# LAS FUNCIONES DISCRIMINANTES EN LA INVESTIGACIÓN PSICOBIOMÉTRICA

Felipe Montemayor
y
María Teresa Jaén

En el campo de la biología y de la psicología el investigador se halla frecuentemente ante la necesidad de colocar a un individuo o a una pequeña muestra de ellos dentro del grupo a que debe pertenecer por tales o cuales características. En el primer caso puede tratarse de tipos genéticos o taxonómicos y en el segundo, de asignar a un sujeto determinado, según sus caracteres psicológicos, a tal o cual grupo profesional o escolar.

La solución a este problema la dio Fisher en 1936,[1] e inmediatamente fue aplicada en el campo de la antropología física por E. S. Martin quien usó las funciones discriminantes en la diferenciación sexual osteométrica de la mandíbula y por Mildred Bernard que en una serie de cráneos fechados pudo determinar el componente genético secular del grupo en cuestión.[2]

El problema de saber si un espécimen determinado pertenece, según características cuantificables dadas, a cierta población previamente estudiada, se puede abordar en términos probabilísticos, a base de las distribuciones de cada una de las características consideradas y viendo el lugar que ocupan dentro de ellas las del sujeto de que se trata. Sin embargo, este tipo de problemas, por lo menos en lo que a ciertos trabajos de antropología física se refiere, son poco tratados. Un buen número de investigaciones se reducen a la estimación de los parámetros de las distribuciones, generalmente normales, de los caracteres objeto de estudio y a la comparación de sus promedios y porcentajes, para obtener criterios de más o de menos con respecto a otros grupos.

En los centros de investigación antropológica se dispone de colecciones de cráneos de diferentes grupos étnicos las cuales por lo general están bien estudiadas.

[1] Fisher, R. A. The Use of Multiple Measusements in Taxonomic Problems. *Annals of Eugenics*, 7, pp. 179-88. England, 1936.

[2] *Ib.*, p. 179.

Pero si en un momento dado se quisiera saber a qué grupo pertenece un nuevo ejemplar cuya procedencia no se revelara, se tendría que andar a tanteos o recurrir al método citado en el párrafo anterior, es decir, a determinar según las características métricas del cráneo, las probabilidades que tiene de pertenecer a cualquiera de las colecciones ya conocidas.

Las funciones discriminantes facilitan este trabajo reduciéndolo a la simple substitución de los datos métricos del ejemplar aislado, en las ecuaciones establecidas para los grupos y, según el valor obtenido en esas, clasificarlo correctamente en la medida de lo posible.

Ahora bien, no es un diámetro ni un índice osteométrico ni la calificación en una prueba psicológica de un individuo, lo que sirve para diagnosticar, sino el conjunto de varias de ellas. El problema es determinar en qué forma e intensidad están combinadas estas medidas para obtener una visión integral que permita la colocación del individuo dentro de un grupo.[3]

El objeto de estas líneas es el de proporcionar a los investigadores, especialmente a los jóvenes, una explicación más o menos detallada y lo más sencilla posible de los procesos numéricos que hay que seguir para el establecimiento de las funciones discriminantes, evitando en lo posible las consideraciones teóricas para las cuales se dan las referencias necesarias. Al mismo tiempo se previene, que no deben generalizarse los resultados obtenidos en el ejemplo a base de datos psicológicos por lo reducido del número de casos empleados y por otros inconvenientes.

## *LAS FUNCIONES DISCRIMINANTES*

El problema consiste en determinar las funciones lineales que ponen en relación un número dado de variables en dos o más poblaciones. En ocasiones una sola variable es suficiente, pero lo más común es que los grupos biológicos o psicológicos que se trata de discriminar difieran por más de un carácter.

Puede decirse que existen dos métodos para el cálculo de dichas funciones, uno que podría llamarse directo y otro indirecto. Este último se aplica cuando se quiere conocer la intensidad y el sentido de la correlación que existe entre las variables consideradas. Los resultados finales de ambos procedimientos, desde luego, son idénticos.

La comprobación final del poder discriminante de las funciones obtenidas se consigue por medio del Análisis de la Variancia, para determinar si la variabilidad entre los dos grupos considerados es significativamente mayor que la que hay dentro de los grupos.

Supongamos la existencia de dos grupos, *a* y *b* de los cuales hemos elegido tres características: X, Y y Z = k.

[3] Moroney, M. J., *Facts from figures.* London, 1955.

Grupo *a*

$$\begin{array}{lllll} Xa_1, & Xa_2, & Xa_3, & \ldots\ldots\ldots\ldots & Xan \\ Ya_1, & Ya_2, & Ya_3, & \ldots\ldots\ldots\ldots & Yan \\ Za_1, & Za_2, & Za_3, & \ldots\ldots\ldots\ldots & Zan \end{array} \quad *$$

Grupo *b*

$$\begin{array}{lllll} Xa_1, & Xa_2, & Xa_3, & \ldots\ldots\ldots\ldots & Xbn \\ Ya_1, & Ya_2, & Ya_3, & \ldots\ldots\ldots\ldots & Ybn \\ Za_1, & Za_2, & Za_3, & \ldots\ldots\ldots\ldots & Zbn \end{array}$$

El primer paso es determinar las diferencias entre las medias de las variables del grupo *a* con las del grupo *b* en la siguiente forma:

$$\begin{array}{l} M_{xa} - M_{xb} = dx \\ M_{ya} - M_{yb} = dy \\ M_{za} - M_{ab} = dz \end{array}$$

Después hay que calcular:

1) La suma de cuadrados o sea la de las desviaciones de cada dato con respecto a su media aritmética:

Para el caso de X se tiene:

$$\text{Suma de cuadrados} = \Sigma\, x^2 = (\Sigma\, X^2a + \Sigma\, X^2b) - \left[\frac{(\Sigma\, Xa)^2}{na}\right] + \left[\frac{(\Sigma\, Xb)^2}{nb}\right]$$

La expresión del segundo término del segundo miembro de la ecuación anterior se conoce como factor de corrección (F. C.).

De manera semejante se calculan las sumas de cuadrados para Y y para Z.

La suma de productos es:

$$\Sigma\, xy = \Sigma\, (Xa\, Ya) + \Sigma\, (XbYb) - \left[\frac{(\Sigma\, Xa)\quad(\Sigma\, Ya)}{na} + \frac{(\Sigma\, Xb)\quad(\Sigma\, Yb)}{nb}\right]$$

Aquí también el término entre paréntesis rectangular se designa por factor de corrección (F. C.).

De modo semejante se calculan las sumas de cuadrados para las combinaciones restantes o sea: $\Sigma xz$ y $\Sigma yz$.

Con estos datos se está en posibilidad de establecer un sistema de ecuaciones con tres incógnitas de la forma:

$$\begin{array}{l} l_1\ (\Sigma\, x^2) + l_2\ (\Sigma\, xy) + l_3\ (\Sigma\, xz) = d_x \\ l_1\ (\Sigma\, xy) + l_2\ (\Sigma\, y^2) + l_3\ (\Sigma\, yz) = d_y \\ l_1\ (\Sigma\, xz) + l_2\ (\Sigma\, yz) + l_3\ (\Sigma\, z^2) = d_z \end{array}$$

* Por dificultades tipográficas se designará por M la media aritmética en lugar de la literal testada usual.

Siendo las incógnitas los coeficientes o ponderaciones de cada término $l_1$, $l_2$ y $l_3$.

Una vez obtenido el valor de los coeficientes se pueden establecer las ecuaciones típicas para cada grupo, substituyendo las medias de sus variables de la manera siguiente:

$$\text{Grupo a: } D_a = l_1(Mx_a) + l_2(My_a) + l_3(Mz_a)$$
$$\text{Grupo b: } D_b = l_1(Mx_b) + l_2(My_b) + l_3(Mz_b)$$

El poder discriminante de la función se obtiene substituyendo en la ecuación las diferencias entre las medias de cada variable:

$$D = l_1(d_x) + l_2(d_y) + l_3(d_z)$$

Que representa la suma de cuadrados dentro de grupos con $n_a + n_b - k - 1$ grados de libertad.

Para proceder al Análisis de la Variancia se obtiene el valor de $D^2$ que multiplicado por $\frac{n_a n_b}{n_a + n_b}$ proporciona la suma de cuadrados entre grupos con k (número de caracteres) grados de libertad.[4]

## *EJEMPLOS*

A continuación se van a dar tres ejemplos, uno de ellos con todos los pasos del cálculo de las funciones discriminantes. Dos de los ejemplos aplicados a la antropología física y uno a la psicología. Para:

1) Discriminar sexos a base de tres diámetros del cráneo.

2) Discriminar dos grupos étnicos a base de cuatro diámetros del cráneo y,

3) Discriminar sexos a base de 3 valores psicológicos.

## *EL MATERIAL OSTEOLÓGICO*

Para el primer ejemplo se seleccionaron al azar 25 cráneos masculinos de una colección de 45 procedentes de la Cueva de la Candelaria, Coah., y 23 femeninos de una colección de 36 de la misma procedencia.

La razón para no tomar las series completas, además de la de facilitar los cálculos, fue la de dejar material no incluido en la elaboración para poder observar si el diagnóstico sexual dado por los antropólogos por los medios tradicionales, quedaba confirmado por las estimaciones a base de las funciones.

[4] Goulden, C. H., *Methods of Statistical Analysis*. New York, 1936.

El material de la Candelaria ha sido estudiado por Romano [5] y a continuación se dan las estimaciones paramétricas obtenidas por dicho autor sobre las colecciones completas (mesurables) y las obtenidas en muestras de 25 cráneos masculinos y 23 femeninos. Las medidas tomadas en cuenta son: Diámetro Antero-posterior Máximo, Diámetro Transverso Máximo y Diámetro Bizigomático.

CRÁNEOS MASCULINOS

| | Romano | Montemayor-Jaén |
|---|---|---|
| *Diámetro antero-posterior máximo* | | |
| n | 43 | 25 |
| M | 182.48 ± 0.45 | 182.64 ± 0.92 |
| s | 4.24 ± 0.45 | 4.60 ± 0.65 |
| *Diámetro transverso máximo* | | |
| n | 42 | 25 |
| M | 132.94 ± 0.63 | 132.68 ± 0.85 |
| s | 4.10 ± 0.44 | 4.28 ± 0.60 |
| *Diámetro bizigomático* | | |
| n | 32 | 25 |
| M | 137.42 ± 0.62 | 136.96 ± 0.82 |
| s | 3.81 ± 0.47 | 4.10 ± 0.57 |

CRÁNEOS FEMENINOS

| | | |
|---|---|---|
| *Diámetro antero-posterior máximo* | | |
| n | 36 | 23 |
| M | 176.92 ± 0.87 | 175.69 ± 1.07 |
| s | 5.20 ± 0.60 | 5.25 ± 0.55 |
| *Diámetro transverso máximo* | | |
| n | 33 | 23 |
| M | 130.00 ± 0.71 | 128.56 ± 0.89 |
| s | 4.16 ± 0.50 | 4.28 ± 0.63 |
| *Diámetro bizigomático* | | |
| n | 27 | 23 |
| M | 128.22 ± 0.75 | 127.95 ± 0.84 |
| s | 3.96 ± 0.53 | 4.06 ± 0.60 |

[5] Romano, A. Los Restos Óseos Humanos de la Cueva de la Candelaria, Coah., Tesis para la Escuela Nacional de Antropología e Historia, 1956 (inédita).

Como se observa, a excepción del Diámetro Transverso en las mujeres, en todas las medidas hay una coincidencia entre las estimaciones de Romano y las establecidas en una muestra de 23 ejemplares. Además, la diferencia que existe entre las estimaciones del diámetro transverso mencionado no es significativa y se debe al azar.[1]

A continuación se presentan los datos brutos y codificados de los 25 cráneos masculinos y de los 23 femeninos de la Cueva de la Candelaria, con las siguientes variables y su respectiva codificación*:

| | |
|---|---|
| Diámetro antero-posterior máximo | −170 mm. = X |
| Diámetro transverso máximo | −130 mm. = Y |
| Diámetro bizigomático | −130 mm. = Z |

CUEVA DE LA CANDELARIA

HOMBRES (Grupo I, con $n_I = 25$)

| *Núm.* | $X$ | $Y$ | $Z$ | $X^2$ | $Y^2$ | $Z^2$ | $XY$ | $XZ$ | $YZ$ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 3 | 7 | 5 | 9 | 49 | 25 | 21 | 15 | 35 |
| 3 | 13 | 4 | 6 | 169 | 16 | 36 | 52 | 78 | 24 |
| 4 | 11 | 0 | 3 | 121 | 0 | 9 | 0 | 33 | 0 |
| 5 | 19 | 0 | 2 | 361 | 0 | 4 | 0 | 38 | 0 |
| 8 | 16 | 5 | 10 | 256 | 25 | 100 | 80 | 160 | 50 |
| 10 | 13 | 5 | 10 | 169 | 25 | 100 | 65 | 130 | 50 |
| 12 | 23 | 5 | 4 | 529 | 25 | 16 | 115 | 92 | 20 |
| 13 | 16 | 12 | 12 | 256 | 144 | 144 | 192 | 192 | 144 |
| 17 | 6 | 0 | 6 | 36 | 0 | 36 | 0 | 36 | 0 |
| 18 | 14 | −1 | 4 | 196 | 1 | 16 | −14 | 56 | −4 |
| 19 | 12 | −3 | 2 | 144 | 9 | 4 | −36 | 24 | −6 |
| 22 | 16 | 0 | 10 | 256 | 0 | 100 | 0 | 160 | 0 |
| 23 | 16 | 12 | 11 | 256 | 144 | 121 | 192 | 176 | 132 |
| 24 | 14 | 8 | 10 | 196 | 64 | 100 | 112 | 140 | 80 |
| 25 | 5 | −1 | 2 | 25 | 1 | 4 | −5 | 10 | −2 |
| 27 | 13 | 2 | 3 | 169 | 4 | 9 | 26 | 39 | 6 |
| 28 | 13 | −2 | 4 | 169 | 4 | 16 | −26 | 52 | −8 |
| 30 | 15 | 1 | 5 | 225 | 1 | 25 | 15 | 75 | 5 |
| 31 | 8 | 6 | 10 | 64 | 36 | 100 | 48 | 80 | 60 |
| 33 | 11 | −3 | 15 | 121 | 9 | 225 | −33 | 165 | −45 |
| 34 | 16 | 7 | 10 | 256 | 49 | 100 | 112 | 160 | 70 |
| 36 | 14 | −1 | 14 | 196 | 1 | 196 | −14 | 196 | −14 |
| 37 | 8 | −1 | 0 | 64 | 1 | 0 | −8 | 0 | 0 |
| 40 | 15 | 3 | 9 | 225 | 9 | 81 | 45 | 135 | 27 |
| 44 | 6 | 2 | 7 | 36 | 4 | 49 | 12 | 42 | 14 |
| Σ | 316 | 67 | 174 | 4 504 | 621 | 1 616 | 951 | 2 284 | 638 |

[1] $$t = \frac{130.00 - 128.56}{\sqrt{\dfrac{(4.16)^2}{33} + \dfrac{(4.28)^2}{23}}} = \frac{1.44}{\sqrt{1.32}} = \frac{1.44}{1.15} = 1.25 \qquad t(0.05\%) = 2.01 \quad t(0.01\%) = 2.68$$

* La codificación consiste en restar una cantidad constante a los datos originales para trabajar con números más chicos. Esto, como se sabe, no altera la variancia.

CUEVA DE LA CANDELARIA

MUJERES (Grupo II, con $n_{II} = 23$)

| *Núm.* | *X* | *Y* | *Z* | $X^2$ | $Y^2$ | $Z^2$ | *XY* | *XZ* | *YZ* |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 4 | 12 | 15 | 16 | 144 | 225 | 48 | 60 | 180 |
| 2 | 14 | 16 | 14 | 196 | 256 | 196 | 224 | 196 | 224 |
| 3 | 11 | 8 | 14 | 121 | 64 | 196 | 88 | 154 | 112 |
| 7 | 12 | 2 | 4 | 144 | 4 | 16 | 24 | 48 | 8 |
| 9 | 1 | 8 | 12 | 1 | 64 | 144 | 8 | 12 | 96 |
| 10 | 9 | 10 | 10 | 81 | 100 | 100 | 90 | 90 | 100 |
| 13 | 6 | 6 | 7 | 36 | 36 | 49 | 36 | 42 | 42 |
| 14 | 10 | 13 | 5 | 100 | 169 | 25 | 130 | 50 | 65 |
| 17 | 6 | 4 | 8 | 36 | 16 | 64 | 24 | 48 | 32 |
| 18 | 6 | 9 | 9 | 36 | 81 | 81 | 54 | 54 | 81 |
| 20 | 9 | 13 | 10 | 81 | 169 | 100 | 117 | 90 | 130 |
| 22 | 6 | 6 | 8 | 36 | 36 | 64 | 36 | 48 | 48 |
| 23 | −1 | 7 | 6 | 1 | 49 | 36 | −7 | −6 | 42 |
| 24 | 4 | 4 | 0 | 16 | 16 | 0 | 16 | 0 | 0 |
| 25 | 11 | 16 | 8 | 121 | 256 | 64 | 176 | 88 | 128 |
| 26 | 1 | 8 | 9 | 1 | 64 | 81 | 8 | 9 | 72 |
| 27 | 2 | 9 | 4 | 4 | 81 | 16 | 18 | 8 | 36 |
| 29 | 4 | 7 | 8 | 16 | 49 | 64 | 28 | 32 | 56 |
| 30 | 16 | 5 | 2 | 256 | 25 | 4 | 80 | 32 | 10 |
| 33 | −3 | 9 | 8 | 9 | 81 | 64 | −27 | −24 | 72 |
| 34 | 2 | 12 | 7 | 4 | 144 | 49 | 24 | 14 | 84 |
| 35 | −2 | 5 | 3 | 4 | 25 | 9 | −10 | −6 | 15 |
| 37 | 3 | 8 | 12 | 9 | 64 | 144 | 24 | 36 | 96 |
| $\Sigma$ | 131 | 197 | 183 | 1 325 | 1 993 | 1 791 | 1 209 | 1 075 | 1 729 |

Los datos necesarios para ambos métodos son:

$$\Sigma\ X^2 = 4504 + 1325 = 5829$$

$$\text{F.C.} = \left[\frac{(316)^2}{25} \quad \frac{(131)^2}{23}\right] = \frac{99856}{25} + \frac{17161}{23} = 3994.24 + 746.13 = 4740.37$$

$$\Sigma\ x^2 = 5829 - 4740.37 = 1088.63$$
$$\sqrt{x^2} = \sqrt{1088.63} = 33.00$$

$$\Sigma\ XY = 951 + 1209 = 2160.00$$

$$\text{F.C.} = \left[\frac{(316)\ (67)}{25} + \frac{(131)\ (197)}{23}\right]$$

$$= \frac{21172}{25} + \frac{25807}{23} = 846.88 + 1122.04 = 1968.92$$

$$\Sigma\ (xy) = 2160.00 - 1968.92 = 191.08$$

$$\Sigma\ Y^2 = 621 + 1993 = 2614$$

$$F.C. = \left[\frac{(67)^2}{25} + \frac{(197)^2}{23}\right] = \frac{4489}{25} + \frac{38809}{23} = 179.56 + 1687.34 = 1866.90$$

$$\Sigma\ y^2 = 2614 - 1866.90 = 747.10$$

$$\sqrt{y^2} = \sqrt{747.10} = 27.34$$

$$\Sigma\ XZ = 2284 + 1075 = 3359$$

$$F.C. = \left[\frac{(316)\ (174)}{25} + \frac{(131)\ (183)}{23}\right]$$

$$= \frac{54984}{25} + \frac{23973}{23} = 2199.36 + 1042.30 = 3241.66$$

$$\Sigma\ (xz) = 3359 - 3241.66 = 117.34$$

$$\Sigma\ Z^2 = 1616 + 1791 = 3407.00$$

$$F.C. = \left[\frac{(174)^2}{25} + \frac{(183)^2}{23}\ \frac{30276}{25} + \frac{33489}{23}\right] = 1211.04 + 1456.04 = 2667.08$$

$$\Sigma\ z^2 = 3407.00 - 2667.08 = 739.92$$

$$\sqrt{z^2} = \sqrt{739.92} = 28.09$$

$$\Sigma\ YZ = 638 + 1729 = 2367.00$$

$$F.C. = \left[\frac{(67)\ (174)}{25} + \frac{(197)\ (183)}{23}\right]$$

$$= \frac{11658}{25} + \frac{36051}{23} = 466.32 + 1567.43 = 2033.75$$

$$\Sigma\ (yz) = 2367 - 2033.75 = 333.25$$

Los datos necesarios para el cálculo de las ecuaciones son:

$$d_x = M_xI - M_xII = (12.64) - (5.69) = 6.95$$
$$d_z = M_yI - M_yII = (2.68) - (-1.44) = 4.12$$
$$d_y = M_zI - M_zII = (6.96) - (-2.05) = 9.01$$

Si designan por *a, b* y *c,* los coeficientes del sistema de ecuaciones. La resultante es:

$$1088.59a + 191.08b + 117.34c = 6.95$$
$$191.08a + 747.10b + 333.25c = 4.12$$
$$117.34a + 333.25b + 739.92c = 9.01$$

Utilizando cualquiera de los procedimientos algebraicos conocidos para su resolución, se obtiene:

$$a = 0.005283$$
$$b = -0.000975$$
$$c = 0.0118$$

Que substituidos en la función discriminante dan:

$$D = 0.005283X - 0.000975Y + 0.0118Z$$

La suma de cuadrados dentro de grupos para el Análisis de la Variancia se obtiene substituyendo en la ecuación anterior $d_x$, $d_y$ y $d_z$ en sus respectivos lugares, o sea:

$$D = 0.005283(6.95) - 0.000975(4.12) + 0.0118(9.01) = 0.1390$$

La suma de cuadrados entre grupos se obtiene por:

$$\frac{n_I \quad n_{II}}{n_I + n_{II}} \qquad D^2 = \frac{(25)\ (23)}{25 + 23}(0.1390)^2 = \frac{575}{48}(0.0193) = 0.2312$$

ANÁLISIS DE LA VARIANCIA

| *Fuente* | *Grados de libertad* | *Suma de cuadrados* | *Cuadrado medio* | *F* | *F(5%)* | *F(1%)* |
|---|---|---|---|---|---|---|
| Entre grupos | 3 | 0.2312 | 0.0770 | 24.83 | 2.76 | 6.17 |
| Dentro de grupos | 44 | 0.1390 | 0.0031 | | | |

Los tres grados de libertad entre grupos corresponden a las variables consideradas $X + Y + Z = k = 3$ y los 44 dentro de grupos a:

$$n_I + n_I - k - 1 = 25 + 23 - 3 - 1 = 48 - 4 = 44$$

El poder discriminante de la función obtenida es altamente significativo, pues los valores de F debidos al azar al 5% y 1% son respectivamente: 2.76 y 6.17 para 3 y 44 grados de libertad.

El procedimiento que consideramos indirecto para el cálculo de las funciones discriminantes, se logra por medio de una matriz inversa de coeficientes de correlación.

Los coeficientes de correlación lineal entre las variables tomadas en cuenta son:

$$r_{xy} = \frac{191.08}{\sqrt{(1088.59)\ (747.10)}} = \frac{191.08}{901.81} = 0.2120 \pm 0.141$$

$$r_{xz} = \frac{117.34}{\sqrt{(1088.59)\ (739.92)}} = \frac{117.34}{\phantom{000}} = 0.1315 \pm 0.144$$

$$r_{yz} = \frac{333.25}{\sqrt{(747.10)\ (739.92)}} = \frac{333.25}{743.50} = 0.4482 \pm 0.118$$

Con ellos se está en posibilidad de calcular la siguiente matriz inversa:[1]

MATRIZ INVERSA DE COEFICIENTES DE CORRELACIÓN

| | | | | | |
|---|---|---|---|---|---|
| 1.0 | 0.2120 | 0.1315 | 1.0 | 0 | 0 |
| 0.2120 | 1.0 | 0.4482 | 0 | 1.0 | 0 |
| 0.1315 | 0.4482 | 1.0 | 0 | 0 | 1.0 |
| | | | | | |
| 1.0 | 0.2120 | 0.1315 | 1.0 | 0 | 0 |
| 1.0 | 0.2120 | 0.1315 | 1.0 | 0 | 0 |
| | | | | | |
| | 0.9551 | 0.4204 | −0.2120 | 1.0 | 0 |
| | 1.0 | 0.4401 | −0.2219 | 1.0470 | 0 |
| | | | | | |
| | | 0.7977 | −0.0382 | −0.4401 | 1.0 |
| | | 1.0 | −0.0479 | −0.5517 | 1.2536 |
| | | | | | |
| | | | 1.0467 | −0.2008 | −0.0479 |
| | | | −0.2008 | 1.2898 | −0.5517 |
| | | | −0.0479 | −0.5517 | 1.2536 |

Prueba:

$$1.0\ (1.0467) - (0.212)(0.2008) - (0.1315)(0.0479) = 1.0000$$
$$-(0.212)(0.2008) + 1.0(1.2898) - (0.4482)(0.5517) = 1.0000$$
$$-(0.1315)(0.0479) - (0.4482)(0.5517) + 1.0\ (1.2536) = 1.0000$$

En este caso el procedimiento varía en cuanto a que los coeficientes por conocer son:

$$l'_1 + l'_2 r_{xy} + l'_3\ r_{xz} = d'_x$$
$$l'_1 r_{xy} + l'_2 + l'_3 r_{yz} = d'_y$$
$$l'_1 r_{xz} + l'_2 r_{yz} + l'_3 = d'_z$$

donde:

$$l'_1 = l_1 \quad \sqrt{\Sigma(x)^2}; \qquad l'_2 = l_2 \quad \sqrt{\Sigma(y)^2}; \qquad l'_3 = l \quad \sqrt{\Sigma(y)^2}$$

[1] El procedimiento para el cálculo de matrices se encuentra en cualquier buen texto de álgebra, o ver Dwyer, P. S. *Psychometrik* 6, 101-129, 1941, o, Goulden, C. H., *Methods of Statistical Analysis.* 2ª. Ed. New York., 1950.

y las diferencias están expresadas en la siguiente forma:

$$d'_x = \frac{dx}{\sqrt{(\Sigma x^2)}}$$

$$d'_y = \frac{dy}{\sqrt{(\Sigma y^2)}}$$

$$d'_z = \frac{dz}{\sqrt{(\Sigma z^2)}}$$

Siendo:

$$D' = l'_1(d'x) + l'_2(d'y) + l'_3(d'z) = l_1(dx) + l_2(dy) + l_3\ (dz) = D$$

o sea la ecuación preliminar con la que se calcularon los otros valores.[1]

Los cálculos correspondientes son:

$$d'_x = \frac{6.95}{33.00} = 0.2106$$

$$d'_y = \frac{4.12}{27.34} = 0.1508$$

$$d'_z = \frac{9.01}{28.09} = 0.3207$$

Los coeficientes a su vez son:

$$l'_x = (0.2106)(1.0467) - (0.1508)(0.2008) - (0.3207)(0.0479 = 0.1748$$
$$l'_y = (0.2106)(-0.2008) + (0.1508)(1.2898) - (0.3207)(0.5517) = 0.0247$$
$$l'_z = (0.2106)(-0.0479) - (0.1508)(0.5517) + (0.3207)(1.2536) = 0.3087$$

Haciendo la conversión queda:

$$l_x = \frac{0.1748}{33.00} = 0.005297$$

$$l_y = \frac{-0.0247}{27.34} = -0.000903$$

$$l_z = \frac{0.3087}{28.09} = 0.010989$$

[1] Goulden, C. H., *op cit.*

y la ecuación final es:

$$D = 0.005297X - 0.000903Y + 0.10989Z$$

o sean casi los mismos valores hallados por el método directo. La diferencia se debe a la exagerada manipulación numérica con relativamente pocos decimales.

## *EL PROBLEMA DE LA TRANSVARIACIÓN*

Es muy común en trabajos de antropología física y psicología que los investigadores enfoquen su atención a los valores medios de sus distribuciones normales y a base de estimaciones de las diferencias entre medias obtengan sus inferencias y conclusiones.

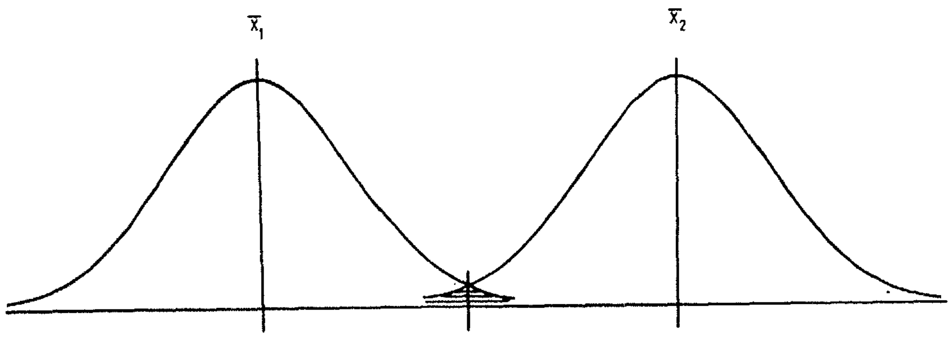


FIG. 1

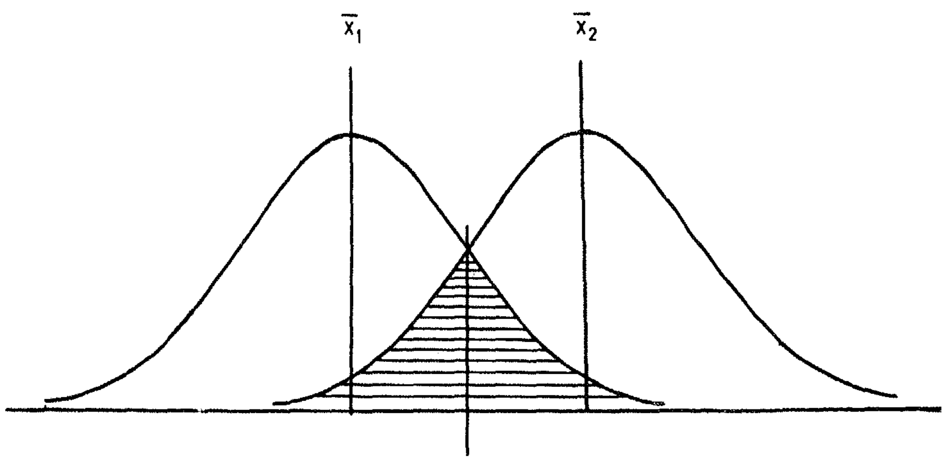


FIG. 2

De este modo cuando los medias son significativamente distintas consideran los grupos como diferentes (fig. 1) sin tomar en cuenta la transvariación o variabilidad transgresiva que es lo que comúnmente se presenta (fig. 2).

Este fenómeno de la transvariación indudablemente que influye en el diagnóstico que se haga de un espécimen determinado por medio de las funciones discriminantes, en el caso en que los valores de las variables consideradas para ur sujeto queden comprendidas en el área de la transvariación.

Por lo que respecta a los tres diámetros que se vienen considerando, las diferencias sexuales son altamente significativas según los valores de la distribución de "Student".

*Diámetro antero-posterior máximo*

$$t = \frac{182.64 - 175.69}{\sqrt{\dfrac{(4.60)^2}{25} + \dfrac{(5.15)^2}{23}}} = \frac{6.95}{\sqrt{1.9994}} = 4.92^{**\,1}$$

*Diámetro transverso máximo*

$$t = \frac{132.68 - 128.56}{\sqrt{\dfrac{(4.28)^2}{25} + \dfrac{(4.28)^2}{23}}} = \frac{4.12}{\sqrt{1.5293}} = 3.35^{**}$$

*Diámetro bizigomático*

$$t = \frac{136.96 - 127.95}{\sqrt{\dfrac{(4.10)^2}{25} + \dfrac{(4.06)^2}{23}}} = \frac{9.01}{\sqrt{1.3389}} = 7.83^{**}$$

Estos resultados serían suficientes por ellos mismos para concluir que ha diferencias sexuales muy marcadas (en esos diámetros) entre los cráneos de l Cueva de la Candelaria, pero la transvariación que existe entre ambas series qued ilustrada en las figuras 3, 4 y 5 y su efecto sobre la clasificación, no sólo en l basada en sus simples medias, sino en las funciones discriminantes será evidente

Por lo que respecta al D. A. P. Máx., las mujeres comparten un 28.11% d la distribución normal teórica de los hombres y éstos, a su vez un 21.48% de l de las mujeres. Es decir, que el área de transvariación en torno al punto de intersección de ambas curvas, obtenido por método gráfico es alrededor de 49.59%. L que corresponde al D. Tr. Mx., es de 62.77% y la del Diám. Bizigomático llega 28.33%. Esto es, que dentro de estas áreas, en términos generales y sin compara las probabilidades, no es posible saber si se trata de un cráneo masculino o femeninc

[1] Los valores tabulares de "t" para $(25 + 23 - 2) = 46$ grados de libertad son
$t(0.05) = 2.02$
$t(0.01) = 2.69$

DIAMETRO ANTEROPOSTERIOR MAXIMO en m.m.

Area de Transvariación teórica.

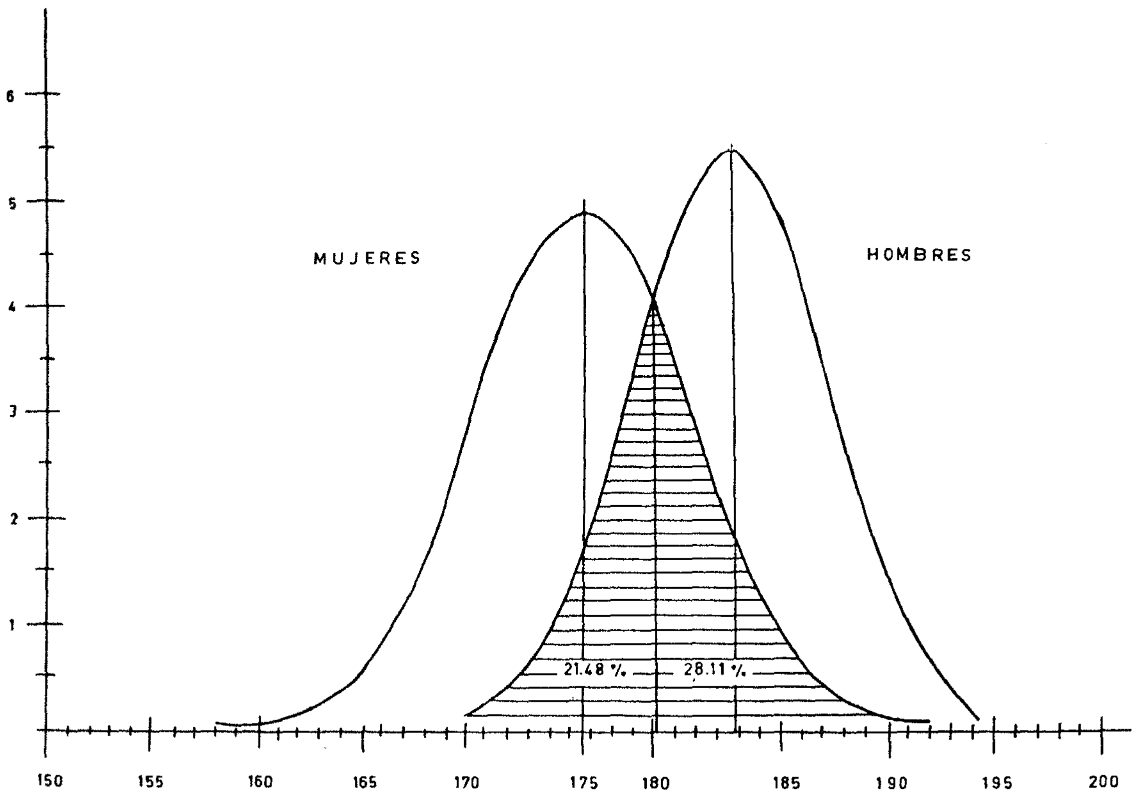


Fig. 3

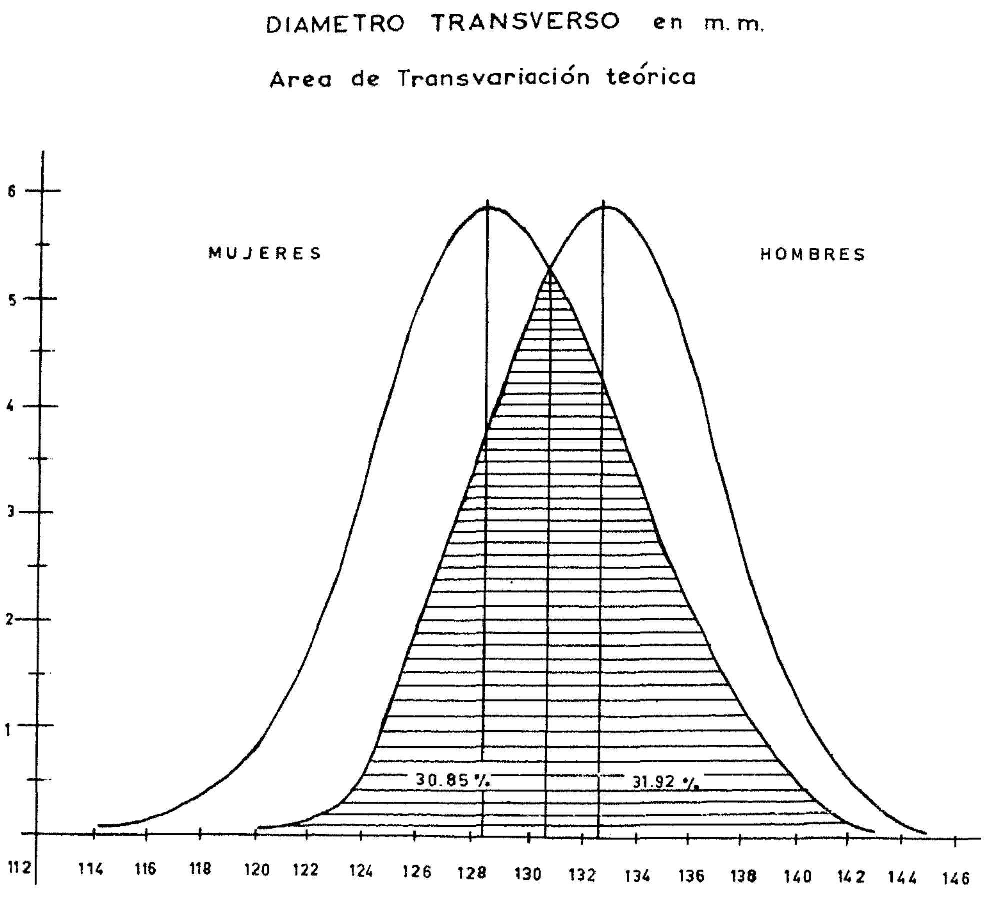
DIAMETRO TRANSVERSO en m.m.
Area de Transvariación teórica
MUJERES
HOMBRES
30.85 %
31.92 %
6
5
4
3
2
1
112
114
116
118
120
122
124
126
128
130
132
134
136
138
140
142
144
146

Fig. 4

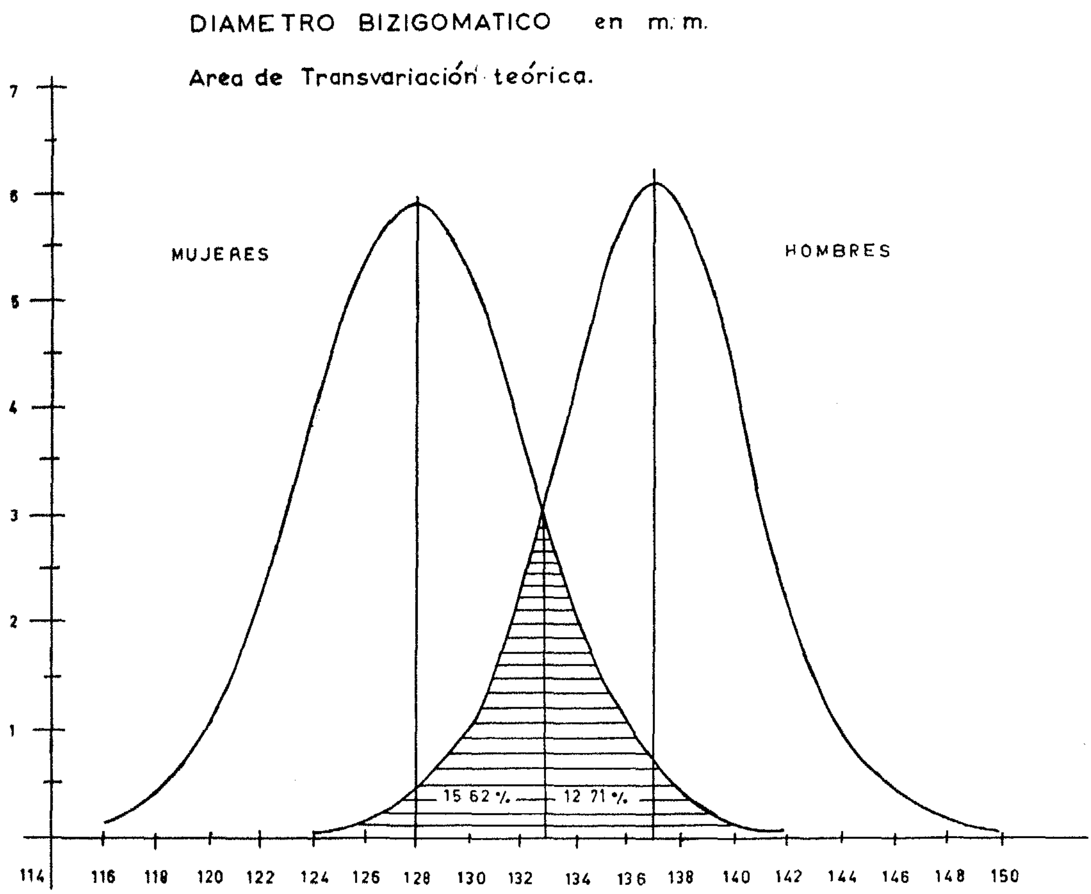
DIAMETRO BIZIGOMATICO en m. m.
Area de Transvariación teórica.
MUJERES
HOMBRES
15 62 %
12 71 %
7
6
5
4
3
2
1
114
116
118
120
122
124
126
128
130
132
134
136
138
140
142
144
146
148
150

Fig. 5

Así que las funciones sólo discriminan correctamente cuando los ejemplares cuyo diagnóstico se prentende, quedan por los caracteres de que se trate, fuera del área de variabilidad transgresiva.

Esto conduce, independientemente del poder discriminante de las funciones, a poder estimar la probabilidad de que un espécimen sea clasificado equivocadamente. Este problema lo resolvió Fisher[1] de la siguiente manera:

Primero se obtiene el valor de las ecuaciones con los valores originales (sumas de cuadrados y productos) multiplicados por sus respectivos coeficientes. En el presente caso sería:[2]

$$0.5283(1088.59) - 0.0975(191.08) + 1.18(117.34) = 694.9329$$
$$0.5283(191.08) - 0.0975(747.10) + 1.18(333.25) = 421.3403$$
$$0.5283(117.34) - 0.0975(333.25) + 1.18(739.92) = 902.6045$$

Estos valores multiplicados por los coeficientes dan:

$$0.5283(694.9329) - 0.0975(421.3403) + 1.18(902.6045) = 1391.1257$$

Esto dividido entre los grados de libertad en conjunto nos da la variancia de un ejemplar aislado.

Los grados de libertad son:

$$25 + 23 - K - 1 = 48 - 4 = 44$$

de donde:

$$s^2 = \frac{1391.1257}{44} = 31.6164$$

y la desviación estándar es:

$$s = \sqrt{31.6164} = 5.57$$

Si la desviación sigmática de un espécimen excede a la mitad de la diferencia entre las ecuaciones características, entonces se puede decir que la clasificación es incorrecta.

Las ecuaciones características en este caso son:

Hombres

$$D = 0.5283(12.64) - 0.0975(2.68) + 1.18(6.96) = 14.6292$$

Mujeres

$$D = 0.5283(5.69) - 0.0975(-1.44) + 1.18(-2.05) = 0.7274$$

[1] Fisher, R. A. *op. cit.*, p. 182.

[2] Los coeficientes han sido multiplicados por 100 para reducir decimales y son los obtenidos por el método directo.

y la diferencia entre el valor de ambas ecuaciones es:

$$\text{Dif.} = 14.6292 - 0.7274 = 13.9718$$

La mitad de tal diferencia:

$$\frac{\text{Dif.}}{2} = \frac{13.9718}{2} = 6.9859$$

La desviación sigmática correspondiente a tal valor es:

$$\frac{x}{s} = \frac{6.9859}{5.57} = 1.25$$

A una desviación sigmática de 1.25 corresponde una área de 0.8944.
La probabilidad de exceder este valor es de:

$$P = 1.0000 - 0.8944 = 0.1056$$

o sea que en este caso particular la probabilidad de atribuir erróneamente un cráneo a determinado sexo, según el valor de la función discriminante es de 10.56%, es decir, muy elevado y atribuible a la exagerada transvariación de las medias consideradas.

## *FUNCIONES PARA DISCRIMINAR CRÁNEOS MASCULINOS PREHISPÁNICOS DE LA CUEVA DE LA CANDELARIA Y DE TLATELOLCO*

La serie masculina de la Cueva de la Candelaria es la misma utilizada en el ejemplo anterior y la de Tlatelolco está formada por 25 ejemplares elegidos al azar de la colección de 32 cráneos normales procedentes de dicho lugar y estudiada por Dávalos.[1]

También en este caso se compararon las estimaciones paramétricas obtenidas por dicho autor con las basadas en la muestra de 25 ejemplares para ver si no había discrepancia. Las medidas tomadas en cuenta son las siguientes:

[1] Dávalos H. E., *La deformación craneana entre los Tlatelolcas.* Escuela Nacional de Antropología e Historia, I.N.A.H. México, 1951.

SERIE MASCULINA DE TLATELOLCO

| Dávalos (Cráneos normales) | Montemayor-Jaén. |
|---|---|
| *Diámetro anteroposterior máximo* | |
| n = 32 | n = 25 |
| M = 173.00 ± 1.31 | M = 172.54 ± 1.06 |
| s = 7.40 ± 0.92 | s = 5.33 ± 0.75 |
| *Diámetro transverso máximo* | |
| n = 32 | n = 25 |
| M = 140.89 ± 1.16 | M = 142.52 ± 1.35 |
| s = 6.55 ± 0.81 | s = 6.74 ± 0.95 |
| *Diámetro bizigomático* | |
| n = 32 | n = 25 |
| M = 130.80 ± 0.54 | M = 133.44 ± 0.72 |
| s = 3.03 ± 0.38 | s = 3.64 ± 0.51 |
| *Altura superior de la cara* | |
| No hay datos. | n = 25 |
| | M = 71.23 ± 0.62 |
| | s = 3.10 ± 0.43 |

No es necesario incluir los datos referentes a la serie de la Candelaria por haberlos presentado antes, excepto la altura superior de la cara que es:

$$n = 25$$
$$M = 72.86 \pm 0.54$$
$$s = 2.74 \pm 0.38$$

Las funciones resultantes considerando las cuatro variables siguientes en milímetros y con su codificación:

| | |
|---|---|
| Diámetro anteroposterior máximo | −170 = W |
| Diámetro transverso máximo | −130 = X |
| Diámetro bizigomático | −130 = Y |
| Altura superior de la cara | − 70 = Z |

La función resultante por el método directo y por la matriz inversa de coeficientes de correlación es la siguiente, que está multiplicada por 100.

$$D = 0.749W - 0.848X + 1.150Y - 0.277Z$$

y las ecuaciones características para cada grupo, al substituir las medias codificadas en las variables, son:

Tlatelolco

$$D_t = 0.749(2.54) - 0.848(12.52) + 1.150(3.44) - 0.277(1.23) = -5.094$$

Candelaria

$$D_c = 0.749(12.64) - 0.848(2.68) + 1.150(6.96) - 0.277(2.86) = 14.406$$

ANÁLISIS DE LA VARIANCIA

| *Fuente* | *Grados de libertad* | *Suma de cuadrados* | *Cuadrado medio* | *F* | *F(5%)* | *F(1%)* |
|---|---|---|---|---|---|---|
| Entre grupos | 4 | 0.4737 | 0.1184 | 27.53 | 2.61 | 3.83 |
| Dentro de grupo | 45 | 0.1949 | 0.0043 | | | |

Es decir, que las funciones tienen un poder altamente discriminante.

La probabilidad de clasificar equivocadamente dos ejemplares es de acuerdo con el método explicado:

Diferencia entre las ecuaciones características:

$$\text{Dif.} = -5.094 - 14.406 = -19.5$$

$$\frac{\text{Dif.}}{2} = \frac{-19.5}{2} = -9.75$$

Para estimar la variancia substituimos en las ecuaciones originales sus coeficientes:

$$1192.22(0.749) - 289.65(-0.848) - 68.05(1.150) + 181.83(-0.277) = 1009.97$$
$$-289.65(0.749) + 1532.18(-0.848) + 470.00(1.150) - 43.38(-0.277) = -963.72$$
$$-68.05(0.749) + 470.00(-0.848) + 724.62(1.150) + 95.49(-0.277) = 357.33$$
$$181.83(0.749) - 43.38(-0.848) + 95.49(1.150) + 412.86(-0.277) = 54.06$$

y multiplicamos por los mismos los valores obtenidos:

$$0.749(1009.97) - 0.848(-963.72) + 1.150(357.33) - 0.277(54.06) = 1969.6569$$

Los grados de libertad son:

$$25 + 25 - 4 - 1 = 50 - 5 = 45$$

Así que la variancia es:

$$s^2 = \frac{1969.6569}{45} = 43.7701$$

y la desviación estándar:

$$s = \sqrt{43.7701} = 6.61$$

La desviación sigmática crítica es de

$$\frac{x}{s} = \frac{9.75}{6.61} = 1.47$$

que corresponde a una área de 0.9292
La probabilidad de clasificar, usando las funciones discriminantes, un cráneo de la Candelaria como de Tlatelolco o viceversa, a base de las cuatro medias consideradas es:

$$1.0000 - 0.9292 = 0.0708$$

o sea de 7%, también muy elevado.

## *LAS FUNCIONES CON DATOS PSICOLÓGICOS*

Para este ensayo se utilizaron 3 caracteres tomados de la Prueba de Valores de G. Allport. Dicha prueba consta de una serie de preguntas encaminadas a descubrir cuales son los aspectos de la vida social que más interesan a un individuo.

El número de valores son seis, a saber: Teórico, Estético, Económico, Social, Político y Religioso. La calificación final se establece con un perfil donde aparecen los campos en que el sujeto alcanza mayor o menor calificación.

Para los fines de este trabajo sólo se escogieron 3 valores: el Teórico, el Social y el Religioso, en dos series, masculina y femenina de 20 sujetos cada una. Las pruebas fueron elegidas al azar de una serie de 28 hombres y 93 mujeres estudiantes de la Facultad de Filosofía y Letras de la U. N. A. M.

El objeto fue observar si era posible discriminar sexos en esos tres valores.

Resultados: las variables y su codificación es:

DATOS PSICOLÓGICOS

| Mujeres | | Hombres |
|---|---|---|
| | *Teórico* | |
| n = 20 | | n = 20 |
| M = 47.15 ± 1.74 | | M = 35.55 ± 1.57 |
| s = 7.80 ± 1.23 | | s = 7.02 ± 1.12 |
| | *Social* | |
| n = 20 | | n = 20 |
| M = 40.25 ± 1.53 | | M = 43.30 ± 1.36 |
| s = 6.87 ± 1.08 | | s = 6.10 ± 0.96 |
| | *Religioso* | |
| n = 20 | | n = 20 |
| M = 36.99 ± 1.70 | | M = 46.00 ± 1.97 |
| s = 7.63 ± 1.20 | | s = 8.85 ± 1.40 |
| | Teórico −40 = X | |
| | Social −40 = Y | |
| | Religioso −40 = Z | |

La función resultante es:

$$D = .29X - .12Y - .18Z$$

y substituyendo las diferencias queda:

$$D = .0337 + .0038 + .0166 = .0541$$

Las funciones características substituyendo en ellas las medias codificadas son:

Hombres

$$D = .29(7.15) - .12(.25) - .18(-3.01) = 2.5853$$

Mujeres

$$D = .29(-4.45) - .12(3.30) - .18(6.00) = -2.7665$$

El poder discriminante se obtiene por:

ANÁLISIS DE LA VARIANCIA

| *Fuente* | *Grados de libertad* | *Suma de cuadrados* | *Cuadrado medio* | *F* | *F(5%)* | *F(1%)* |
|---|---|---|---|---|---|---|
| Entre grupos | 3 | .0290 | .0097 | 6.46 | 2.87 | 4.38 |
| Dentro de grupos | 36 | .0541 | .0015 | | | |

El valor discriminante de la función es significativo al 5% y al 1%.

Los valores originales de las ecuaciones con sus coeficientes son:

$$2177.50(.29) - 1209.95(-.12) - 2036.43\ (-.18)\ 1143.2264 = 1143.2264$$
$$-1209.95(.29) + 1623.95(-.12) - 1360.25(-.18)\ -300.9145 = -300.9145$$
$$-2036.43(.29) - 1360.25(-.12) + 2616.55(-.18)\ -898.3137 = -898.3137$$

y los productos con los coeficientes dan:

$$.29(1143.2264)\ -.12(-300.9145)\ -.18(-898.3137) = 529.3418$$

Los grados de libertad son:

$$20 + 20 - 3 - 1 = 40 - 4 = 40$$

La variancia es entonces:

$$s^2 = \frac{529.3418}{40} = 13.2335$$

y la desviación estándar:

$$s = \sqrt{13.2335} = 3.63$$

La diferencia entre las ecuaciones características es de:

$$\text{Dif.} = (2.5853) - (-2.7665) = 5.3118$$

y la mitad de ella:

$$\frac{\text{Dif.}}{2} = \frac{5.3118}{2} = 2.6559$$

El valor sigmático crítico para clasificar los sexos equivocadamente es:

$$\frac{x}{s} = \frac{2.6559}{3.63} = 0.7316$$

al cual corresponde una área de 0.7673 y la probabilidad de exceder esa área es de:

$$P = 1.0000 - 0.7673 = 0.2327$$

o sea 23.27% de probabilidades de error.

Las reservas que al principio de este ensayo se hicieron para no generalizar estos resultados se deben a que la población a la que corresponden las pruebas seleccionadas es de edades muy variables y a que no se dispone de información sobre las condiciones en que tales pruebas fueron administradas.

El objeto de incluir estos datos, fue el de insistir en los procedimientos del cálculo de las funciones discriminantes, reduciéndolos, en este caso, a sus principales pasos.

## *COMENTARIO*

En el presente ensayo sólo se pretendió mostrar la aplicación de un recurso técnico de valor indiscutible para el investigador y no sacar conclusiones o exponer puntos de vista antropológico físicos o psicológicos.

La elevada probabilidad de discriminar equivocadamente un sujeto que aparece en los ejemplos presentados puede deberse a muchos factores, entre ellos el número de variables, lo reducido de las muestras, la variabilidad transgresiva, etc., pero la consistencia del método es indudable y más sólida que cualquier procedimiento subjetivo para encarar el problema.

Para los elementos técnicos habituales en el trabajo del antropólogo o el psicólogo, el cálculo de ecuaciones con más de tres incógnitas resulta engorroso por los continuos errores personales. Pero habiendo la posibilidad de usar computadoras electrónicas, la labor del investigador se limita al establecimiento de las funciones iniciales y deja en manos del programador la solución de las mismas en un lapso mínimo. De este modo puede usar el número de variables que considere apropiado para el caso.