

**Distribuciones en el muestreo para una normal multivariada****Distribución Normal Multivariante**

La distribución normal p-dimensional tiene en el análisis multivariante un papel similar al de la normal en el caso unidimensional, siendo la base de muchas técnicas estadísticas.

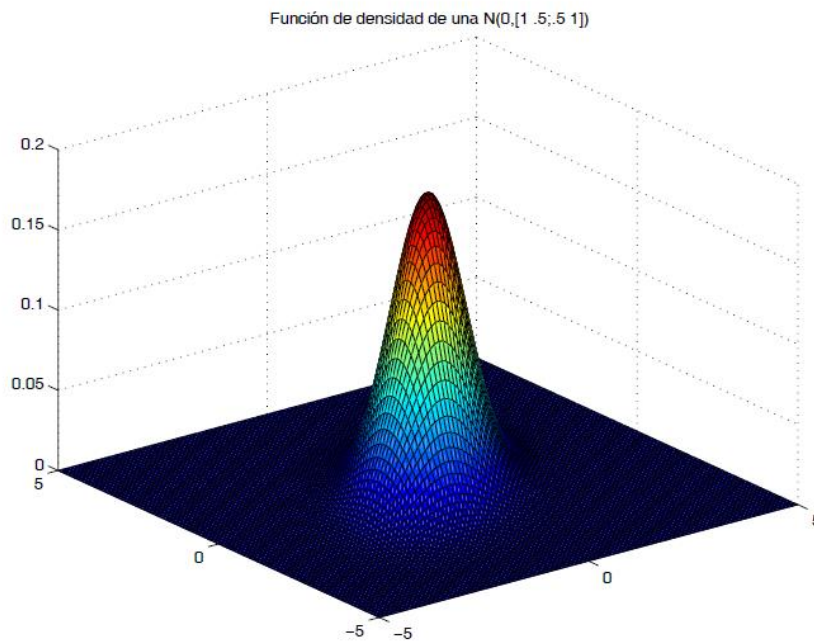
La expresión de su función de densidad de probabilidad es:

$$f(x) = (2\pi)^{\frac{-p}{2}} |\Sigma|^{-1/2} \exp\left(-\frac{(x-\mu)' \Sigma^{-1} (x-\mu)}{2}\right) \text{ con } x \in R^p$$

Esta expresión es claramente una generalización de la distribución normal univariada, o bien la expresión univariada es un caso particular de esta.

Para p=2 la gráfica tiene la siguiente forma:

## La distribución normal multivariante



Entre sus ventajas se pueden destacar las siguientes:

1. Es una generalización sencilla del caso univariante, tiene propiedades matemáticas que la hacen muy manejable
2. Depende de un número relativamente reducido de parámetros:  $p$  para la media y  $p(p+1)/2$  para la varianza
3. Independencia lineal es equivalente en este caso a independencia Estadística
4. Si bien, los datos nunca son exactamente normales, la distribución normal suele ser una aproximación útil de la verdadera distribución.

5. Es el límite de la suma de vectores aleatorios independiente y con la misma distribución (versión multivariada del TCL)

### **Propiedades de la Distribución Normal Multivariante**

Si  $\mathbf{X} \sim \mathbf{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  entonces cualquier combinación lineal de sus coordenadas también tiene distribución normal:  $\mathbf{Y} = \mathbf{c}'\mathbf{X} \sim \mathbf{N}(\mathbf{c}'\boldsymbol{\mu}, \mathbf{c}'\boldsymbol{\Sigma}\mathbf{c})$

Siendo  $\mathbf{c}$  un vector de constantes conocidas.

*En particular cada distribución marginal ( vale decir la proyección sobre una dirección principal) es normal.*

Es decir, podemos dividir la variable  $\mathbf{x}$  como sigue:

$$\mathbf{X} = (\underbrace{x_1, x_2, \dots, x_q}_{\mathbf{y}_1}, \underbrace{x_{q+1}, \dots, x_p}_{\mathbf{y}_2})$$

lo que permite definir las variables  $\mathbf{y}_1 = (x_1, \dots, x_q)$  y  $\mathbf{y}_2 = (x_{q+1}, \dots, x_p)$  normales.

### **Ejemplo 1:**

Sea  $\mathbf{x} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  donde  $\boldsymbol{\mu} = (1, 2, 3)'$  y  $\boldsymbol{\Sigma}$  es la identidad.

Sean las combinaciones lineales:

$$y_1 = 2x_1 + 3x_2 \text{ e } y_2 = -x_1 + 2x_3.$$

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 2 & 3 & 0 \\ -1 & 0 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 2x_1 + 3x_2 \\ -x_1 + 2x_3 \end{pmatrix}$$

Entonces, la distribución (conjunta) de la variable  $y = (y_1, y_2)$  es normal multivariante con parámetros:

$$\mu_Y = \begin{pmatrix} 2 & 3 & 0 \\ -1 & 0 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 8 \\ 5 \end{pmatrix}$$

$$\Sigma_Y = \begin{pmatrix} 2 & 3 & 0 \\ -1 & 0 & 2 \end{pmatrix} I \begin{pmatrix} 2 & -1 \\ 3 & 0 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} 13 & -2 \\ -2 & 5 \end{pmatrix}$$

### **Estimaciones de máxima verosimilitud**

Los estimadores máximo – verosímiles de la distribución  $N_p(\mu, \Sigma)$  se obtienen maximizando la función de verosimilitud con respecto a los parámetros  $\mu$  y  $\Sigma$ .

La función de verosimilitud de una variable es la función de distribución conjunta de la misma pero considerada como función de los parámetros en lugar de función de las variables. La estimación de máxima verosimilitud es la que se obtiene maximizando la verosimilitud, es decir hallando los valores de parámetro que la hacen máxima.

Se puede demostrar que estos estimadores para el caso normal multivariado son:

$$\hat{\mu} = \bar{x} \quad \text{y} \quad \hat{\Sigma} = S.$$

El estimador máximo – verosímil de la normal multivariante verifican:

- ♥  $\bar{x} \sim N_p(\mu, \frac{1}{n}\Sigma)$
- ♥  $nS$  se distribuye Wishart con  $n-1$  grados de libertad
- ♥  $\bar{x}$  y  $S$  son independientes

### Distribución asintótica de los estimadores de máxima verosimilitud:

#### Ley de los grandes números:

Si  $x$  es una variable multivariante con esperanza  $\mu$  y matriz de covarianzas  $\Sigma$ , entonces:

⌚  $\bar{x}$  converge a  $\mu$ .

⌚  $S$  converge a  $\Sigma$ .

#### El teorema central del límite:

Si  $x$  es una variable multivariante con esperanza  $\mu$  y matriz de covarianzas  $\Sigma$ , entonces:

$$\sqrt{n}(x - \mu) \sim N_p(0, \Sigma).$$

### Distribución de Wishart y distribución de Hotelling

**Definición** Se dice que una matriz  $V$  cuadrada de orden  $p$  sigue una distribución de **Wishart**  $Wp(n, \Sigma)$  si y solo  $V$  puede escribirse como:

$$V = \sum_{i=1}^n x_i \cdot x_i'$$

donde los  $x_i$  son i.i.d. con  $x_i \sim N_p(0, \Sigma)$ .

Es decir que la distribución de Wishart obedece a la suma de los productos de normales multivariadas de media 0 y varianza común  $\Sigma$  por sus respectivas traspuestas.

La distribución de Wishart generaliza la distribución  $\chi_n^2$ .

### **Proposición:**

La suma de variables aleatorias independientes con distribución Wishart cada una de la misma dimensión y matriz de covarianzas es una nueva distribución de Wishart con la misma dimensión y la misma matriz de covarianzas y parámetro  $n$  suma de los parámetros de las dos variables sumadas.

Si  $V_1 \sim W_p(n_1, \Sigma)$ ,  $V_2 \sim W_p(n_2, \Sigma)$  siendo  $V_1$  y  $V_2$  son independientes, entonces

$$V_1 + V_2 \sim W_p(n_1 + n_2, \Sigma).$$

### **Distribución de la media y la varianza muestrales**

En el caso de una variable aleatoria normal unidimensional  $X$ , la media muestral  $\bar{x}$  sigue una distribución normal con parámetros:

$$\bar{x} \sim N(\mu; \sigma^2) \quad \text{y} \quad \frac{(n-1)s^2}{\sigma^2} \sim \chi_{n-1}^2$$

Generalicemos esta propiedad para el caso un vector de normal multivariado.

Si las  $x_i$  son i.i.d. y  $x_i \sim N_p(\mu, \Sigma)$ , la distribución del vector de medias muestrales es:

$$\bar{x} \sim N_p(\mu, \Sigma/n).$$

La distribución de Wishart es la distribución de una matriz de varianza-covarianza empírica.

Si  $\{x_1, x_2, \dots, x_n\}$  es una muestra aleatoria de la  $N(\mu, \sigma^2)$  con  $\sigma > 0$ , entonces:

$$\frac{\sum (x_i - \bar{x})^2}{\sigma^2} \sim \chi_{n-1}^2$$

Consideremos una muestra aleatoria simple de tamaño  $n$  de un vector aleatorio de  $p$  componentes con distribución  $N_p(\mu, \sigma^2)$ .

$$X = \begin{pmatrix} x_1 \\ \dots \\ x_p \end{pmatrix}$$

Sea  $M$  la matriz que tiene en las filas las realizaciones independientes

$$X_i \sim N_p(\mu, \sigma^2).$$

Consideremos la matriz de varianzas y covarianzas

$$D = \sum_{i=1}^n (x_i - \mu)(x_i - \mu)'$$

vale decir la matriz de las observaciones centradas en las medias de la población.

### Propiedades:

De la misma manera que en el caso  $p = 1$ ,  $\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$  es una estimación insesgada de la varianza  $\sigma^2$  de la población cuando la media  $\mu$  es conocida, la matriz:

$$\frac{1}{n} D = \frac{1}{n} \sum_{i=1}^n (x_i - \mu) (x_i - \mu)'$$

Es una estimación insesgada de  $\Sigma$  cuando el vector de medias  $\mu$  es conocido.

La matriz  $D$  es semi-definida positiva; es definida positiva (p.s.) cuando  $\Sigma$  es invertible.

La matriz  $D$  tiene una distribución de Wishart  $W_p(n, \Sigma)$ .

### Proposición

Sean  $\bar{x}$  el vector de medias empíricas y  $V$  la matriz de varianzas-covarianzas empíricas:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{y} \quad V = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})'$$

entonces

$$\bar{x} \sim N_p(\mu; \Sigma) \quad \text{y} \quad V \sim W_{p(n-1; \Sigma)}$$

### La distribución $T^2$ de Hotelling



Esta distribución generaliza la distribución *t*-Student. La v.a. *t de Student* es el cociente entre una variable aleatoria normal standard y la raíz cuadrada de una variable aleatoria Chi cuadrado independiente de la variable normal del numerador, de *n* grados de libertad dividida por sus grados de libertad.

$$\frac{\sqrt{n}(\bar{x}_n - \mu)}{\sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^2}} \sim t_{n-1}.$$

### Definición

Si el vector  $x \sim N_p(0, I_p)$ ,  $V \sim W_p(n, I_p)$  y  $x$  es independiente de  $D$ , entonces

$$n x' V^{-1} x$$

sigue una distribución de  **$T^2$  de Hotelling de parámetro  $n$**  denotada  $T^2_p(n)$ .

**Proposición** Se puede escribir la  $T^2_p(n)$  de Hotelling en función de una  $F$  de Fisher:

$$T^2_p(n) = np/(n-p+1) F_{p, n-p+1}$$

Entonces

Si  $x \sim N_p(\mu, \Sigma)$ ,  $D \sim W_p(n, \Sigma)$  y  $x$  independiente de  $D$ , entonces:

$$n(x-\mu)' D^{-1} (x-\mu)$$

sigue una distribución de Hotelling  $T^2_p(n)$ .

**Proposición** Sean una muestra  $\{x_1, x_2, \dots, x_n\}$  i.i.d.,

$x \sim N_p(\mu, \Sigma)$ , y  $S$  es la matriz de varianzas y covarianzas muestrales;  
entonces:

$$n(\bar{x} - \mu)' S^{-1} (\bar{x} - \mu) \sim T_p^2 (n-1)$$

### Test de vector de medias para una población

Sea ahora una muestra aleatoria multivariada  $\{x_1, x_2, \dots, x_n\}$  donde cada  $x_i \sim N_p(\mu, \Sigma)$  con  $\Sigma$  invertible.

Queremos construir un test para la hipótesis nula

$H_0 : \mu = \mu_0$  versus

$H_1 : \mu \neq \mu_0$ .

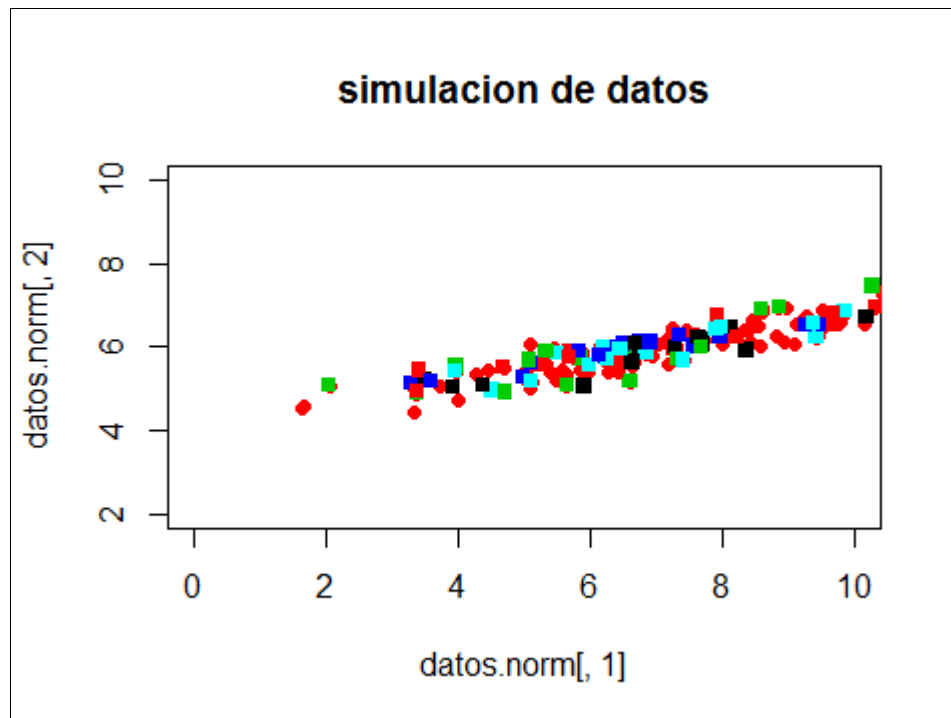
Utilizamos el estadístico:

$$\frac{(n-p)}{n} (\bar{x} - \mu_0)' S^{-1} (\bar{x} - \mu_0) \sim F_{p, n-p}$$

### Ejemplo 2

Si nos interesa contrastar las hipótesis:

$H_0: \mu = (\sqrt{50}; 6)'$  vs  $H_1: \mu \neq (\sqrt{50}; 6)'$



para una muestra de 40 individuos de una variable bidimensional con matriz de varianzas y covarianzas muestral

$$S = \begin{bmatrix} 4,288 & 1,244 \\ 1,244 & 0,428 \end{bmatrix}$$

Y vector de medias muestrales:

$$\bar{x} = (6,9; 6,2)'$$

$$\frac{40-2}{2} \left[ (6.069, 6.6225) - (\sqrt{50}; 6) \right]' \begin{bmatrix} 4,288 & 1,244 \\ 1,244 & 0,428 \end{bmatrix}^{-1} (6.069, 6.6225) - (\sqrt{50}; 6) = 17.77$$

El estadístico construido resulta igual a 17,77; como  $P(F_{2,38} > 17.77) < 0.001$  entonces se rechaza la hipótesis de nulidad.

No podemos sostener que la media es igual a  $\mu = (\sqrt{50}; 6)'$

### Test para comparar medias de dos poblaciones

Sea una variable  $X$  observada en dos poblaciones  $P_1$  y  $P_2$  y dos muestras independientes con distribuciones multivariadas. Suponiendo que ambas poblaciones tienen distribución normal con la misma matriz de varianzas y covarianzas; nos interesa comparar sus vectores medios; es decir:

$$H_0: \mu_1 = \mu_2$$

contra la hipótesis alternativa

$$H_1: \mu_1 \neq \mu_2.$$

Sean  $\bar{x}_1$  y  $\bar{x}_2$  los dos vectores de medias muestrales,  $A_1$  y  $A_2$  las dos matrices de covarianzas muestrales.

Entonces  $S=(A_1+A_2)/(n_1+n_2-2)$  es un estimador insesgado de la matriz de varianzas común  $\Sigma$ .

Además  $\bar{x}_1$  es independiente de  $A_1$   $\bar{x}_2$  es independiente de  $A_2$ ; entonces:

$$\frac{n_1 n_2}{n_1 + n_2} (\bar{x}_1 - \bar{x}_2)' S^{-1} (\bar{x}_1 - \bar{x}_2) \sim T_p^2(n_1 + n_2 - 2)$$

$$\frac{n_1 + n_2 - p - 1}{(n_1 + n_2 - 2)p} T_p^2(n_1 + n_2 - 2) \sim T_p^2(n_1 + n_2 - 2)$$

### Ejemplo 3:

Se desean comparar dos especies de moscas de agua: *Amerohelea* fascinata,

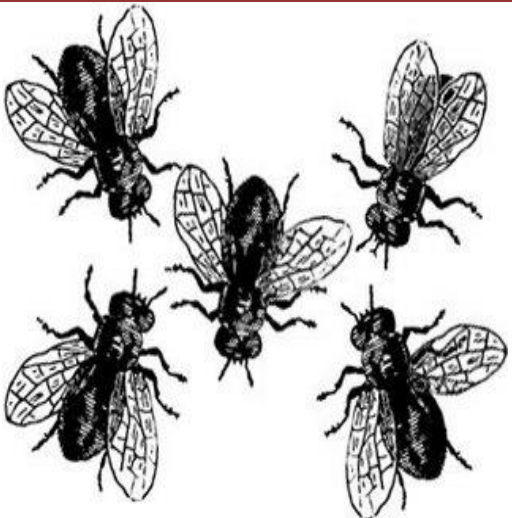
*Amerohelea pseudofascinata*. En relación a las variables:

$X_1$  = long.antena,

$X_2$  = long. ala (en mm)

para dos muestras independientes de tamaños  $n_1 = 9$  y  $n_2 = 6$ , se han obtenido las matrices de datos:

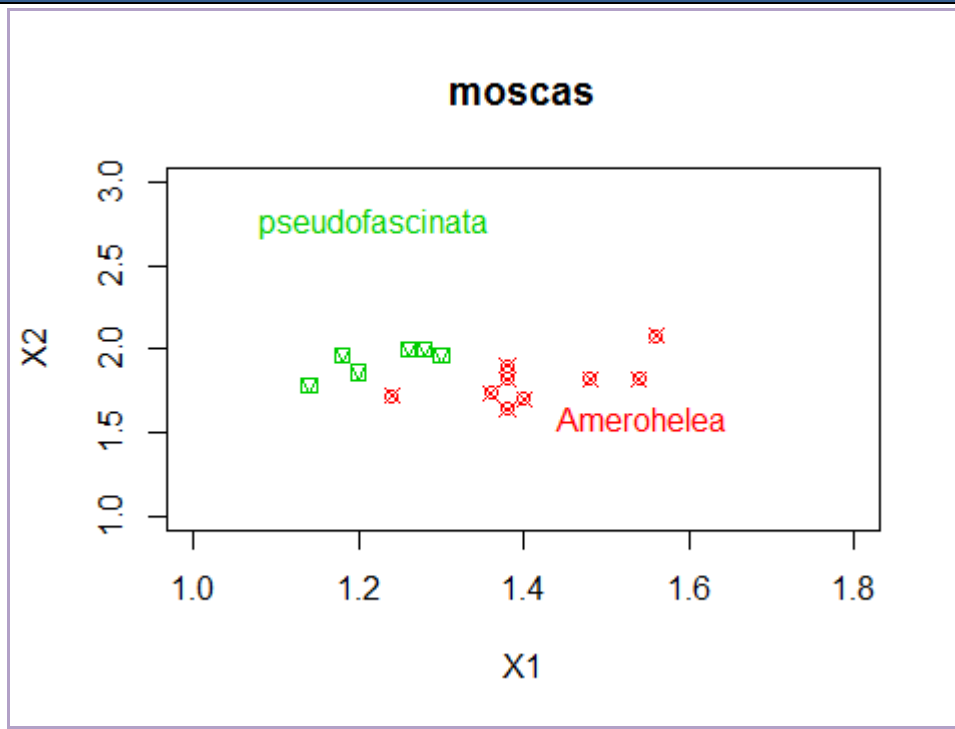
| $X_1$ | $X_2$ | grupo                  |
|-------|-------|------------------------|
| 1.38  | 1.64  | <i>Amerohelea</i>      |
| 1.4   | 1.7   | <i>Amerohelea</i>      |
| 1.24  | 1.72  | <i>Amerohelea</i>      |
| 1.36  | 1.74  | <i>Amerohelea</i>      |
| 1.38  | 1.82  | <i>Amerohelea</i>      |
| 1.48  | 1.82  | <i>Amerohelea</i>      |
| 1.54  | 1.82  | <i>Amerohelea</i>      |
| 1.38  | 1.9   | <i>Amerohelea</i>      |
| 1.56  | 2.08  | <i>Amerohelea</i>      |
| 1.14  | 1.78  | <i>pseudofascinata</i> |
| 1.2   | 1.86  | <i>pseudofascinata</i> |
| 1.18  | 1.96  | <i>pseudofascinata</i> |
| 1.3   | 1.96  | <i>pseudofascinata</i> |
| 1.26  | 2     | <i>pseudofascinata</i> |
| 1.28  | 2     | <i>pseudofascinata</i> |



Vectores de medias (valores multiplicados por 100):

$$\bar{x} = (141.33, 180.44),$$

$$\bar{y} = (122.67, 192.67)$$



Las matrices de covarianzas son:

$$S_1 = \begin{pmatrix} 98 & 80.83 \\ 80.83 & 167.78 \end{pmatrix}$$

$$S_2 = \begin{pmatrix} 39.47 & 43.47 \\ 43.47 & 77.87 \end{pmatrix}$$

Estimación centrada de la matriz de covarianzas común:

$$S = \frac{1}{13} (8S_1 + 5S_2) = \begin{pmatrix} 75.49 & 66.46 \\ 66.46 & 133.81 \end{pmatrix}$$

La distancia de Mahalanobis entre las dos muestras:

$$D^2 = (x - y)' S^{-1} (x - y) = 15.52$$

El estadístico  $T^2$ :

$$T^2 = \frac{6 \cdot 9}{6+9} D^2 = 55.87$$

$$\frac{9 + 6 - 1 - 2}{2(9 + 6 - 2)} T^2 \sim F_{2,12}$$

$F^{\text{obs}} = 25.78$ , luego rechazamos la  $H_0$ .

### Análisis de perfiles



Por ejemplo si consideramos dos grupos de personas, un grupo con  $n_1$  individuos y el otro con  $n_2$  individuos.

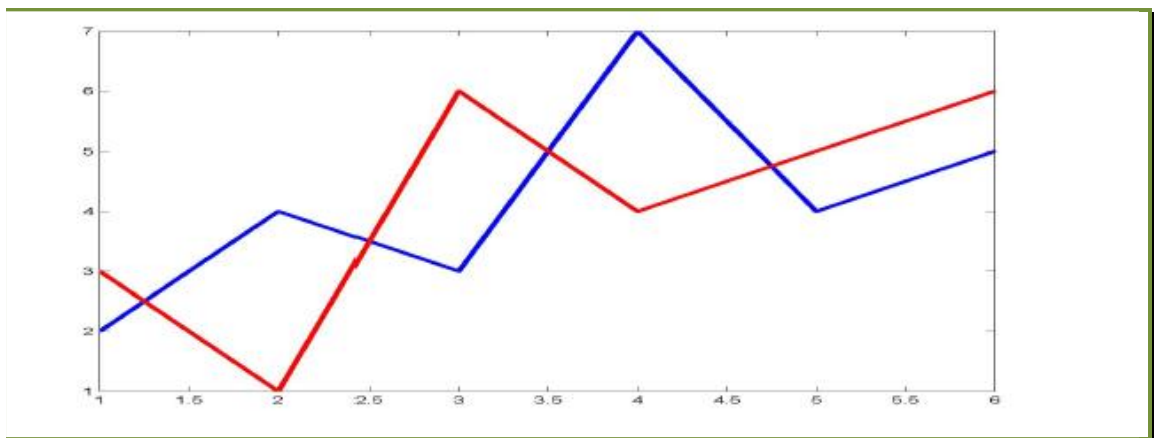
***En cada uno de los grupos se aplica a un tratamiento distinto y se mide el resultado del tratamiento con un examen en  $p$  diferentes instantes.***

En la figura se presenta un ejemplo de las medias en cada instante para cada uno de los grupos. En abscisa tenemos los instantes o repeticiones y en la ordenada el valor de la media en este instante.

Las observaciones correspondientes al seguimiento de cada grupo constituye un **perfil**.

Lo importante a observar aquí es que longitudinalmente las observaciones no son independientes ya que son medidas sobre el mismo individuo u objeto.

Más allá que estudiar la igualdad de las curvas (o los perfiles), queremos contestar a tres preguntas:







Trataremos el caso de dos grupos

### Caso de dos perfiles

Sean  $\mu_1$  y  $\mu_2$  los vectores de medias poblacionales correspondientes a las dos poblaciones consideradas.

$$\mu_1 = \begin{pmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{13} \end{pmatrix} \text{ y } \mu_2 = \begin{pmatrix} \mu_{21} \\ \mu_{22} \\ \mu_{23} \end{pmatrix}$$

Si nos interesa saber si los perfiles **son idénticos** para las dos poblaciones se debe realizar el test para

$$H_0: \mu_1 = \mu_2 \text{ versus}$$

$$H_1: \mu_1 \neq \mu_2$$

pero para testear la hipótesis de paralelismo...

$$H_0: \mu_{11} - \mu_{21} = \mu_{12} - \mu_{22} = \dots = \mu_{p1} - \mu_{p2}$$

O bien

$$H_0 = C(\mu_1 - \mu_2) = 0$$

Siendo la matriz C la siguiente:

$$C = \begin{pmatrix} 1 & -1 & 0 & 0 & \dots & 0 \\ 0 & 1 & -1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 & -1 \end{pmatrix}$$

Y entonces el estadístico de contraste es:

$$\frac{n_1 n_2}{n_1 + n_2} (\bar{x}_1 - \bar{x}_2)^t C^t (C S C^t)^{-1} C (\bar{x}_1 - \bar{x}_2) \sim T_{p-1}^2(n_1 + n_2 - 2)$$

$$\frac{n_1 + n_2 - p}{(n_1 + n_2 - 2)(p-1)} T_{p-1}^2(n_1 + n_2 - 2) \sim F_{p-1, n_1 + n_2 - p}$$

#### Ejemplo 4:

Deseamos comparar dos grupos mediante las variables  $V_1$ ,  $V_2$  y  $V_3$

Disponemos de 5 observaciones de estas variables para cada uno de los dos grupos.

El vector medio total (estimando la misma media para los dos grupos) es:

#### Vector medio total

| <u>V1</u> | <u>V2</u> | <u>V3</u> |
|-----------|-----------|-----------|
| 24.60     | 36.10     | 42.30     |

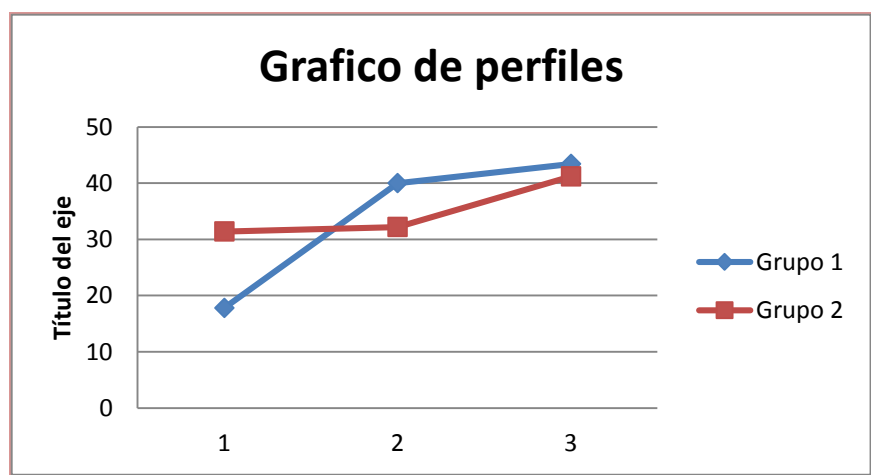
Estimamos los vectores medios de cada grupo, para ver si son similares o diferentes los grupos en estas variables.

Si observamos el vector de medias por grupo se aprecian diferencias importantes entre los grupos.

pero ...serán significativas??

#### Vectores medios por grupo

| <u>Grupo</u> | <u>V1</u> | <u>V2</u> | <u>V3</u> |
|--------------|-----------|-----------|-----------|
| 1.00         | 17.80     | 40.00     | 43.40     |
| 2.00         | 31.40     | 32.20     | 41.20     |



Utilizaremos la matriz de covarianzas común para realizar la prueba:

### Matriz de covarianzas común (insesgada)

|    | V1    | V2    | V3    |
|----|-------|-------|-------|
| V1 | 4.25  | -1.93 | -4.38 |
| V2 | -1.93 | 3.10  | 3.98  |
| V3 | -4.38 | 3.98  | 67.50 |

Aprovechamos el test de Hotelling que provee InfoStat

### Cuadro de Análisis de la Varianza (Lawley-Hotelling)

| F.V.  | Estadístico | F     | gl (num) | gl (den) | p      |
|-------|-------------|-------|----------|----------|--------|
| Grupo | 14.88       | 29.75 | 3        | 6        | 0.0005 |

Las diferencias entre los vectores medios de ambos grupos resultaron estadísticamente significativas.

Mirando el grafico de los perfiles, no tiene sentido preguntarnos si los perfiles son paralelos, si lo tuviera podemos proceder de la siguiente manera:

Introducimos los contrastes en InfoStat:

### Coeficientes de la matriz C'

| Grupo | C' (1) | C' (2) | C' (3) |
|-------|--------|--------|--------|
| 1.00  | 1.00   | -1.00  | 0.00   |
| 2.00  | 0.00   | 1.00   | -1.00  |

### Cuadro de análisis de la varianza para combinaciones lineales (Lawley-Hotelling)

| Grupo       | Estadístico | F       | gl (num) | gl (den) | p       |
|-------------|-------------|---------|----------|----------|---------|
| Fila 1 de C | 698.15      | 1396.30 | 3        | 6        | <0.0001 |
| Fila 2 de C | 14.88       | 29.75   | 3        | 6        | 0.0005  |
| Fila 3 de C | 759.95      | 1519.90 | 3        | 6        | <0.0001 |

|       |         |         |   |    |         |
|-------|---------|---------|---|----|---------|
| Total | 1458.10 | 1215.08 | 6 | 10 | <0.0001 |
|-------|---------|---------|---|----|---------|

Se rechaza la hipótesis nula de paralelismo, **los perfiles no son paralelos.**

**Cuando las medias de dos grupos son significativamente distintas, puede ser de utilidad considerar estas variables para asignar a un individuo a uno de los dos grupos**

## Análisis Discriminante

El análisis discriminante tiene por **objetivo** encontrar una **función** , tal que al aplicarla a un nuevo individuo, nos permita clasificarlo de acuerdo con el valor que este presenta en un conjunto de variables que denominaremos **variables discriminantes** asignándolo a uno de los grupos previamente conocidos o definidos.

Esta técnica se llama **supervisada** pues se conoce la pertenencia de cada individuo a uno de los grupos así como el valor que toman sus variables.

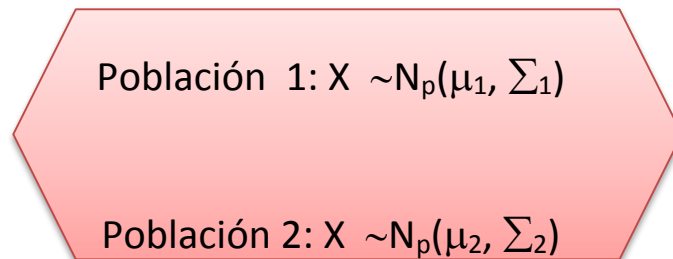
Veremos luego que existen algoritmos de clasificación no supervisada como el análisis de cluster donde se desconoce la existencia de un grupo de pertenencia.

Partimos de una tabla de datos donde se registraron sobre  $N$  individuos  $p$  variables y el grupo de pertenencia; es decir una tabla de  $N \times (p+1)$ .

El análisis discriminante muestra para cada una de las variables consideradas si **tienen o no un poder discriminante**, al cuantificar su peso en la función discriminante.

Consideremos en primera instancia el caso de **dos poblaciones** cuya distribución es **normal multivariada**.

Supongamos que nuestros individuos de interés provienen de dos poblaciones normales multivariadas:



Y sabemos que un nuevo individuo  $X$ , proviene de alguna de estas dos poblaciones a las que denominaremos  $\pi_1$  y  $\pi_2$  respectivamente.

**Buscamos una regla para predecir a cuál de las dos poblaciones es más probable que pertenezca este nuevo individuo.**

Podemos pensar esto de distintas maneras:

**Opción 1:** en función de la verosimilitud de  $x$  en cada población

Asigno al sujeto a  $\pi_1$  si  $L(X, \mu_1, \Sigma_1) > L(X, \mu_2, \Sigma_2)$

Esto es asignar al sujeto a  $\pi_1$  si su función de verosimilitud (al ser una sola observación es su función de densidad de probabilidad) toma valor superior para la población 1.

**$X \in \pi_1$**  cuando

$$(2\pi)^{-\frac{p}{2}} |\Sigma_1|^{-1/2} \exp\left(-\frac{(x-\mu_1)'\Sigma_1^{-1}(x-\mu_1)}{2}\right) > \\ (2\pi)^{-\frac{p}{2}} |\Sigma_2|^{-1/2} \exp\left(-\frac{(x-\mu_2)'\Sigma_2^{-1}(x-\mu_2)}{2}\right)$$

En el caso particular que  $\Sigma_1 = \Sigma_2$

Esta regla se simplifica

$X \in \pi_1$  cuando

$$\exp\left(-\frac{(x-\mu_1)'\Sigma^{-1}(x-\mu_1)}{2}\right) > \exp\left(-\frac{(x-\mu_2)'\Sigma^{-1}(x-\mu_2)}{2}\right)$$

Y aplicando la propiedad distributiva, se obtiene:

$X \in \pi_1$  cuando

$$(\mu_1 - \mu_2)'\Sigma^{-1}x > (\mu_1 - \mu_2)'\Sigma^{-1}(\mu_1 + \mu_2)$$

Que suele escribirse de la siguiente manera:

$X \in \pi_1$  cuando  $b x > k$  siendo  $b = (\mu_1 - \mu_2)'\Sigma^{-1}$  y  $k = (\mu_1 - \mu_2)'\Sigma^{-1}(\mu_1 + \mu_2)$

**Opción 2:** en función de la distancia de Mahalanobis

$X \in \pi_1$  cuando  $d_1 < d_2$  siendo  $d_i$  el cuadrado de la distancia de Mahalanobis al grupo  $i$ .



$$d_i = (x - \mu_i)' \Sigma^{-1} (x - \mu_i)$$

**$D_i$**  es una medida de lo lejos que esta  $x$  de  $\mu_i$

**$D_i$**  es una distancia que considera las matrices de varianzas y covarianzas.

**Opción 3:** regla de la probabilidad a posteriori

**$X \in \pi_1$**  cuando  $P(\pi_1 | X) > P(\pi_2 | X)$

La principal ventaja de la regla de la probabilidad a posteriori es que brinda una indicación de cuanta confianza se podría tener en que este investigador este tomando la decisión correcta.

**Observación:** se le dice “probabilidad a posteriori” pero en realidad no es una probabilidad. La observación pertenece a una población u otra, la incertidumbre proviene de la capacidad de la regla creada por el investigador para elegir la población correcta.

Por ejemplo si:  $P(\pi_1 | X) = 0.53$  y  $P(\pi_2 | X) = 0.47$  entonces la clasificación de ese individuo no es tan confiable como si  $P(\pi_1 | X) = 0.96$  y  $P(\pi_2 | X) = 0.04$ .

**Importante:** para el caso en que las matrices de covarianza son iguales, las tres reglas coinciden.

**Reglas basadas en estimaciones de los parámetros**

Cuando disponemos de un par de muestras de  $\pi_1$  y de  $\pi_2$  en realidad desconocemos el verdadero valor de  $\mu_1$  de  $\mu_2$  de  $\Sigma_1$  y de  $\Sigma_2$ , entonces trabajamos con sus estimaciones:

$$\widehat{\mu}_1 = \bar{x}_1 \quad \widehat{\mu}_2 = \bar{x}_2 \quad \widehat{\Sigma} = \frac{(n_1-1)S_1 + (n_2-1)S_2}{n_1+n_2-2}$$

**Ejemplo 5** Consideremos nuevamente el caso de las moscas.

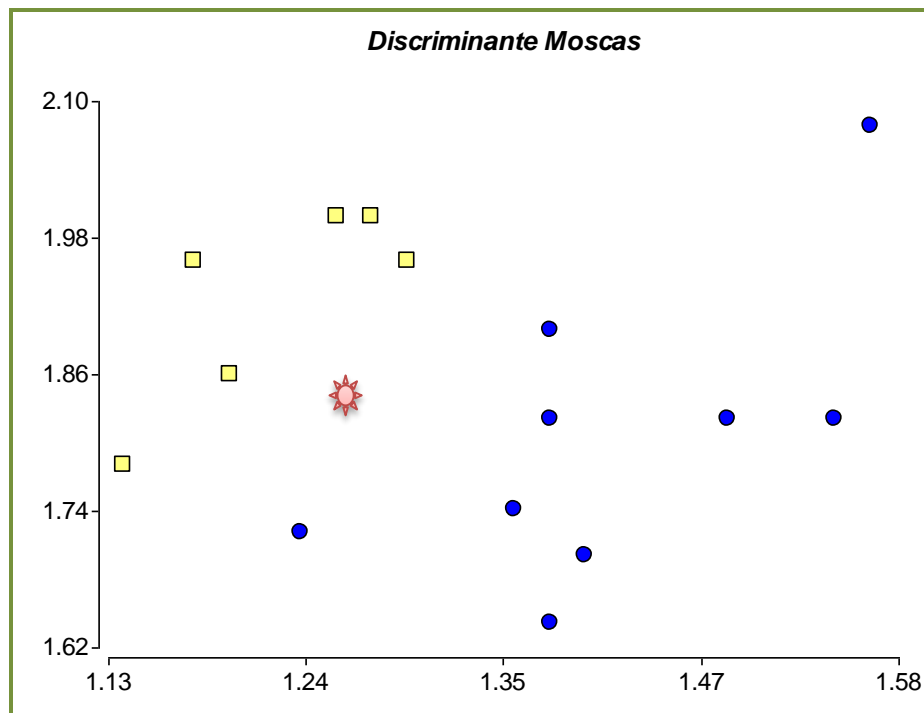
|       |          |         |          |          |          |          |          |          |          |          |          |          |          |          |          |
|-------|----------|---------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| $x_1$ | 1.3<br>8 | 1.<br>4 | 1.2<br>4 | 1.3<br>6 | 1.3<br>8 | 1.4<br>8 | 1.5<br>4 | 1.3<br>8 | 1.5<br>6 | 1.1<br>4 | 1.2      | 1.1<br>8 | 1.3      | 1.2<br>6 | 1.2<br>8 |
| $x_2$ | 1.6<br>4 | 1.<br>7 | 1.7<br>2 | 1.7<br>4 | 1.8<br>2 | 1.8<br>2 | 1.8<br>2 | 1.9      | 2.0<br>8 | 1.7<br>8 | 1.8<br>6 | 1.9<br>6 | 1.9<br>6 | 2        | 2        |
| grupo | A        | A       | A        | A        | A        | A        | A        | A        | A        | B        | B        | B        | B        | B        | B        |

Hemos probado con el test de Hotelling que las medias de ambas poblaciones son significativamente distintas.

Y supongamos que tenemos una nueva observación y queremos clasificarla:

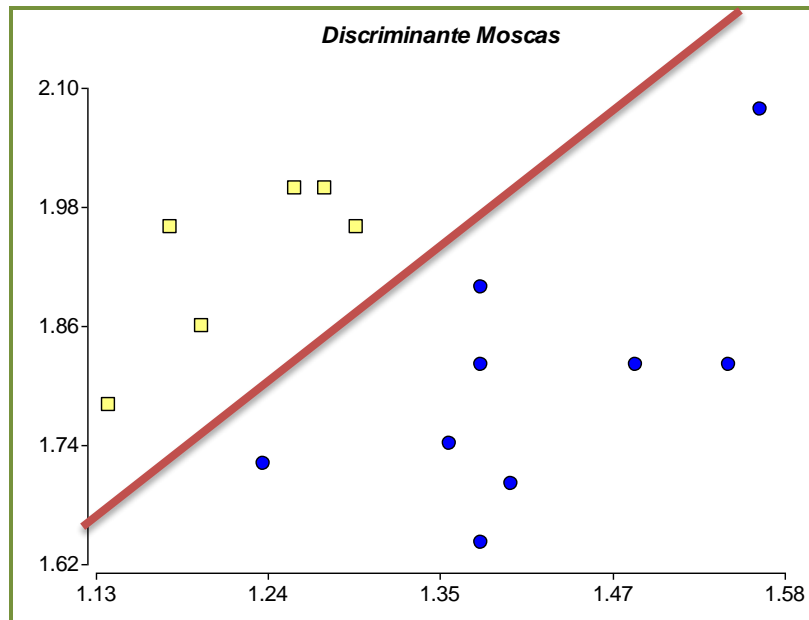
$$x = (1.3, 1.8)$$

Inicialmente graficamos las observaciones en un color cada una de las poblaciones. ( en este caso al tener solo dos variables es sencillo hacerlo, esto para más variables no es posible!!).

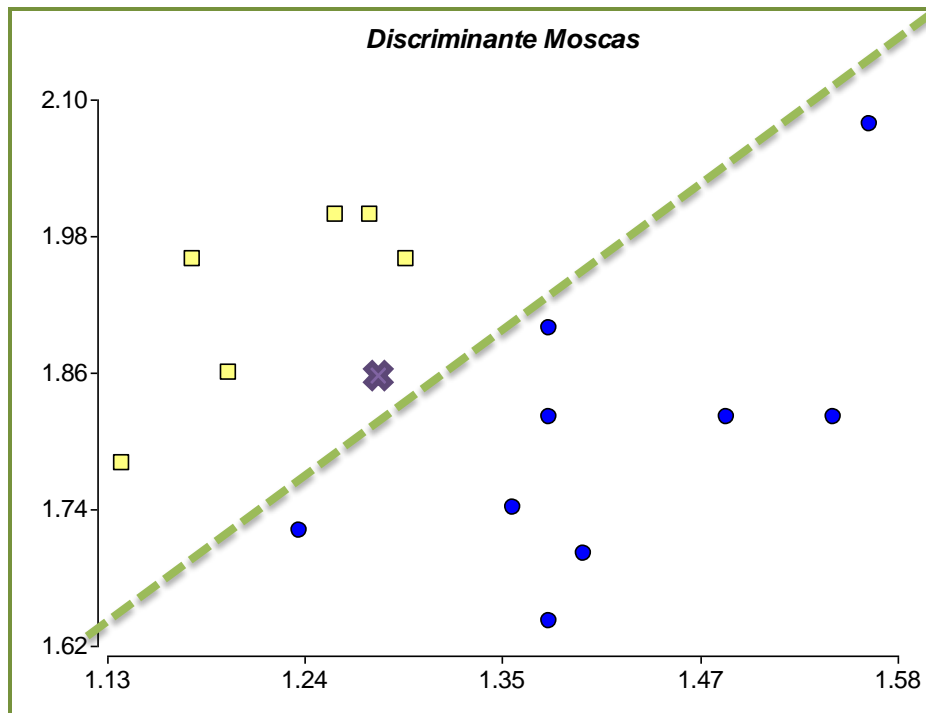


Del gráfico queda claro que estas dos variables permiten discriminar entre las poblaciones.

Donde trazarían una línea para partir entre las poblaciones??



En qué dirección convendría proyectar los datos de modo tal que se discrimine mejor entre las poblaciones? Cuál sería el punto límite de las proyecciones que permitiría clasificarlas?



Estimamos los parámetros de interés.

**Vector medio total**

| X     | Y     | . |
|-------|-------|---|
| 1.339 | 1.853 |   |

**Vectores medios por grupo**

| grupo | X     | Y     | .                       |
|-------|-------|-------|-------------------------|
| A     | 1.413 | 1.804 | $\rightarrow \bar{X}_A$ |
| B     | 1.227 | 1.927 | $\rightarrow \bar{X}_B$ |

**Matriz de covarianzas común (insesgada)**

|   | X     | Y     | . |
|---|-------|-------|---|
| X | 0.008 | 0.007 |   |
| Y | 0.007 | 0.013 |   |

**Variables** X1 y X2

**Variables de clasificación:** grupo

**Testeamos el supuesto de igualdad de varianzas de los grupos**

**Prueba de Homogeneidad de Matrices de Covarianzas**

| Grupos | N  | Estadístico | gl | p-valor |
|--------|----|-------------|----|---------|
| 2      | 15 | 2.73        | 3  | 0.4359  |

Como no se rechaza la hipótesis de nulidad, suponemos que las varianzas son iguales en ambos grupos.

### Testeamos el supuesto de normalidad

#### Shapiro-Wilks (modificado)

| Variable | n  | Media | D.E. | W*   | p (Unilateral D) |
|----------|----|-------|------|------|------------------|
| X1       | 15 | 1.34  | 0.13 | 0.94 | 0.5390           |
| X2       | 15 | 1.85  | 0.13 | 0.95 | 0.6940           |

Como no se rechazan las hipótesis de normalidad de las dos variables asumimos normalidad multivariada, si bien esto debería probarse en forma más rigurosa no lo haremos en este ejemplo.

### Calculamos la función discriminante lineal

#### Funciones discriminantes canónicas

|           |       |   |
|-----------|-------|---|
|           | 1     | . |
| Constante | -1.88 |   |
| X1        | 14.78 |   |
| X2        | -9.66 |   |

$$D_i = -1.88 + 14.78 \cdot X - 9.66 \cdot Y$$

| X1   | X2   | grupo | Eje Canónico 1 | Clasificación |
|------|------|-------|----------------|---------------|
| 1.38 | 1.64 | A     | 2.67           | A             |
| 1.4  | 1.7  | A     | 2.39           | A             |
| 1.24 | 1.72 | A     | -0.17          | A             |
| 1.36 | 1.74 | A     | 1.41           | A             |
| 1.38 | 1.82 | A     | 0.93           | A             |
| 1.48 | 1.82 | A     | 2.41           | A             |
| 1.54 | 1.82 | A     | 3.3            | A             |
| 1.38 | 1.9  | A     | 0.16           | A             |
| 1.56 | 2.08 | A     | 1.08           | A             |
| 1.14 | 1.78 | B     | -2.23          | B             |
| 1.2  | 1.86 | B     | -2.11          | B             |
| 1.18 | 1.96 | B     | -3.38          | B             |
| 1.3  | 1.96 | B     | -1.6           | B             |
| 1.26 | 2    | B     | -2.58          | B             |
| 1.28 | 2    | B     | -2.28          | B             |

Calculamos la proyección del vector de medias de cada uno de los grupos sobre el espacio discriminante:

$$D_{\bar{X}_A} = -1.88 + 14.78*1.413 - 9.66*1.804 = 1.5777$$

$$D_{\bar{X}_B} = -1.88 + 14.78*1.227 - 9.66*1.927 = -2.3597$$

El programa nos informa estos valores como los centroides!!!

#### Centroides en el espacio discriminante

| Grupo | Eje 1 |
|-------|-------|
| A     | 1.58  |
| B     | -2.36 |

**Cómo clasificamos a nuestra nueva observación?**

**Cuál es el punto medio entre los dos centroides?**

$$(C_a + C_b) / 2 = (1.58 - 2.36) / 2 = -0.39$$

Entonces si una observación tiene su proyección sobre la coordenada discriminante superior a -0.39 se la clasifica en el grupo A, caso contrario se la clasifica en el grupo B.

Para nuestro ejemplo:

$$D(\text{nueva observación}) = -1.88 + 14.78*1.3 - 9.66*1.8 = \boxed{-0.054}$$

-0.054 > -0.39 luego la clasificamos en el grupo A.

Esta solución tal como la hemos presentado, corresponde al **Análisis Discriminante Lineal de Fisher**.

Que es válido solamente bajo el cumplimiento de dos **supuestos**:

- a- Normalidad en la distribución de ambos grupos
- b- Homocedasticidad entre los grupos. (igualdad de matrices de varianzas y covarianzas) mediante su función discriminante lineal:

$$D = a_1X_1 + a_2X_2 + a_3X_3 + \dots + a_pX_p.$$

Se llaman **puntuaciones discriminantes** a los valores que se obtienen al evaluar la función D para un individuo en la ecuación anterior.

La función discriminante puede expresarse matricialmente de la siguiente forma:

$$\begin{pmatrix} D_1 \\ D_2 \\ \dots \\ D_N \end{pmatrix} = \begin{pmatrix} X_{11} & X_{21} & \dots & X_{p1} \\ X_{12} & X_{22} & \dots & X_{p2} \\ \dots & \dots & \dots & \dots \\ X_{1N} & X_{2N} & \dots & X_{pN} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \dots \\ a_p \end{pmatrix}$$

### Estimación de las probabilidades de clasificación errónea

Después de construir la regla, interesa conocer su capacidad para discriminar. Si no tenemos regla alguna, tiramos la moneda y asignamos a un grupo a los sujetos al azar, la capacidad de discriminar correctamente seria del 50%.

Obviamente preferiríamos que la regla no se equivoque en ningún caso o bien que clasifique correctamente en el 95% de los casos, pero esto no siempre es posible de lograr.

**Matriz de confusión:** da una idea de la tasa de clasificaciones incorrectas. Como se sabe el grupo al que pertenece cada objeto, se puede comprobar la efectividad del método de clasificación usando la máxima probabilidad a posteriori, cuando se observa el porcentaje de casos bien clasificados.

Para estimar la probabilidad de clasificación correcta disponemos de tres alternativas:

### **1- Clasificación ingenua**

Utilizar los mismos datos con los que se construyó la regla para estimar la probabilidad de clasificación correcta.

Es decir calcular la proporción de observaciones bien clasificadas con la regla construida a partir de ellas mismas.

### **2- Muestra de entrenamiento y de validación**

Se parte el conjunto de datos disponibles en dos submuestras. (Por ejemplo  $2/3$  y  $1/3$ , elegidos al azar) con la mayor de las submuestras(training sample o muestra de entrenamiento) se construye la regla de clasificación y con la menor de las submuestras( muestra de validación) se estima la probabilidad de buena clasificación.

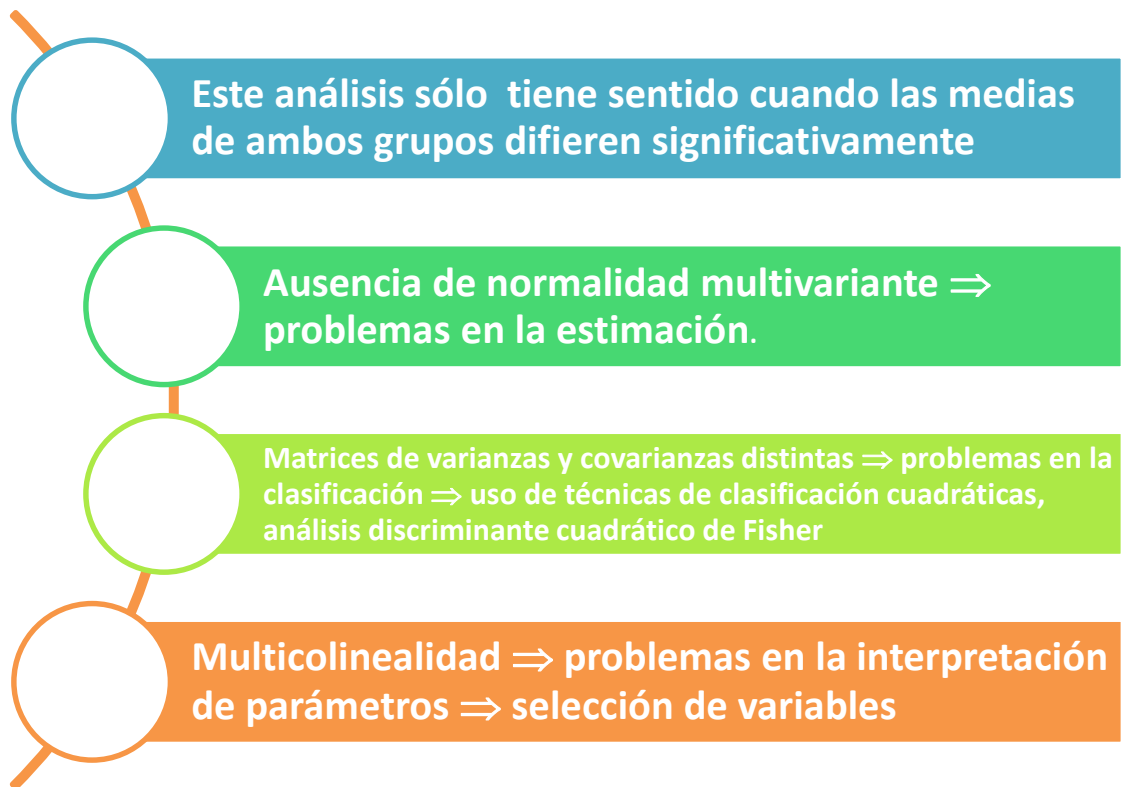
### **3- Validación cruzada (cross validation - leave one out)**

Se elimina la primer observación, se construye la regla sin ella y se la clasifica a esta con dicha regla, luego se reincorpora la primera observación y se elimina la segunda...se continua de esta manera hasta la última observación.

Luego se estima la probabilidad de buena clasificación considerando la proporción de observación bien clasificadas de esta manera.



## Observaciones



Cuando hay  $g > 2$  de dos grupos bajo supuesto de homocedasticidad:

### **Primera propuesta**

Se calcula la distancia de Mahalanobis al centroide(media) de cada grupo y un nuevo individuo se clasifica en el grupo  $i$  si el valor de la distancia de Mahalanobis a su grupo es la menor de todas.

### **Segunda propuesta**

Calcular la probabilidad a posteriori de la nueva observación(o la función de verosimilitud) para cada una de los grupos.

**Clasificar a la observación en el grupo que maximice dicha probabilidad.**

### Validación de los supuestos paramétricos del análisis discriminante

El objetivo del análisis discriminante **es minimizar la probabilidad de equivocarse al clasificar los individuos en cada grupo.**

Sin embargo el análisis solo es válido cuando las variables originales se distribuyen con normal multivariante y las matrices de covarianzas son iguales en todos los grupos.

Cuando la distribución conjunta es normal multivariante, entonces cualquier combinación lineal de ellas se distribuye como una normal univariante. En particular esto vale para cada una de las coordenadas del vector de observaciones. Por ello, si alguna de las variables originales no se distribuye como una normal, entonces es seguro que todas las variables conjuntamente no se distribuirán como una normal multivariante.

Para comprobar el supuesto de homocedasticidad, es decir igualdad de las matrices de varianzas y covarianza, se puede aplicar la prueba **M de Box**.

La hipótesis nula de la prueba es que las matrices de covarianzas de los grupos son iguales.

Para comparar las matrices de covarianzas se utiliza el determinante de la matriz de covarianza de cada uno de los grupos.

Como el test de **Bartlett** en el caso univariado, el test M de Box es sensible a la falta de normalidad multivariante, es decir, matrices iguales pueden

aparecer como significativamente diferentes por la falta de cumplimiento del supuesto de normalidad.

Por otra parte, si las muestras son grandes, pierde efectividad (es más fácil rechazar la hipótesis nula).

### Que podemos hacer cuando se rechaza la homocedasticidad?

Cuando se rechaza la hipótesis nula  $H_0: \Sigma_1 = \Sigma_2$ , una alternativa es calcular la misma dirección discriminante asumiendo matrices iguales pero calculando la distancia estandarizada de la puntuación discriminante del dato que queremos clasificar a los centroides.

En este caso se proyectan todos los datos y se estima la varianza de las puntuaciones discriminantes dentro de cada grupo.

Otra opción es utilizar análisis discriminante cuadrático de Fisher que construye la regla:

$x \in G_1$  cuando:

$$(2\pi)^{-\frac{p}{2}} |\Sigma_1|^{-1/2} \exp\left(-\frac{(x-\mu_1)'\Sigma_1^{-1}(x-\mu_1)}{2}\right) > (2\pi)^{-\frac{p}{2}} |\Sigma_2|^{-1/2} \exp\left(-\frac{(x-\mu_2)'\Sigma_2^{-1}(x-\mu_2)}{2}\right)$$

Considerando  $\Sigma_1 \neq \Sigma_2$

### **Algunas consideraciones sobre la selección de las variables para introducir al modelo:**

Primero se puede realizar un análisis descriptivo univariado calculando las medias y las desviaciones estándar de las variables originales para cada uno de los grupos por separado. Si para alguna variable las medias de los grupos son diferentes y la variabilidad es pequeña, se considera que dicha variable será importante a la hora de discriminar a los grupos.

Luego se testea la igualdad o no de los vectores medios de los grupos para las variables seleccionadas.

También resulta útil comparar las matrices de correlaciones por ser más fácilmente interpretables que las de covarianzas. Se estiman estas matrices para el grupo general y para cada uno de los subgrupos.

### **Interpretación de los coeficientes de la función discriminante**

Si usamos **variables originales tipificadas**, se obtienen los coeficientes  $a_{ij}$  que relacionan las variables con las funciones discriminantes:

Se pueden interpretar las magnitudes de los coeficientes como indicadores de la importancia relativa de las variables en cada función discriminante. Así, si  **$a_{ij}$  es grande en valor absoluto**, entonces hay una fuerte asociación entre la variable  $x_j$  y la función  $y_i$ , en relación al resto de variables.

De cualquier manera, al existir en general correlaciones significativas entre las variables originales, se debe tener cuidado al hacer interpretaciones precipitadas.

### Costos de clasificación

En ocasiones es necesario ponderar los errores. Por ejemplo, es más costoso no indicar que un paciente tiene rechazo a un órgano trasplantado, cuando efectivamente lo tiene, que indicar que si lo tiene cuando en realidad esto no es así.

Una manera de diferenciar en la regla discriminante entre los dos tipos de error posibles es asignar un costo a cada error.

También podría utilizarse información previa, como por ejemplo si uno sabe que el rechazo ocurre en a lo sumo el 20% de los pacientes.

Supongamos que tenemos dos poblaciones:

- $\pi_1$  con función de densidad  $f(x, \theta_1)$  donde  $\theta_1$  es un vector de parámetros que caracteriza a la densidad de la primera población.
- $\pi_2$  con función de densidad  $f(x, \theta_2)$  donde  $\theta_2$  es un vector de parámetros que caracteriza a la densidad de la segunda población.

Una regla discriminante general, partirá al espacio p-dimensional en dos partes  $R_1$  y  $R_2$  de modo tal que si una observación cae en  $R_1$  será clasificada en el primer grupo y en caso contrario será clasificada en el segundo grupo.

Llamemos  $C(i/j)$  al costo de clasificar a un individuo en la población  $i$  cuando en realidad pertenece a la población  $j$  y  $P(i/j)$  es la probabilidad de clasificar a un individuo en la población  $i$  cuando pertenece en realidad a la población  $j$ .

También designemos con  $p_i$  a la probabilidad de que un individuo de la población general pertenezca al  $i$ -simo grupo.

Cuál sería el costo promedio de clasificación errónea de una observación seleccionada de forma aleatoria de la población??

$$p_1 C(2/1) P(2/1) + p_2 C(1/2) P(1/2)$$

Utilizando la regla de Bayes:

$$X \in \pi_1$$

$$\text{si} \quad p_1 C(2/1) f(x, \theta_2) < p_2 C(1/2) f(x, \theta_1)$$

Si las probabilidades de pertenencia a ambos grupos fueran iguales:

$$\text{Si} \quad C(2/1) f(x, \theta_2) < C(1/2) f(x, \theta_1)$$

Si los costos de mala clasificación para ambos grupos fueran iguales:

$$X \in \pi_1$$

$$\text{si} \quad p_1 f(x, \theta_2) < p_2 f(x, \theta_1)$$

Si los costos de mala clasificación y las probabilidades a priori para ambos grupos fueran iguales:

$$X \in \pi_1$$

$$\text{si} \quad f(x, \theta_2) < f(x, \theta_1)$$

Es decir se reduce a maximizar la función de verosimilitud.

### **Ejemplo 6**

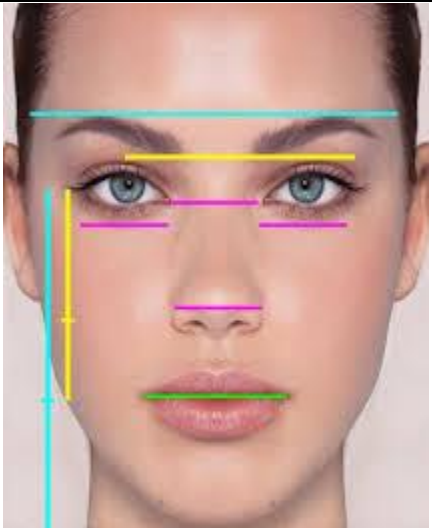
Consideramos una muestra de 32 cráneos pertenecientes a dos regiones, 17 de una de las regiones y 15 de la otra. (Ver archivo Cráneos.xls).

Sobre estos 32 ejemplares se midieron las variables:

**Longitud,  
Anchura, Altura del cráneo y  
Altura y Anchura de la cara**

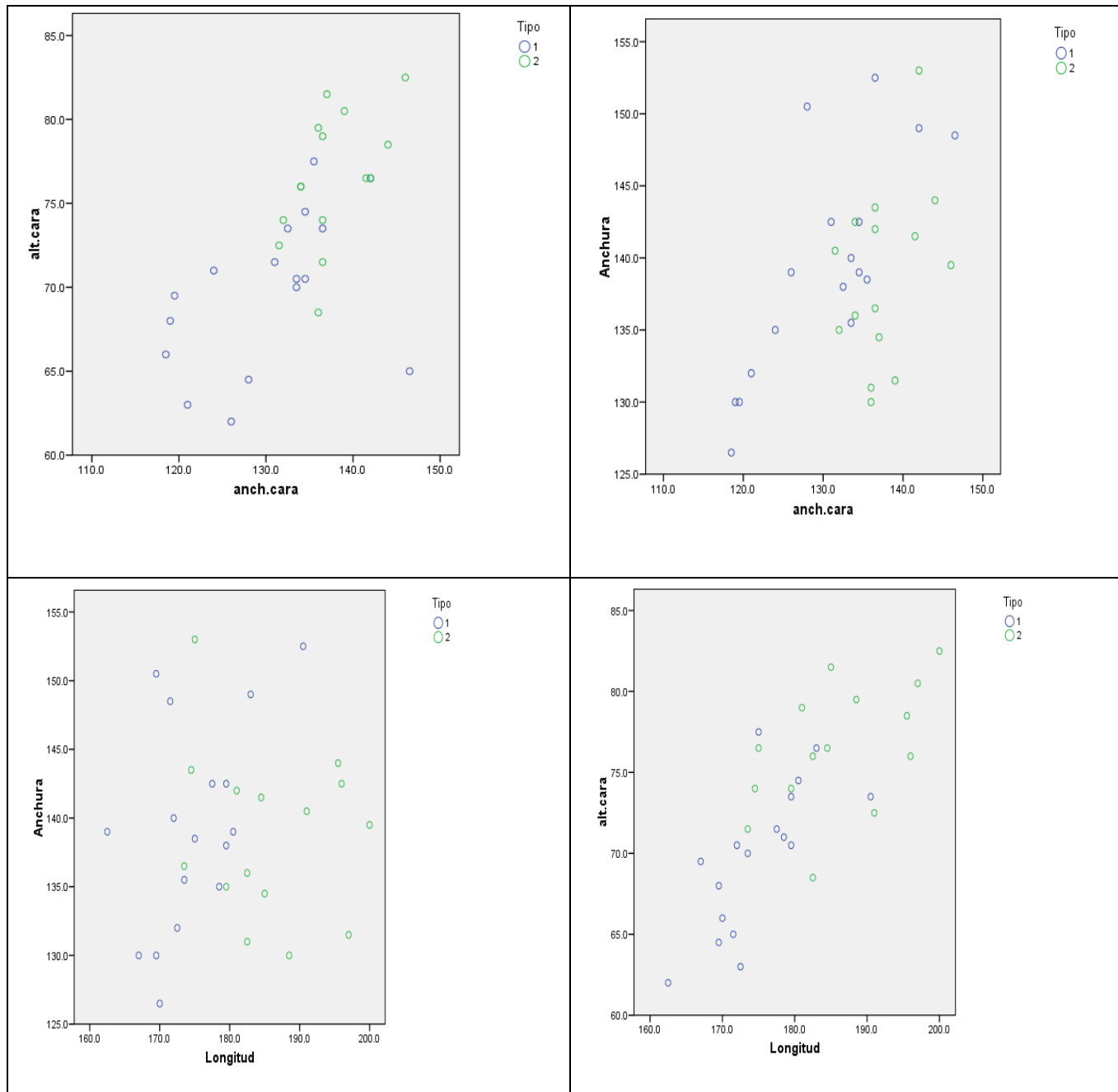
Los resúmenes (univariados) para cada una de las variables se presentan en la siguiente tabla:

| Estadísticos de grupo |           |         |            |                        |            |
|-----------------------|-----------|---------|------------|------------------------|------------|
| Tipo                  |           | Media   | Desv. típ. | N válido (según lista) |            |
|                       |           |         |            | No ponderados          | Ponderados |
| 1                     | Longitud  | 174.824 | 6.7475     | 17                     | 17.000     |
|                       | Anchura   | 139.353 | 7.6030     | 17                     | 17.000     |
|                       | Altura    | 132.000 | 6.0078     | 17                     | 17.000     |
|                       | alt.cara  | 69.824  | 4.5756     | 17                     | 17.000     |
|                       | anch.cara | 130.353 | 8.1370     | 17                     | 17.000     |
| 2                     | Longitud  | 185.733 | 8.6269     | 15                     | 15.000     |
|                       | Anchura   | 138.733 | 6.1117     | 15                     | 15.000     |
|                       | Altura    | 134.767 | 6.0263     | 15                     | 15.000     |
|                       | alt.cara  | 76.467  | 3.9118     | 15                     | 15.000     |
|                       | anch.cara | 137.500 | 4.2384     | 15                     | 15.000     |
| Total                 | Longitud  | 179.938 | 9.3651     | 32                     | 32.000     |
|                       | Anchura   | 139.063 | 6.8412     | 32                     | 32.000     |
|                       | Altura    | 133.297 | 6.0826     | 32                     | 32.000     |
|                       | alt.cara  | 72.938  | 5.3908     | 32                     | 32.000     |
|                       | anch.cara | 133.703 | 7.4443     | 32                     | 32.000     |



Notamos diferencias entre los valores medios de estas variables pero no nos queda claro si las diferencias que notamos son o no significativas. Graficamos pares de variables para ver si alguno de estos pares permite discriminar mejor que otros.





Comparamos los vectores de medias de ambos grupos, para ello en el InfoStat, aplicamos análisis de la varianza multivariado, que correspondería para más de dos grupos, pero como caso especial también es adecuado para dos grupos:

#### Cuadro de Análisis de la Varianza (Lawley-Hotelling)

| F.V. | Estadístico | F    | gl (num) | gl (den) | p      |
|------|-------------|------|----------|----------|--------|
| Tipo | 0.93        | 4.84 | 5        | 26       | 0.0029 |

Se rechaza la hipótesis de igualdad de medias entre los grupos, por lo tanto, se puede inferir que las medias de los grupos son diferentes. En virtud de lo cual **tiene sentido realizar el análisis discriminante**.

Si realizamos las pruebas para contrastar las medias en forma univariada, encontramos que no todas son significativas.

#### Pruebas de igualdad de las medias de los grupos

|           | Lambda de Wilks | F      | gl1 | gl2 | Sig. |
|-----------|-----------------|--------|-----|-----|------|
| Longitud  | .651            | 16.072 | 1   | 30  | .000 |
| Anchura   | .998            | .063   | 1   | 30  | .803 |
| Altura    | .947            | 1.685  | 1   | 30  | .204 |
| alt.cara  | .610            | 19.210 | 1   | 30  | .000 |
| anch.cara | .763            | 9.315  | 1   | 30  | .005 |

Antes de realizar el análisis discriminante, podemos testear si las matrices de covarianza de los grupos son similares, si es así, utilizaremos análisis discriminante lineal, en caso contrario utilizaremos análisis discriminante cuadrático o bien transformaremos las variables.

### Prueba de Box sobre la igualdad de las matrices de covarianza

#### Logaritmo de los determinantes

| Tipo                   | Rango | Logaritmo del determinante |
|------------------------|-------|----------------------------|
| 1                      | 5     | 16.164                     |
| 2                      | 5     | 15.773                     |
| Intra-grupos combinada | 5     | 16.727                     |

Los rangos y logaritmos naturales de los determinantes impresos son los de las matrices de covarianzas de los grupos.

#### Resultados de la prueba

|          |              |
|----------|--------------|
| M de Box | 22.371       |
| F        | Aprox. 1.218 |
|          | gl1 15       |
|          | gl2 3489.901 |
| Sig.     | .249         |

Contrasta la hipótesis nula de que las matrices de covarianzas poblacionales son iguales.

No se rechaza la hipótesis de homocedasticidad, luego aplicamos análisis discriminante lineal de Fisher.

**Coefficientes estandarizados de las funciones discriminantes canónicas (para datos estandarizados con matriz de covarianza común)**

|           | Función |
|-----------|---------|
|           | 1       |
| Longitud  | .367    |
| Anchura   | -.578   |
| Altura    | -.017   |
| alt.cara  | .405    |
| anch.cara | .627    |

**Funciones en los centroides de los grupos**

|      | Función |
|------|---------|
| Tipo | 1       |
| 1    | -.877   |
| 2    | .994    |

Funciones discriminantes canónicas no tipificadas evaluadas en las medias de los grupos

**Probabilidades previas para los grupos**

| Tipo  | Previas | Casos utilizados en el análisis |               |
|-------|---------|---------------------------------|---------------|
|       |         | Ponderados                      | No ponderados |
| 1     | .531    | 17                              | 17.000        |
| 2     | .469    | 15                              | 15.000        |
| Total | 1.000   | 32                              | 32.000        |

**Coefficientes de la función de clasificación**

|             | Tipo     |          |
|-------------|----------|----------|
|             | 1        | 2        |
| Longitud    | 1.468    | 1.558    |
| Anchura     | 2.361    | 2.205    |
| Altura      | 2.752    | 2.747    |
| alt.cara    | .775     | .952     |
| anch.cara   | .195     | .372     |
| (Constante) | -514.895 | -545.484 |

Funciones discriminantes lineales de Fisher

**Tabla de clasificación cruzada (tasa de error aparente)**

| Grupo | 1  | 2  | Total | Error(%) |
|-------|----|----|-------|----------|
| 1     | 14 | 3  | 17    | 17.65    |
| 2     | 3  | 12 | 15    | 20       |
| Total | 17 | 15 | 32    | 18.75    |

Existen otras alternativas para la clasificación:

- ✓ **Análisis Discriminante lineal de Fisher (LDA)**
- ✓ **Análisis Discriminante cuadrático de Fisher (QDA)**
- ✓ **Regresión logística**
- ✓ **K vecinos más cercanos (KNN)**
- ✓ **Árboles de clasificación (CART)**