Coordenadas Discriminantes

Graciela Boente

Ejemplo

Variables medidas sobre árboles de manzana de 6 injertos. Para cada injerto hay 8 árboles. Las variables son:
x_1 =Diámetro del tronco a los 4 años en unidades de 10cm,
x_2 =Largo a los 4 años,
x_3 =Diámetro del tronco a los 15 años en unidades de 10cm,
x_4 =Peso del árbol a los 15 años, en unidades de 1000 libras.

Inj.	1	1	1	1	1	1	1	1	2	2	2	2
x_1	1.11	1.19	1.09	1.25	1.11	1.08	1.11	1.16	1.05	1.17	1.11	1.25
x_2	2.569	2.928	2.865	3.844	3.027	2.336	3.211	3.037	2.074	2.885	3.378	3.906
x_3	3.58	3.75	3.93	3.94	3.60	3.51	3.98	3.62	4.09	4.06	4.87	4.98
x_4	0.760	0.821	0.928	1.009	0.766	0.726	1.209	0.750	1.036	1.094	1.635	1.517
Inj.	2	2	2	2	3	3	3	3	3	3	3	3
x_1	1.17	1.15	1.17	1.19	1.07	0.99	1.06	1.02	1.15	1.20	1.20	1.17
x_2	2.782	3.018	3.383	3.447	2.505	2.315	2.667	2.390	3.021	3.085	3.308	3.231
x_3	4.38	4.65	4.69	4.40	3.76	4.44	4.38	4.67	4.48	4.78	4.57	4.56
x_4	1.197	1.244	1.495	1.026	0.912	1.398	1.197	1.613	1.476	1.571	1.506	1.458
Inj.	4	4	4	4	4	4	4	4	5	5	5	5
x_1	1.22	1.03	1.14	1.01	0.99	1.11	1.20	1.08	0.91	1.15	1.14	1.05
x_2	2.838	2.351	3.001	2.439	2.199	3.318	3.601	3.291	1.532	2.552	3.083	2.330
x_3	3.89	4.05	4.05	3.92	3.27	3.95	4.27	3.85	4.04	4.16	4.79	4.42
x_4	0.944	1.241	1.023	1.067	0.693	1.085	1.242	1.017	1.084	1.151	1.381	1.242
Inj.	5	5	5	5	6	6	6	6	6	6	6	6
x_1	0.99	1.22	1.05	1.13	1.11	0.75	1.05	1.02	1.05	1.07	1.13	1.11
x_2	2.079	3.366	2.416	3.100	2.813	0.840	2.199	2.132	1.949	2.251	3.064	2.469
x_3	3.47	4.41	4.64	4.57	3.76	3.14	3.75	3.99	3.34	3.21	3.63	3.95
x_4	0.673	1.137	1.455	1.325	0.800	0.606	0.790	0.853	0.610	0.562	0.707	0.952

Ejemplo

Hemos estudiado si las medias de los Injertos 1, 2 y 3 eran iguales, o sea, testeamos $H_3 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \boldsymbol{\mu}_3$. Para ello supusimos que las matrices de covarianza eran iguales. Teníamos que

$\bar{\mathbf{x}}_1 = (1.1375, 2.9771, 3.7388, 0.8711)^{\mathrm{T}}$

$\bar{\mathbf{x}}_2 = (1.1575, 3.1091, 4.5150, 1.2805)^{\mathrm{T}}$

$\bar{\mathbf{x}}_3 = (1.1075, 2.8152, 4.4550, 1.3914)^{\mathrm{T}}$

$\bar{\mathbf{x}} = (1.1342, 2.9672, 4.2363, 1.1810)^{\mathrm{T}}$

Las matrices de covarianza estimadas son

$$\mathbf{S}_1 = \begin{pmatrix} 0.0034 & 0.0203 & 0.0037 & 0.0018 \\ 0.0203 & 0.2007 & 0.0580 & 0.0458 \\ 0.0037 & 0.0580 & 0.0352 & 0.0285 \\ 0.0018 & 0.0458 & 0.0285 & 0.0283 \end{pmatrix} \quad \mathbf{S}_2 = \begin{pmatrix} 0.0034 & 0.0258 & 0.0088 & 0.0032 \\ 0.0258 & 0.3048 & 0.1498 & 0.0832 \\ 0.0088 & 0.1498 & 0.1157 & 0.0711 \\ 0.0032 & 0.0832 & 0.0711 & 0.0565 \end{pmatrix}$$

$$\mathbf{S}_3 = \begin{pmatrix} 0.0068 & 0.0314 & 0.0087 & 0.0060 \\ 0.0314 & 0.1543 & 0.0480 & 0.0329 \\ 0.0087 & 0.0480 & 0.0951 & 0.0680 \\ 0.0060 & 0.0329 & 0.0680 & 0.0534 \end{pmatrix}$$

Ejemplo

Construímos $\mathbf{U} = \mathbf{Q}_1 + \mathbf{Q}_2 + \mathbf{Q}_3 = \sum_{i=1}^{3} \sum_{j=1}^{n_i} (\mathbf{x}_{i,j} - \overline{\mathbf{x}}_i)(\mathbf{x}_{i,j} - \overline{\mathbf{x}}_i)^{\mathrm{T}}$ y $\mathbf{H} = \sum_{i=1}^{k} n_i (\overline{\mathbf{x}}_i - \overline{\mathbf{x}})(\overline{\mathbf{x}}_i - \overline{\mathbf{x}})^{\mathrm{T}}$ obteniendo

$$\mathbf{U} = \begin{pmatrix} 0.0956 & 0.5422 & 0.1490 & 0.0771 \\ 0.5422 & 4.6184 & 1.7911 & 1.1326 \\ 0.1490 & 1.7911 & 1.7221 & 1.1731 \\ 0.0771 & 1.1326 & 1.1731 & 0.9670 \end{pmatrix} \quad \mathbf{H} = \begin{pmatrix} 0.0101 & 0.0592 & -0.0079 & -0.0346 \\ 0.0592 & 0.3466 & 0.0111 & -0.1674 \\ -0.0079 & 0.0111 & 2.9845 & 1.8233 \\ -0.0346 & -0.1674 & 1.8233 & 1.2014 \end{pmatrix}$$

El estadístico del test era

$$V = \frac{|\mathbf{U}|}{|\mathbf{U} + \mathbf{H}|} = \Lambda(23, 4, 2)$$

y rechazamos H_3.

Propiedad

Sea $\mathbf{x} \in \mathbb{R}^p$ un vector aleatorio y G una variable aleatoria que indica la pertencia al grupo, tales que para $1 \leq j \leq k$

$$\mathbb{P}(G = j) = \pi_j \qquad \mathbb{E}(\mathbf{x}|G = j) = \boldsymbol{\mu}_j \qquad \mathrm{VAR}(\mathbf{x}|G = j) = \boldsymbol{\Sigma}_j$$

entonces si $\overline{\boldsymbol{\mu}} = \mathbb{E}(\mathbf{x})$ y $\overline{\boldsymbol{\Sigma}} = \mathrm{VAR}(\mathbf{x})$ se cumple que

$$\overline{\boldsymbol{\mu}} = \sum_{j=1}^{k} \pi_j \boldsymbol{\mu}_j \qquad \overline{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}_{\mathrm{W}} + \boldsymbol{\Sigma}_{\mathrm{B}}$$

donde

$$\boldsymbol{\Sigma}_{\mathrm{W}} = \sum_{j=1}^{k} \pi_j \boldsymbol{\Sigma}_j \qquad \boldsymbol{\Sigma}_{\mathrm{B}} = \sum_{j=1}^{k} \pi_j (\boldsymbol{\mu}_j - \overline{\boldsymbol{\mu}})(\boldsymbol{\mu}_j - \overline{\boldsymbol{\mu}})^{\mathrm{T}}$$

Por lo tanto, si $\boldsymbol{\Sigma}_j = \boldsymbol{\Sigma}$ para $1 \leq j \leq k$ tenemos que

- $\boldsymbol{\Sigma}_{\mathrm{W}} = \sum_{j=1}^{k} \pi_j \boldsymbol{\Sigma}_j = \boldsymbol{\Sigma}$ mide la variabilidad dentro de grupos

- $\boldsymbol{\Sigma}_{\mathrm{B}}$ mide la variabilidad entre grupos.

Es decir, descompusimos
la variabilidad total como la variabilidad dentro de grupos más la variabilidad entre grupos.

Observemos que

- $\boldsymbol{\Sigma}_{\mathrm{B}}$ es definida no–negativa de rango $s \leq \min(k-1, p)$.
- Bajo H_3, $\boldsymbol{\Sigma}_{\mathrm{B}} = \mathbf{0}$

Luego, tenemos que

$$\overline{\boldsymbol{\Sigma}} \geq \boldsymbol{\Sigma}_{\mathrm{W}} \qquad \text{y bajo } H_3 : \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_k \qquad \overline{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}_{\mathrm{W}}$$

o sea, para realizar un test para H_3 basta comparar cuan distinta es $\overline{\boldsymbol{\Sigma}}$ de $\boldsymbol{\Sigma}_{\mathrm{W}}$.

Por otra parte, si las poblaciones son normales $\widehat{\boldsymbol{\Sigma}}_j = \mathbf{Q}_j / n_j$ es el EMV de $\boldsymbol{\Sigma}_j$ y $\overline{\mathbf{x}}_j$ es el de $\boldsymbol{\mu}_j$. Luego, si $\widehat{\pi}_j = n_j/n$

$$\widehat{\overline{\boldsymbol{\mu}}} = \sum_{j=1}^{k} \widehat{\pi}_j \widehat{\boldsymbol{\mu}}_j \qquad \widehat{\overline{\boldsymbol{\Sigma}}} = \widehat{\boldsymbol{\Sigma}}_{\mathrm{W}} + \widehat{\boldsymbol{\Sigma}}_{\mathrm{B}}$$

son los EMV de $\overline{\boldsymbol{\mu}}$ y $\overline{\boldsymbol{\Sigma}}$, donde

$$\widehat{\boldsymbol{\Sigma}}_{\mathrm{W}} = \sum_{j=1}^{k} \widehat{\pi}_j \widehat{\boldsymbol{\Sigma}}_j \qquad \widehat{\boldsymbol{\Sigma}}_{\mathrm{B}} = \sum_{j=1}^{k} \widehat{\pi}_j (\widehat{\boldsymbol{\mu}}_j - \widehat{\overline{\boldsymbol{\mu}}})(\widehat{\boldsymbol{\mu}}_j - \widehat{\overline{\boldsymbol{\mu}}})^{\mathrm{T}}$$

O sea,

- $\mathbf{U}/n$ es el EMV de $\boldsymbol{\Sigma}_{\mathrm{W}}$,
- $\mathbf{H}/n$ es el EMV de $\boldsymbol{\Sigma}_{\mathrm{B}}$ y
- $(\mathbf{U} + \mathbf{H})/n$ es el EMV de $\overline{\boldsymbol{\Sigma}}$,

lo que explica el estadístico del test para H_3.

Caso de dos poblaciones

En el caso de dos poblaciones, nosotros vimos que la dirección a la que se atribuía la responsabilidad del rechazo era

$$\widehat{\boldsymbol{\alpha}} = \mathbf{S}^{-1}(\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2)$$

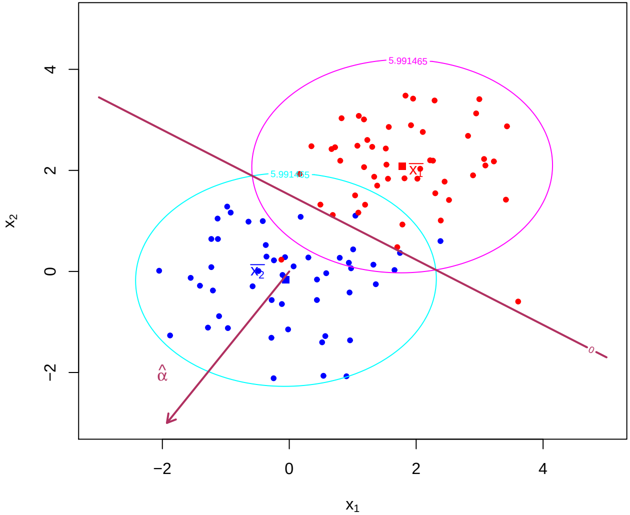
con $\mathbf{S} = ((n_1 - 1)\mathbf{S}_1 + (n_2 - 1)\mathbf{S}_2)/(n_1 + n_2 - 2)$

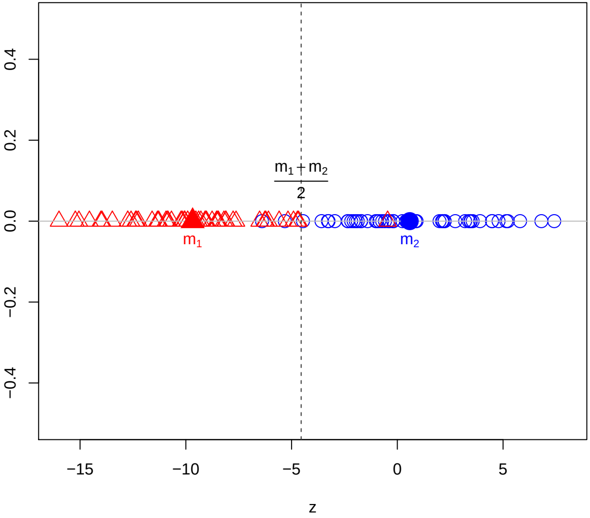
- La función $H(\mathbf{u}) = \widehat{\boldsymbol{\alpha}}^{\mathrm{T}}\mathbf{u}$ es la función discriminante lineal.
- Asignamos $\mathbf{u}$ a la población 1 si

$$z = \widehat{\boldsymbol{\alpha}}^{\mathrm{T}}\mathbf{u} > \widehat{\boldsymbol{\alpha}}^{\mathrm{T}}\left(\frac{\overline{\mathbf{x}}_1 + \overline{\mathbf{x}}_2}{2}\right) = \frac{m_1 + m_2}{2}$$

o equivalentemente, si $z - m_2 > m_1 - z$

Observemos que $m_1 - m_2 > 0$, luego, esto es análogo a clasificar $\mathbf{u}$ en aquella población donde la distancia $|z - m_j|$ sea mínima.


5.991465
5.991465
$\overline{x_1}$
$\overline{x_2}$
$\hat{\alpha}$
0
x_1
x_2
−2
0
2
4
−2
0
2
4


$\frac{m_1+m_2}{2}$
m_1
m_2
0.4
0.2
0.0
−0.2
−0.4
−15
−10
−5
0
5
z

Varias poblaciones: Problema poblacional

Teníamos $\mathbf{x} \in \mathbb{R}^p$ un vector aleatorio y G una variable aleatoria que indica la pertencia al grupo, tales que para $1 \leq i \leq k$

$$\mathbb{P}(G = i) = \pi_i \qquad \mathbb{E}(\mathbf{x}|G = i) = \boldsymbol{\mu}_i \qquad \mathrm{VAR}(\mathbf{x}|G = i) = \boldsymbol{\Sigma}_i$$

entonces si $\overline{\boldsymbol{\mu}} = \mathbb{E}(\mathbf{x})$ y $\overline{\boldsymbol{\Sigma}} = \mathrm{VAR}(\mathbf{x})$ se cumple que

$$\overline{\boldsymbol{\mu}} = \sum_{j=1}^{k} \pi_j \boldsymbol{\mu}_j \qquad \overline{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}_{\mathrm{W}} + \boldsymbol{\Sigma}_{\mathrm{B}}$$

donde

$$\boldsymbol{\Sigma}_{\mathrm{W}} = \sum_{i=1}^{k} \pi_i \boldsymbol{\Sigma}_i \qquad \boldsymbol{\Sigma}_{\mathrm{B}} = \sum_{i=1}^{k} \pi_i (\boldsymbol{\mu}_i - \overline{\boldsymbol{\mu}})(\boldsymbol{\mu}_i - \overline{\boldsymbol{\mu}})^{\mathrm{T}}$$

El Problema

Sea $z = \mathbf{a}^{\mathrm{T}}\mathbf{x}$, luego

$$\mathrm{VAR}(z) = \mathbf{a}^{\mathrm{T}}\mathrm{VAR}(\mathbf{x})\mathbf{a} = \mathbf{a}^{\mathrm{T}}\mathbf{\Sigma}_{\mathrm{W}}\mathbf{a} + \mathbf{a}^{\mathrm{T}}\mathbf{\Sigma}_{\mathrm{B}}\mathbf{a}$$

es decir, descompusimos a la varianza de z en una componente que mide la variabilidad *dentro* de grupos y otra que mide la variablidad *entre* grupos.

Nos interesan combinaciones lineales, o sea, vectores **a** tales que la varianza de z es mucho más grande que la varianza *dentro* de grupos ya que de esto indica que la variabilidad dentro de grupos se ve aumentada por diferencias en posición.

Como $\overline{\mathbf{\Sigma}} = \mathbf{\Sigma}_{\mathrm{W}} + \mathbf{\Sigma}_{\mathrm{B}}$, nos interesarán direcciones **a** que maximizan

$$F_{\mathbf{a}} = \frac{\mathbf{a}^{\mathrm{T}}\mathbf{\Sigma}_{\mathrm{B}}\mathbf{a}}{\mathbf{a}^{\mathrm{T}}\mathbf{\Sigma}_{\mathrm{W}}\mathbf{a}}$$

Caso $k = 2$

Si $k = 2$

$$\boldsymbol{\Sigma}_{\mathrm{B}} = \pi_1 \pi_2 (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^{\mathrm{T}}$$

luego basta maximizar

$$\frac{(\mathbf{a}^{\mathrm{T}}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2))^2}{\mathbf{a}^{\mathrm{T}} \boldsymbol{\Sigma}_{\mathrm{W}} \mathbf{a}}$$

y vimos que el máximo se alcanza en

$$\boldsymbol{\alpha} = \boldsymbol{\Sigma}_{\mathrm{W}}^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$$

que se estima por

$$\widehat{\boldsymbol{\alpha}} = \mathbf{S}^{-1} (\widehat{\boldsymbol{\mu}}_1 - \widehat{\boldsymbol{\mu}}_2)$$

Caso $k > 2$

Supongamos $\mathbf{\Sigma}_{\mathrm{W}} > 0$. Sea $\mathbf{C}$ triangular tal que $\mathbf{\Sigma}_{\mathrm{W}} = \mathbf{C}^{\mathrm{T}}\mathbf{C}$ y definamos

$$\mathbf{B} = (\mathbf{C}^{-1})^{\mathrm{T}}\mathbf{\Sigma}_{\mathrm{B}}\mathbf{C}^{-1}$$

Sean $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_p$ los autovectores de $\mathbf{B}$ asociados a los autovalores $\lambda_1 \geq \cdots \geq \lambda_p$.

$$\mathbf{B} = \sum_{j=1}^{p} \lambda_j \boldsymbol{\beta}_j \boldsymbol{\beta}_j^{\mathrm{T}} \qquad \boldsymbol{\beta}_j^{\mathrm{T}} \boldsymbol{\beta}_\ell = 0 \text{ si } j \neq \ell \quad \|\boldsymbol{\beta}_j\| = 1$$

Como rango$(\mathbf{B}) = s \leq \min(k-1, p)$, $\lambda_j = 0$ si $j > s$. Por simplicidad supondremos que los autovalores no nulos son distintos, o sea, $\lambda_1 > \cdots > \lambda_s$. Es decir, tenemos que

$$\mathbf{B} = \sum_{j=1}^{s} \lambda_j \boldsymbol{\beta}_j \boldsymbol{\beta}_j^{\mathrm{T}} \,, \quad \boldsymbol{\beta}_j^{\mathrm{T}} \boldsymbol{\beta}_\ell = 0 \text{ si } 1 \leq j \neq \ell \leq p \quad \|\boldsymbol{\beta}_j\| = 1 \,, 1 \leq j \leq p$$

Caso $k > 2$

Luego,

$$F_{\mathbf{a}} = \frac{\mathbf{a}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathrm{B}}\mathbf{a}}{\mathbf{a}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathrm{W}}\mathbf{a}} = \frac{(\mathbf{Ca})^{\mathrm{T}}\mathbf{B}(\mathbf{Ca})}{\|\mathbf{Ca}\|^2}$$

Por lo tanto, si $\mathbf{b} = \mathbf{Ca}$

$$\max_{\mathbf{a}\neq\mathbf{0}} F_{\mathbf{a}} = \max_{\mathbf{b}\neq\mathbf{0}} \frac{\mathbf{b}^{\mathrm{T}}\mathbf{B}\mathbf{b}}{\|\mathbf{b}\|^2}$$

El máximo de la expresión del lado derecho se alcanza en $\boldsymbol{\beta}_1$ el autovector de $\mathbf{B}$ asociado a su mayor autovalor λ_1.

Con lo cual,

$$\max_{\mathbf{a}\neq\mathbf{0}} F_{\mathbf{a}} = F_{\boldsymbol{\alpha}_1} = \lambda_1$$

donde

$$\boldsymbol{\alpha}_1 = \mathbf{C}^{-1}\boldsymbol{\beta}_1$$

La combinación lineal $z_1 = \boldsymbol{\alpha}_1^{\mathrm{T}}\mathbf{x}$ se llama la primer **coordenada discriminante** y da la mejor separación entre grupos.

Caso $k > 2$

Tenemos que si rechazamos H_3

$$0 < \text{rango}(\mathbf{B}) = \text{rango}(\boldsymbol{\Sigma}_{\mathrm{B}}) = s \leq \min(k-1, p)\,.$$

Luego, para elegir las siguientes direcciones y no repetir información, buscaremos maximizar $F_{\mathbf{a}}$ sujeto a la condición de que las nuevas coordenadas sean no–correlacionadas con z_1, es decir, tales que

$$\mathrm{Cov}(\mathbf{a}^{\mathrm{T}}\mathbf{x}, \boldsymbol{\alpha}_1^{\mathrm{T}}\mathbf{x}) = \mathbf{a}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathrm{W}}\boldsymbol{\alpha}_1 = 0\,.$$

Este problema se puede escribir como

$$\max_{\mathbf{a}\neq\mathbf{0}\,;\,\mathbf{a}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathrm{W}}\boldsymbol{\alpha}_1=0} F_{\mathbf{a}} = \max_{\mathbf{b}\neq\mathbf{0}\,;\,\mathbf{b}^{\mathrm{T}}\boldsymbol{\beta}_1=0} \frac{\mathbf{b}^{\mathrm{T}}\mathbf{B}\mathbf{b}}{\|\mathbf{b}\|^2}$$

El lado derecho de la expresión se alcanza en $\boldsymbol{\beta}_2$ el autovector de $\mathbf{B}$ asociado al segundo mayor autovalor λ_2.

Caso $k > 2$

$$\max_{\mathbf{a}\neq\mathbf{0}\,;\,\mathbf{a}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathrm{W}}\boldsymbol{\alpha}_1=0} F_{\mathbf{a}} = F_{\boldsymbol{\alpha}_2} = \lambda_2 \qquad \boldsymbol{\alpha}_2 = \mathbf{C}^{-1}\boldsymbol{\beta}_2$$

donde $\boldsymbol{\beta}_2$ el autovector de $\mathbf{B}$ asociado a λ_2.

Definición:

Definimos la $j-$ésima variable canónica o variable discriminante z_j como

$$z_j = \boldsymbol{\alpha}_j^{\mathrm{T}}\mathbf{x} \qquad \textit{donde} \qquad \boldsymbol{\alpha}_j = \mathbf{C}^{-1}\boldsymbol{\beta}_j$$

El vector $\mathbf{z} = (z_1, \ldots, z_p)^{\mathrm{T}}$ es el vector de variables canónicas o variables discriminantes.

$$\mathbf{z} = \mathbf{A}^{\mathrm{T}}\mathbf{x} \qquad \textit{con} \qquad \mathbf{A} = (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_p)$$

Observemos que $\boldsymbol{\alpha}_j$ es un autovector de $\boldsymbol{\Sigma}_{\mathrm{W}}^{-1}\boldsymbol{\Sigma}_{\mathrm{B}}$.

Caso $k > 2$ y $s < p$

Como rango$(\mathbf{B}) = s \leq \min(k-1, p)$ si $s < p$, para todo $s + 1 \leq m \leq p$ tendremos que

$$\max_{\mathbf{a}\neq\mathbf{0}\,;\,\mathbf{a}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathrm{W}}\boldsymbol{\alpha}_\ell=0\,,\,\ell<m} F_{\mathbf{a}} = F_{\boldsymbol{\alpha}_m} = \lambda_m = 0$$

Es decir, una combinación lineal no correlacionada con $z_1, \ldots, z_s$ no dará información sobre las diferencias en posición.

En particular, $z_{s+1}, \ldots, z_p$ no dan información sobre las diferencias de posición.

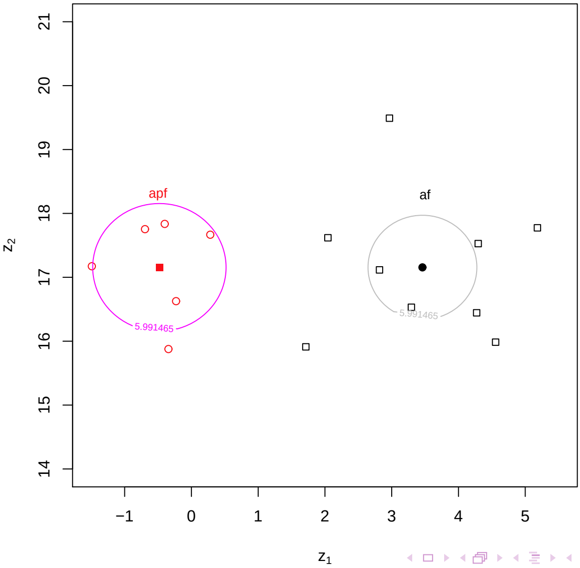
O sea, si $\overline{\boldsymbol{\nu}} = \mathbf{A}^{\mathrm{T}}\overline{\boldsymbol{\mu}}$ y si $\boldsymbol{\nu}_i = \mathbf{A}^{\mathrm{T}}\boldsymbol{\mu}_i$, entonces

$$\boldsymbol{\nu}_{im} = \overline{\boldsymbol{\nu}}_m \qquad \text{para} \quad s + 1 \leq m \leq p$$

Más aún, si $\boldsymbol{\Lambda}_1 = \mathrm{diag}\,(\lambda_1, \ldots, \lambda_s)$

$$\mathbf{A}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathrm{B}}\mathbf{A} = \begin{pmatrix} \boldsymbol{\Lambda}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} = \sum_{j=1}^{k} \pi_j(\boldsymbol{\nu}_j - \overline{\boldsymbol{\nu}})(\boldsymbol{\nu}_j - \overline{\boldsymbol{\nu}})^{\mathrm{T}}$$

Ejemplo



Caso $k > 2$ y $s < p$

Por lo tanto, sólo basta considerar el vector de variables discriminantes

$$\mathbf{z}^{(1)} = (z_1, \ldots, z_s) = \mathbf{A}_1^{\mathrm{T}}\mathbf{x}$$

donde

$$\mathbf{A} = (\mathbf{A}_1, \mathbf{A}_2) \quad \text{con} \quad \mathbf{A}_1 = (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_s)$$

Sean además, $\boldsymbol{\nu}_i^{(1)} = \mathbf{A}_1^{\mathrm{T}}\boldsymbol{\mu}_i$ $\mathbf{z}^{(2)} = \mathbf{A}_2^{\mathrm{T}}\mathbf{x}$, $\mathbf{z} = \begin{pmatrix} \mathbf{z}^{(1)} \\ \mathbf{z}^{(2)} \end{pmatrix}$

Si $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}$, $1 \le i \le k$ entonces $\boldsymbol{\Sigma}_{\mathrm{W}} = \boldsymbol{\Sigma}$ por lo tanto,

$$\mathrm{VAR}(\mathbf{z}|G = i) = \mathbf{I}_p \qquad \mathbb{E}(\mathbf{z}|G = i) = \boldsymbol{\nu}_i$$

Si $\mathbf{x}|G = i \sim N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$ entonces

$$\mathbf{z}|G = i \sim N(\boldsymbol{\nu}_i, \mathbf{I}_p)$$

Caso $k > 2$

En particular

$$\mathbf{z}^{(1)}|G = i \sim N(\boldsymbol{\nu}_i^{(1)}, \mathbf{I}_s) \Longrightarrow \|\mathbf{z}^{(1)} - \boldsymbol{\nu}_i^{(1)}\|^2 \sim \chi_s^2$$

Luego, una manera heurística de definir una región para clasificar una nueva observación es asignar el vector $\mathbf{x}_0$ al grupo i si la media $\boldsymbol{\nu}_i^{(1)}$ de las variables transformadas es la más cercana a $\mathbf{v}_0 = \mathbf{A}_1^{\mathrm{T}}\mathbf{x}_0$.

Asigno $\mathbf{x}_0$ al grupo i si $\mathbf{v}_0 = \mathbf{A}_1^{\mathrm{T}}\mathbf{x}_0 \in \mathcal{G}_i$ donde

$$\begin{aligned} \mathcal{G}_i &= \{\mathbf{v} \in \mathbb{R}^s : \quad \|\mathbf{v} - \boldsymbol{\nu}_i^{(1)}\| < \|\mathbf{v} - \boldsymbol{\nu}_\ell^{(1)}\| \quad \forall \ell \neq i\} \\ &= \{\mathbf{v} \in \mathbb{R}^s : \quad (\boldsymbol{\nu}_i^{(1)})^{\mathrm{T}}(\mathbf{v} - \frac{1}{2}\boldsymbol{\nu}_i^{(1)}) > (\boldsymbol{\nu}_\ell^{(1)})^{\mathrm{T}}(\mathbf{v} - \frac{1}{2}\boldsymbol{\nu}_\ell^{(1)}) \quad \forall \ell \neq i\} \end{aligned}$$

Si $s = 1$ partimos la recta de las observaciones transformadas en dos semi–rectas.

Si $s = 2$ la regla de clasificación genera una partición de $\mathbb{R}^2$ en regiones limitadas por semi–rectas.

Se cortan en el punto equidistante de todos los $\boldsymbol{\nu}_\ell^{(1)}$, $\ell = 1, \ldots, k$.

Inferencia

Supongamos tener ahora observaciones $\mathbf{x}_{ij} \sim N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, $1 \leq j \leq n_i$, $1 \leq i \leq k$, donde

$$\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma} \quad \text{para todo} \quad 1 \leq i \leq k$$

entonces

- $\overline{\mathbf{x}}_i$ es el EMV de $\boldsymbol{\mu}_i$.
- El EMV de $\boldsymbol{\Sigma}_{\mathrm{W}} = \boldsymbol{\Sigma}$ es
$$\frac{\mathbf{U}}{n}$$
- El EMV de $\boldsymbol{\Sigma}_{\mathrm{B}}$ es
$$\frac{\mathbf{H}}{n}$$
y $\mathbb{P}(\text{rango}(\mathbf{H}) = \min(p, k-1)) = 1$
- Un estimador insesgado de $\boldsymbol{\Sigma}_{\mathrm{W}}$ es
$$\frac{\mathbf{U}}{n-k}$$

Inferencia

- $\mathbf{U} = \mathbf{T}^{\mathrm{T}}\mathbf{T}$

- $\mathbf{b}_j$ los autovectores de $\widehat{\mathbf{B}} = (\mathbf{T}^{-1})^{\mathrm{T}}\mathbf{H}\mathbf{T}^{-1}$ tales que $\|\mathbf{b}_j\| = 1$, $\mathbf{b}_j^{\mathrm{T}}\mathbf{b}_\ell = 0$ si $\ell \neq j$

  $\mathbf{b}_j$ es el autovector asociado al $j-$ésimo autovalor $\widehat{\lambda}_j$ de $\widehat{\mathbf{B}}$, donde

$$\mathbb{P}(\widehat{\lambda}_1 > \widehat{\lambda}_2 > \cdots > \widehat{\lambda}_s > 0 \quad \text{y} \quad s = \min(p, k-1)) = 1$$

- $\mathbf{a}_j = \mathbf{T}^{-1}\mathbf{b}_j\sqrt{n-k}$, $\widehat{\mathbf{A}}_1 = (\mathbf{a}_1, \ldots, \mathbf{a}_s)$

- $\overline{\mathbf{z}}_i^{(1)} = \widehat{\mathbf{A}}_1^{\mathrm{T}}\overline{\mathbf{x}}_i$. Entonces, si n_i es grande $\overline{\mathbf{z}}_i^{(1)} \approx N(\boldsymbol{\nu}_i^{(1)}, \mathbf{I}_s/n_i)$.

Inferencia

Una región de confianza asintótica de nivel $1-\alpha$ para $\boldsymbol{\nu}_i^{(1)}$ está dada por

$$\{\mathbf{v} \in \mathbb{R}^s : \quad n_i \|\overline{\mathbf{z}}_i^{(1)} - \mathbf{v}\|^2 \leq \chi^2_{s,\alpha}\} = \{\mathbf{v} \in \mathbb{R}^s : \quad \|\overline{\mathbf{z}}_i^{(1)} - \mathbf{v}\| \leq \sqrt{\frac{\chi^2_{s,\alpha}}{n_i}}\}$$

Además, asignamos $\mathbf{x}_0$ al grupo i si $\widehat{\mathbf{v}}_0 = \widehat{\mathbf{A}}_1^{\mathrm{T}} \mathbf{x}_0 \in \widehat{\mathcal{G}}_i$ donde

$$\begin{aligned}
\widehat{\mathcal{G}}_i &= \{\mathbf{v} \in \mathbb{R}^s : \quad \|\mathbf{v} - \overline{\mathbf{z}}_i^{(1)}\| < \|\mathbf{v} - \overline{\mathbf{z}}_\ell^{(1)}\| \quad \forall \ell \neq i\} \\
&= \{\mathbf{v} \in \mathbb{R}^s : \quad (\overline{\mathbf{z}}_i^{(1)})^{\mathrm{T}}(\mathbf{v} - \frac{1}{2}\overline{\mathbf{z}}_i^{(1)}) > (\overline{\mathbf{z}}_\ell^{(1)})^{\mathrm{T}}(\mathbf{v} - \frac{1}{2}\overline{\mathbf{z}}_\ell^{(1)}) \quad \forall \ell \neq i\} \\
&= \{\mathbf{v} \in \mathbb{R}^s : \quad (\overline{\mathbf{z}}_i^{(1)} - \overline{\mathbf{z}}_\ell^{(1)})^{\mathrm{T}} \left[\mathbf{v} - \frac{\overline{\mathbf{z}}_i^{(1)} + \overline{\mathbf{z}}_\ell^{(1)}}{2}\right] > 0 \quad \forall \ell \neq i\}
\end{aligned}$$

Inferencia

Si las poblaciones no son balanceadas, asignamos $\mathbf{x}_0$ al grupo i si $\widehat{\mathbf{v}}_0 = \widehat{\mathbf{A}}_1^{\mathrm{T}} \mathbf{x}_0 \in \widehat{\mathcal{G}}_i$ donde

$$\widehat{\mathcal{G}}_i = \left\{ \mathbf{v} \in \mathbb{R}^s : \; (\overline{\mathbf{z}}_i^{(1)} - \overline{\mathbf{z}}_\ell^{(1)})^{\mathrm{T}} \left[\mathbf{v} - \frac{\overline{\mathbf{z}}_i^{(1)} + \overline{\mathbf{z}}_\ell^{(1)}}{2} \right] > \log \left(\frac{\widehat{\pi}_\ell}{\widehat{\pi}_j} \right) \quad \forall \ell \neq i \right\}$$

$$\widehat{\pi}_j = \frac{n_j}{n} \qquad n = \sum_{i=1}^{k} n_i$$

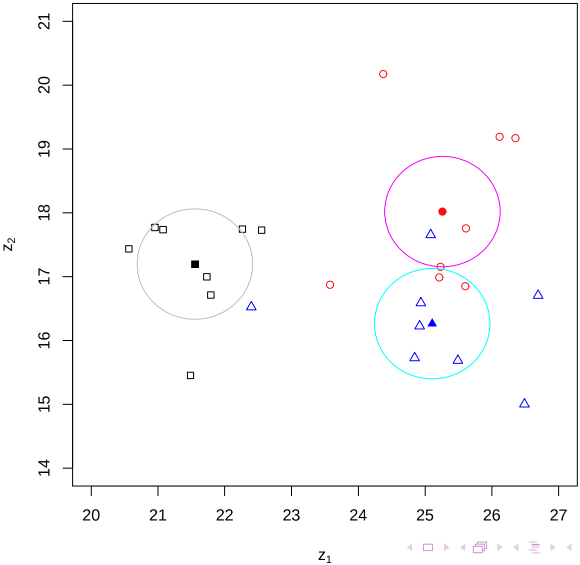
Ejemplo

En el ejemplo de los árboles tenemos $k = 3$, $p = 4$ luego $s = 2$, los autovalores no nulos de $\widehat{\mathbf{B}} = (\mathbf{T}^{-1})^{\mathrm{T}}\mathbf{H}\mathbf{T}^{-1}$ son 3.3522, 0.5879 y

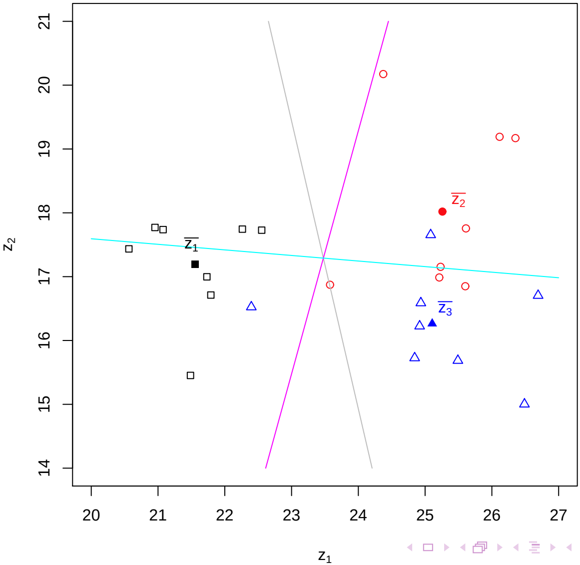
$$\widehat{\mathbf{B}} = \begin{pmatrix} 0.0139 & -0.4232 & 0.0000 & 0.9059 \\ -0.0139 & -0.0207 & 0.9996 & -0.0094 \\ -0.9845 & 0.1523 & -0.0098 & 0.0863 \\ 0.1742 & 0.8929 & 0.0248 & 0.4145 \end{pmatrix}$$

$$\widehat{\mathbf{A}}_1 = \begin{pmatrix} 7.7167 & 0.0711 \\ -2.7909 & 0.8135 \\ 6.1042 & 6.3137 \\ -1.9958 & -10.2292 \end{pmatrix}$$

Ejemplo: Se grafican los círculos de radio $\{\chi^2_{2,0.05}/n_i\}^{1/2} = 2.4477/\sqrt{n_i}$

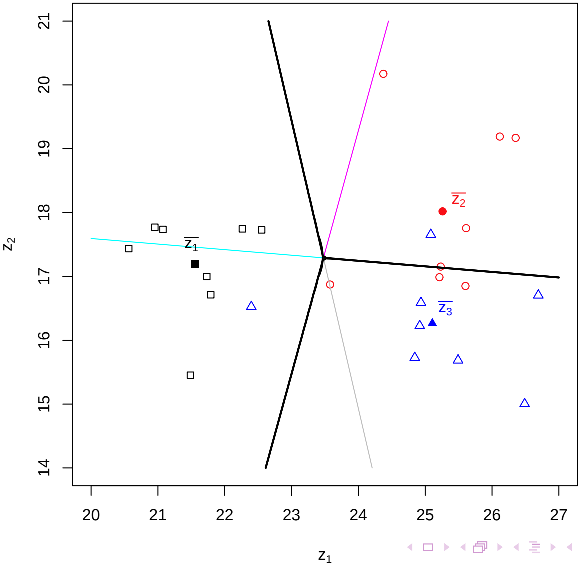


Ejemplo



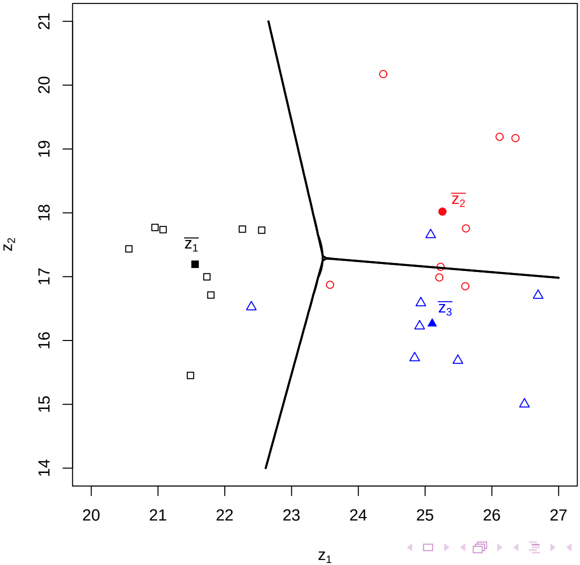
21
20
19
18
17
16
15
14
z_2
20
21
22
23
24
25
26
27
z_1
$\overline{z_1}$
$\overline{z_2}$
$\overline{z_3}$

Ejemplo



$\overline{z_1}$
$\overline{z_2}$
$\overline{z_3}$
14
15
16
17
18
19
20
21
z_2
20
21
22
23
24
25
26
27
z_1

Ejemplo



21
20
19
18
17
16
15
14
20 21 22 23 24 25 26 27
z_2
z_1
$\overline{z_1}$
$\overline{z_2}$
$\overline{z_3}$

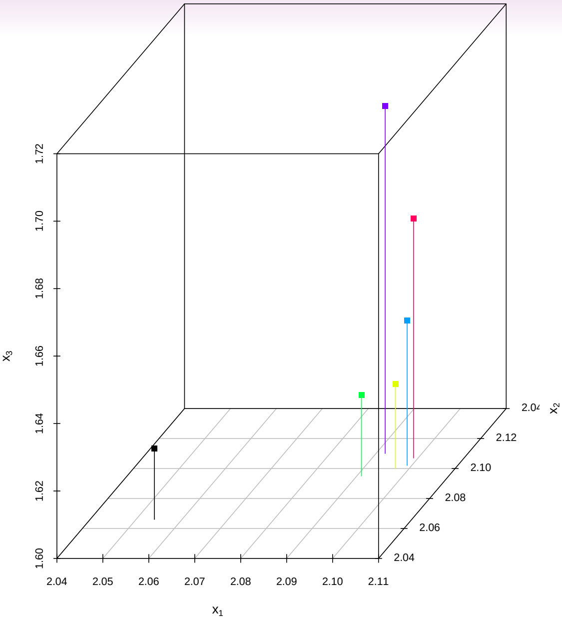
Ejemplo 2

Medidas tridimensionales en cráneos de 4 subespecies de oso hormiguero. Las mediciones hechas fueron

- $y_1 =$ Largo de base, excluyendo premaxilar
- $y_2 =$ Largo oxipitonasal
- $y_3 =$ Largo de los nasales.

Los datos consisten en $x_j = \log(y_j)$ para las subespecies

- *Instabilis*, (Colombia) $n_1 = 21$, $\bar{\mathbf{x}}_1 = (2.054, 2.066, 1.621)^{\mathrm{T}}$.
- *Chapadensis*, en tres localidades
  - Minas Gerais: $n_2 = 6$, $\bar{\mathbf{x}}_2 = (2.097, 2.1, 1.625)^{\mathrm{T}}$,
  - Matto Grosso: $n_3 = 9$, $\bar{\mathbf{x}}_3 = (2.091, 2.095, 1.624)^{\mathrm{T}}$,
  - Santa Cruz: $n_4 = 3$ $\bar{\mathbf{x}}_4 = (2.099, 2.102, 1.643)^{\mathrm{T}}$
- *Chiriquensis*, (Panamá) $n_5 = 4$, $\bar{\mathbf{x}}_5 = (2.092, 2.11, 1.703)^{\mathrm{T}}$
- *Mexicana* $n_6 = 5$, $\bar{\mathbf{x}}_6 = (2.099, 2.107, 1.671)^{\mathrm{T}}$


x1
x2
x3
2.04
2.05
2.06
2.07
2.08
2.09
2.10
2.11
2.04
2.06
2.08
2.10
2.12
1.60
1.62
1.64
1.66
1.68
1.70
1.72

Ejemplo 2

La matrices **U** y **H** son

$$\mathbf{U} = \begin{pmatrix} 0.01363 & 0.01277 & 0.01644 \\ & 0.01293 & 0.01714 \\ & & 0.03615 \end{pmatrix} \qquad \mathbf{H} = \begin{pmatrix} 0.02002 & 0.01744 & 0.01308 \\ & 0.01585 & 0.01507 \\ & & 0.03068 \end{pmatrix}$$

Luego, $\nu_1 = pk + \frac{p(p+1)}{2} = 24$ y $\nu_2 = p + \frac{p(p+1)}{2} = 9$ de donde

$$-48 \log \left(\frac{|\mathbf{U}|}{|\mathbf{U}+\mathbf{H}|} \right) \approx \chi^2_{15}$$

$$\frac{|\mathbf{U}|}{|\mathbf{U}+\mathbf{H}|} = 0.1468 \quad -48 \log \left(\frac{|\mathbf{U}|}{|\mathbf{U}+\mathbf{H}|} \right) = 92.08 \quad \chi^2_{15,0.01} = 30.57791$$

y rechazo la igualdad de medias.

Ejemplo 2

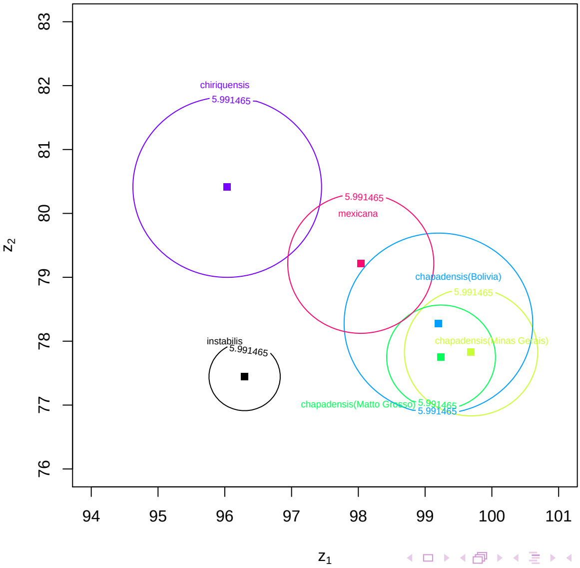
En este ejemplo $s = 3$. Los autovalores de $\widehat{B}$ son

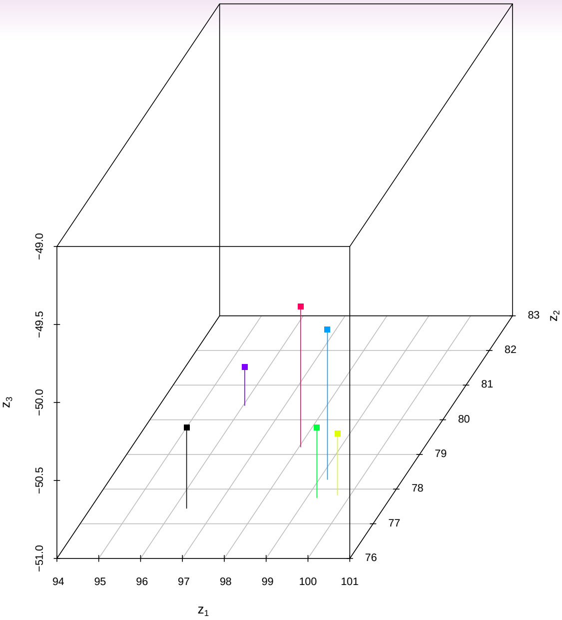
$$\widehat{\lambda}_1 = 2.4001 \quad \widehat{\lambda}_2 = 0.9050 \quad \widehat{\lambda}_3 = 0.0515$$

$$\widehat{\mathbf{A}}_1 = \begin{pmatrix} 108.9220 & -40.6120 & 169.3528 \\ -33.8285 & 59.9577 & -222.1998 \\ -35.4962 & 22.8195 & 37.4651 \end{pmatrix}$$

Graficaremos las dos primeras coordenadas discriminantes.

Ejemplo 2: Se grafican los círculos de radio $\{\chi^2_{2,0.05}/n_i\}^{1/2} = 2.4477/\sqrt{n_i}$




−49.0
−49.5
−50.0
−50.5
−51.0
z_3
94
95
96
97
98
99
100
101
z_1
76
77
78
79
80
81
82
83
z_2