en centímetros, de la altura de inserción de la mazorca superior tomada desde el nivel del suelo, hasta el nudo donde nace la mazorca superior.

Largo de la hoja de la mazorca (X_3): Promedio de las 10 plantas, medidas en centímetros del largo de la hoja de la mazorca, desde la ligula hasta la punta de la hoja.

Ancho de la hoja de la mazorca (X_4): Promedio de las 10 plantas, del ancho de la hoja de la mazorca, medida en centímetros en el punto medio de la distancia entre la ligula y la punta.

Diámetro del tallo (X_5): Promedio de las 10 plantas, medida en centímetros con la ayuda del calibrador del diámetro mayor tomado sobre el tallo principal en el punto medio del primer entrenudo emergente del suelo.

Se verificó (Depaz, 1999) que las muestras provenían de poblaciones normales multivariantes con matrices de covarianzas diferentes. (Ver Anexo 1)

Resultados :

Medias de las muestras de las razas Mochero y Tambo:

$$\bar{X}^{(1)} = \begin{bmatrix} 175.800 \\ 97.030 \\ 80.730 \\ 9.524 \\ 2.078 \end{bmatrix}_{(5 \times 1)} \quad \bar{X}^{(2)} = \begin{bmatrix} 196.770 \\ 106.190 \\ 79.858 \\ 10.093 \\ 2.307 \end{bmatrix}_{(5 \times 1)}$$

Matriz de covarianza de la 1ª muestra: Variedad Mochero

$$S_1 = \begin{bmatrix} 693.0105 & 504.8129 & 74.4507 & 13.9565 & 3.5584 \\ 504.8129 & 419.8410 & 55.9354 & 12.0594 & 2.9399 \\ 74.4507 & 55.9354 & 30.2355 & 2.2559 & 0.4373 \\ 13.9565 & 12.0594 & 2.2559 & 0.8051 & 0.1148 \\ 3.5584 & 2.9399 & 0.4373 & 0.1148 & 0.0332 \end{bmatrix}$$

Matriz de covarianza de la 2^o muestra: Variedad Tambo-pateño

$$S_2 = \begin{bmatrix} 183.1509 & 122.1484 & 6.6440 & 4.1513 & 0.9465 \\ 122.1484 & 178.2635 & 5.2590 & 4.8132 & 1.1705 \\ 6.6440 & 5.2590 & 23.4375 & 0.5887 & 0.2543 \\ 4.1513 & 4.8132 & 0.5887 & 0.4895 & 0.0649 \\ 0.9465 & 1.1705 & 0.2543 & 0.0649 & 0.0264 \end{bmatrix}$$

$$S_1^{-1} S_2 = \begin{bmatrix} 0.4942 & -1.0428 & -0.1003 & -0.0129 & -0.0065 \\ -0.3093 & -1.7717 & -0.0802 & 0.0031 & 0.0004 \\ -0.6623 & -0.4461 & 1.0459 & -0.0205 & 0.0046 \\ 2.7924 & -1.3099 & -1.8443 & 0.7244 & -0.0495 \\ 1.9974 & -0.5387 & 18.1418 & 0.8320 & 1.5641 \end{bmatrix}$$

Autovalores de la matriz $S_1^{-1} S_2$:

$$r_1 = 1.9658 \quad r_2 = 1.7405 \quad r_3 = 0.9621 \quad r_4 = 0.6881 \quad r_5 = 0.2438$$

Autovectores de la matriz $S_1^{-1} S_2$

$$\begin{bmatrix} a_1 & a_2 & a_3 & a_4 & a_5 \\ \mathbf{0.0984} & \mathbf{0.0033} & \mathbf{0.0009} & \mathbf{-0.0063} & \mathbf{0.0729} \\ \mathbf{-0.1487} & \mathbf{0.0026} & \mathbf{-0.0045} & \mathbf{-0.0010} & \mathbf{0.0192} \\ \mathbf{-0.0018} & \mathbf{-0.0138} & \mathbf{-0.0388} & \mathbf{0.0238} & \mathbf{0.0714} \\ \mathbf{0.3443} & \mathbf{0.0794} & \mathbf{0.1307} & \mathbf{0.4406} & \mathbf{-0.1978} \\ \mathbf{0.9218} & \mathbf{-0.9967} & \mathbf{0.9906} & \mathbf{-0.8973} & \mathbf{-0.9747} \end{bmatrix}$$

Combinaciones lineales que maximizan y minimizan el cociente de varianzas

$$\begin{aligned} Y_{\max}^{(g)} &= 0.0984x_1^{(g)} - 0.1487x_2^{(g)} - 0.0018x_3^{(g)} + 0.3443x_4^{(g)} + 0.9218x_5^{(g)} ; \\ Y_{\min}^{(g)} &= 0.0729x_1^{(g)} + 0.0192x_2^{(g)} + 0.0714x_3^{(g)} - 0.1978x_4^{(g)} - 0.9747x_5^{(g)} ; \end{aligned} \quad (3)$$

Estimaciones de los Vectores de medias ψ_1 y ψ_2 .

$$\bar{Y}^{(1)} = \begin{bmatrix} 8.0130 \\ 15.348 \end{bmatrix} \quad \bar{Y}^{(2)} = \begin{bmatrix} 9.0300 \\ 17.840 \end{bmatrix}$$

Matriz de covarianzas estimadas para los datos transformados en los grupos 1 y 2

$$\hat{\phi}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \hat{\phi}_2 = \begin{bmatrix} 1.9658 & 0 \\ 0 & 0.2438 \end{bmatrix}$$

$$\hat{\phi}_1^{-1} - \hat{\phi}_2^{-1} = \begin{bmatrix} 0.4913 & 0 \\ 0 & -3.10172 \end{bmatrix}$$

$$\hat{\phi}_1^{-1} \bar{Y}_1 - \hat{\phi}_2^{-1} \bar{Y}_2 = \begin{bmatrix} 3.32650 \\ -56.640 \end{bmatrix}$$

$$\text{Ln}|\hat{\phi}_2^{-1}| = -0.7355$$

$$\bar{Y}_1^T \hat{\phi}_1^{-1} \bar{Y}_1 - \bar{Y}_2^T \hat{\phi}_2^{-1} \bar{Y}_2 = -999.20$$

$$D(\bar{y}) = -\frac{1}{2} \bar{Y}^T [\hat{\phi}_1^{-1} - \hat{\phi}_2^{-1}] \bar{Y} + \bar{Y}^T [\hat{\phi}_1^{-1} \bar{Y}_1 - \hat{\phi}_2^{-1} \bar{Y}_2] + d$$

donde : $d = \frac{1}{2} \text{Ln}|\hat{\phi}_2^{-1}| - \frac{1}{2} [\bar{Y}_1^T \hat{\phi}_1^{-1} \bar{Y}_1 - \bar{Y}_2^T \hat{\phi}_2^{-1} \bar{Y}_2]$ $\bar{Y} = \begin{bmatrix} Y_{\max} \\ Y_{\min} \end{bmatrix}$

Finalmente, la función discriminante cuadrática será:

$$D(\bar{Y}) = -0.24565Y_{\max}^2 + 1.55086Y_{\min}^2 + 3.3265Y_{\max} - 56.6407Y_{\min} + 499.60 \quad (4)$$

Esta metodología reduce de manera significativa el manejo de las variables originales en la función discriminante sin dejar de lado la importancia de ellas puesto que en el cálculo de las variables Y_{\max} , Y_{\min} están involucradas todas las variables originales.

Suponiendo costos y probabilidades a priori iguales (Anderson, 1984) y partiendo de la ecuación (4) se establece la siguiente regla de clasificación:

Si $D(\bar{Y}) \geq 0$ entonces \bar{x} se clasifica en π_1 caso contrario, \bar{x} se clasifica en π_2

La regla de clasificación propuesta es :

$$\text{Si } -0.24565Y_{\max}^2 + 1.55086Y_{\min}^2 + 3.3265Y_{\max} - 56.6407Y_{\min} + 499.60 \geq 0 \Rightarrow$$

\bar{x} será clasificado como perteneciente a la variedad mochero.

$$\text{Si } -0.24565Y_{\max}^2 + 1.55086Y_{\min}^2 + 3.3265Y_{\max} - 56.6407Y_{\min} + 499.60 < 0 \Rightarrow$$

\bar{x} será clasificado como perteneciente a la variedad tambo pateño. (5)

Cabe señalar que, toda vez que se hace un trabajo de discriminación será necesario evaluar la función discriminante. En ese sentido el método mas usado y el más sencillo consiste en estimar la probabilidad total de mala clasificación.

La estimación del costo de mala clasificación se resume en el siguiente cuadro :

**TABLA N° 1. CLASIFICACIÓN DE VARIEDADES DE MAÍZ.
MÉTODO DE DISCRIMINACIÓN CUADRÁTICA.**

GRUPO OBSERVADO	CLASIFICADO EN		TOTAL
	π_1	π_2	
π_1	40	3	43
π_2	4	41	45
TOTAL	44	44	88

Calculando la estimación de la probabilidad de mala clasificación.

$$PTM = \frac{3+4}{43+45} = \frac{7}{88} = 0.0795$$

La tasa de mala clasificación de la regla planteada es igual a 0.0795, esto significa que el 7.95% de los tipos de maíz han sido mal clasificados.

Por tanto, la función discriminante obtenida discrimina bien los dos grupos, lo que nos permite usarla para clasificar nuevas plantas en una de las dos especies consideradas.

Se consideró conveniente evaluar los datos usando el software MINITAB, que trabaja la discriminación cuadrática, con la metodología de Anderson (1958, 1984). La probabilidad total de mala clasificación se resume en el siguiente cuadro:

**TABLA N° 2. CLASIFICACIÓN DE VARIEDADES DE MAÍZ.
REPORTE DE MINITAB.**

GRUPO OBSERVADO	CLASIFICADO EN		TOTAL
	π_1	π_2	
π_1	35	8	43
π_2	7	38	45
TOTAL	42	46	88

$$PTM = \frac{7+8}{43+45} = \frac{15}{88} = 0.170$$

Esto significa que el 17% de las plantas de maíz han sido clasificados incorrectamente.

Recordemos que la probabilidad de mala clasificación con la metodología usada fue de 7.95%, entonces podemos decir que en la presente aplicación la regla de Discriminación con la metodología planteada, discrimina mejor que la propuesta de Anderson (1958), siempre que sus matrices de covarianzas sean diferentes.

CONCLUSIONES

Para aplicar la metodología de Discriminación Cuadrática los datos tienen que ser previamente transformados de R^p a R^2 . La función Discriminante usa sólo la información de los datos transformados en R^2 .

De la comparación se tiene que la probabilidad de mala clasificación, usando la Discriminación Cuadrática resulta menor que la que se obtiene con el método clásico de Anderson (1984) en el que trabaja con los datos originales en la respectiva dimensión.

BIBLIOGRAFIA

- [1] Anderson, T. W. *An Introduction to Multivariate Statistical Analysis*. Wiley & Sons, New York, (1984).
- [2] FISHER. R. A. (1936). *The Use of Multiple Measurements in Taxonomic Problems*. Ann. Eugenics, 17, 179 - 188.
- [3] FLURY, B. *Some relations between the comparison of covariance matrices principal component analysis*. Computational Statistics & Data Analysis 1 : 97 - 109 (1983).
- [4] GÓMEZ, T. D. *Componentes Principales Generalizados Vía comparación de matrices de covarianzas*. Tesis de pre-grado (1990).
- [5] GÓMEZ, T. D. *Discriminación Cuadrática mediante matrices de covarianza*. PESQUIMAT Vol. I N° 2 p.p. 111-116 (1988).
- [6] LAINES, B. L. *Componentes Principales en K Grupos*. Tesis de Post-grado (1996).
- [7] MARDIA, K. V. *Multivariate Analysis*. Academic Press (1976).
- [8] MORRISON, *Multivariate Statistical Methods*. Mc Graw Hill, New York, (1976).

ANEXO 1

PRUEBA DE MULTINORMALIDAD (Mardia, 1970)

Muestra 1 : Promedio de 10 plantas de maíz de la raza Mochero, tomadas en cada una de las 43 parcelas.

H_0 : La distribución poblacional no tiene sesgo y es normal multivariante.

H_1 : La distribución poblacional es sesgada y no normal multivariante.

$$H_0 : B_{1p} = 0 ; \quad B_{2p} = p(p+2)$$

$$n_1 = 43 \quad p = 5$$

$$b_{1p} = 5.62998 \quad b_{2p} = 34.3163$$

$$A = \frac{1}{6} n_1 b_{1p} = 40.3482 \quad B = \frac{b_{2p} - p(p+2)}{\sqrt{8p(p+2)/n_1}} = -0.26793$$

para $\alpha = 5\%$ los valores teóricos son :

$$\chi^2_{\alpha, [p(p+3)(p+2)/6]} = 49.802 \quad Z_{\alpha} = 1.96$$

Luego como :

$$\frac{1}{6} n_1 b_{1p} = 40.3482 < \chi^2_{\alpha, [p(p+3)(p+2)/6]} = 49.802 \quad y$$

$$Z = \frac{b_{2p} - p(p+2)}{\sqrt{8p(p+2)/n_1}} = -0.26793 > -1.96 = Z_{0.05}$$

No se rechaza la hipótesis de inesgamiento y de multinormalidad de la muestra 1, ($\alpha = 0.05$).

Muestra 2 : Promedio de 10 plantas de maíz de la raza Tambo pateño tomadas en cada una de las 45 parcelas.

Planteamiento de la Hipótesis : (Muestra 2)

H_0 : La distribución poblacional no tiene sesgo y es normal multivariante.

H_1 : La distribución poblacional es sesgada y no normal multivariante.

$$n_2 = 45 \quad p = 5$$

$$b_{1p} = 4.9462 \quad b_{2p} = 31.40$$

$$A = \frac{1}{6} n_2 b_{1p} = 37.09 \quad B = \frac{b_{2p} - p(p+2)}{\sqrt{8p(p+2)/n_2}} = -1.44$$

para $\alpha = 5\%$ los valores teóricos son :

$$\chi^2_{\alpha, [p(p+3)(p+2)/6]} = 49.802 \quad Z_{\alpha} = 1.96$$

Luego como :

$$\frac{1}{6} n_2 b_{1p} = 37.09 < \chi^2_{\alpha, [p(p+3)(p+2)/6]} = 49.802 \quad y$$

$$Z = \frac{b_{2p} - p(p+2)}{\sqrt{8p(p+2)/n_2}} = -1.44 > 1.96 = Z_{0.05}$$

No se rechaza la hipótesis de insesgamiento y de multinormalidad de la muestra 2, ($\alpha = 0.05$).