

ESTUDIO DE UN CASO

Los datos que estudiaremos corresponden al experimento de Park Grass realizado por John Bennett Lawes en la estación experimental de Rothamsted en 1856.

Los datos son los rendimientos relativos de 44 especies de prado en 38 parcelas.

| Especie | 3d | 3a | ... | 11/2d | 11/2c | 11/2a | 11/2b |
|------------------------------|-----------|-----------|------------|--------------|--------------|--------------|--------------|
| <i>Agrostis tenuis</i> | 15,50 | 2,30 | ... | 1,60 | 1,60 | 0,00 | 0,20 |
| <i>Alopecurus pratensis</i> | 2,50 | 1,00 | ... | 0,00 | 7,40 | 28,70 | 15,20 |
| <i>Anthoxanthum odoratum</i> | 7,20 | 6,50 | ... | 0,30 | 0,05 | 0,20 | 0,00 |
| ⋮ | ⋮ | ⋮ | | ⋮ | ⋮ | ⋮ | ⋮ |
| <i>Veronica chamaedrys</i> | 0,00 | 0,05 | ... | 0,00 | 0,00 | 0,00 | 0,00 |
| Rendimiento total | 0,80 | 1,60 | ... | 7,00 | 8,60 | 7,30 | 7,80 |
| pH | 5,20 | 6,80 | ... | 3,80 | 4,20 | 5,10 | 5,50 |

Esquemas de fertilización y encalado

| Id. | Fertilizante | Id. | Fertilizante | Id. | Fertilizante |
|-----------|----------------|-----------|--------------|-------------------|--------------|
| 3 (d, a) | No fertilizado | 1 (d, c) | N1 | 4 (d, c, a, b) | N2P |
| 8 (d, a) | PNaMg | 18 (d, c) | N2KNaMg | 10 (d, c, a, b) | N2PNaMg |
| 7 (d, a) | PKNaMg | 13 (d, c) | Orgánico | 9 (d, c, a, b) | N2PKNaMg |
| 17 (d, a) | N4 | | | 11/1 (d, c, a, b) | N3PKNaMg |
| 16 (d, a) | N4PKNaMg | | | 11/2 (d, c, a, b) | N3PKNaMgSi |
| 14 (d, a) | N5PKNaMg | | | | |

- N1, N2 y N3: suministro de 48, 96 y 144 kg N ha⁻¹ de sulfato de amonio, respectivamente.
- N4 y N5: suministro de 48 y 96 kg N ha⁻¹ de nitrato de sodio, respectivamente.
- P: suministro de 35 kg P ha⁻¹ de superfosfato en polvo.
- K: suministro de 255 kg K ha⁻¹ de sulfato potásico.
- Na: suministro de 15 kg Na ha⁻¹ de sulfato de sodio.
- Mg: suministro de 11 kg Mg ha⁻¹ de sulfato de magnesio.
- Si: suministro de 450 kg ha⁻¹ de silicato de sodio en polvo.
- Orgánico: suministro de 35 t ha⁻¹ de abono orgánico.
- Las parcelas a recibieron 2.2 t CaO ha⁻¹ cada 4 años, y las parcelas d no fueron encaladas. Las parcelas b y c originalmente recibieron un tratamiento igual a las a y d pero años antes de la obtención de la muestra recibieron un tratamiento adicional de encalado.

Matrices de datos - I

Los datos del estudio pueden dar lugar a distintas matrices de datos:

Fichero Excel

| | A | B | C | D | E |
|---|------------------------------|-----------|-----------|-----------|-----------|
| 1 | ESPECIE | d3 | a3 | d8 | a8 |
| 2 | <i>Agrostis tenuis</i> | 15,50 | 2,30 | 4,60 | 4,00 |
| 3 | <i>Alopecurus pratensis</i> | 2,50 | 1,00 | 2,80 | 1,00 |
| 4 | <i>Anthoxanthum odoratum</i> | 7,20 | 6,50 | 9,60 | 13,10 |
| 5 | <i>Arrhenatherum elatius</i> | 0,20 | 0,20 | 2,00 | 6,10 |
| 6 | <i>Briza media</i> | 1,00 | 2,00 | 0,10 | 1,60 |
| 7 | Rendimiento total | 0,80 | 1,60 | 2,80 | 2,30 |
| 8 | Ph | 5,20 | 6,80 | 5,20 | 6,80 |
| 9 | | | | | |

- Podemos construir una matriz de datos \mathbf{X} de dimensión 44×38 , donde el elemento genérico x_{ij} represente el rendimiento relativo de la especie i en la parcela j .

$$\mathbf{X} = \begin{bmatrix} 15,50 & 2,30 & 4,60 & 4,00 \\ 2,50 & 1,00 & 2,80 & 1,00 \\ 7,20 & 6,50 & 9,60 & 13,10 \\ 0,20 & 0,20 & 2,00 & 6,10 \\ 1,00 & 2,00 & 0,10 & 1,60 \end{bmatrix}$$

Matrices de datos - II

- Podemos construir una matriz de datos \mathbf{Y} de dimensión 38×46 , donde el elemento genérico y_{ij} represente:
 - ▷ al rendimiento relativo en la parcela i de la especie j si $j < 45$,
 - ▷ al rendimiento total en la parcela i si $j = 45$,
 - ▷ al pH de la parcela i si $j = 46$.

$$\mathbf{Y} = \begin{bmatrix} 15,5 & 2,5 & 7,2 & 0,2 & 1,0 & 0,8 & 5,2 \\ 2,3 & 1,0 & 6,5 & 0,2 & 2,0 & 1,6 & 6,8 \\ 4,6 & 2,8 & 9,6 & 2,0 & 0,1 & 2,8 & 5,2 \\ 4,0 & 1,0 & 13,1 & 6,1 & 1,6 & 2,3 & 6,8 \end{bmatrix}$$

$\text{¿} \mathbf{Y}' = \mathbf{X} \text{?}$

Matrices de datos - III

Fichero Excel

| | A | Barra de fórmulas | | D | E |
|----|------------------------------|-------------------|-----------|-----------|-----------|
| 1 | ESPECIE | d3 | a3 | d8 | a8 |
| 2 | <i>Agrostis tenuis</i> | 15,50 | 2,30 | 4,60 | 4,00 |
| 3 | <i>Alopecurus pratensis</i> | 2,50 | 1,00 | 2,80 | 1,00 |
| 4 | <i>Anthoxanthum odoratum</i> | 7,20 | 6,50 | 9,60 | 13,10 |
| 5 | <i>Arrhenatherum elatius</i> | 0,20 | 0,20 | 2,00 | 6,10 |
| 6 | <i>Briza media</i> | 1,00 | 2,00 | 0,10 | 1,60 |
| 7 | Rendimiento total | 0,80 | 1,60 | 2,80 | 2,30 |
| 8 | Ph | 5,20 | 6,80 | 5,20 | 6,80 |
| 9 | | | | | |
| 10 | | 0,12 | 0,04 | 0,13 | 0,09 |
| 11 | | 0,02 | 0,02 | 0,08 | 0,02 |
| 12 | | 0,06 | 0,10 | 0,27 | 0,30 |
| 13 | | 0,00 | 0,00 | 0,06 | 0,14 |
| 14 | | 0,01 | 0,03 | 0,00 | 0,04 |
| 15 | | | | | |

- Podemos construir una matriz de datos \mathbf{Z} de dimensión 44×38 , donde el elemento genérico z_{ij} represente el rendimiento absoluto de la especie i en la parcela j . Notemos que $z_{ij} = \frac{x_{ij}y_{j45}}{100}$.

$$\mathbf{Z} = \begin{bmatrix} 0,12 & 0,04 & 0,13 & 0,09 \\ 0,02 & 0,02 & 0,08 & 0,02 \\ 0,06 & 0,10 & 0,27 & 0,30 \\ 0,00 & 0,00 & 0,06 & 0,14 \\ 0,01 & 0,03 & 0,00 & 0,04 \end{bmatrix}$$

Matrices de datos - IV

Fichero Excel

| | A | B | C | D | E | F |
|----|------------------------------|-----------|-----------|-----------|-----------|------|
| 1 | ESPECIE | d3 | a3 | d8 | a8 | |
| 2 | <i>Agrostis tenuis</i> | 15,50 | 2,30 | 4,60 | 4,00 | |
| 3 | <i>Alopecurus pratensis</i> | 2,50 | 1,00 | 2,80 | 1,00 | |
| 4 | <i>Anthoxanthum odoratum</i> | 7,20 | 6,50 | 9,60 | 13,10 | |
| 5 | <i>Arrhenatherum elatius</i> | 0,20 | 0,20 | 2,00 | 6,10 | |
| 6 | <i>Briza media</i> | 1,00 | 2,00 | 0,10 | 1,60 | |
| 7 | Rendimiento total | 0,80 | 1,60 | 2,80 | 2,30 | |
| 8 | Ph | 5,20 | 6,80 | 5,20 | 6,80 | |
| 9 | | | | | | |
| 10 | | 1,00 | 1,00 | 1,00 | 1,00 | |
| 11 | | 1,00 | 1,00 | 1,00 | 1,00 | |
| 12 | | 1,00 | 1,00 | 1,00 | 1,00 | |
| 13 | | 0,00 | 0,00 | 1,00 | 1,00 | |
| 14 | | 1,00 | 1,00 | 0,00 | 1,00 | |
| 15 | | | | | | |
| 16 | | 4,00 | 4,00 | 4,00 | 2,00 | 3,00 |
| 17 | | 4,00 | 4,00 | 2,00 | 3,00 | |
| 18 | | 4,00 | 2,00 | 3,00 | | |
| 19 | | 2,00 | 1,00 | | | |
| 20 | | 3,00 | | | | |

- Podemos construir una matriz de datos W de dimensión 44×44 , donde el elemento genérico w_{ij} represente el número de veces que las especies i y j cohabitan.

$$W = \begin{bmatrix} 4,00 & 4,00 & 4,00 & 2,00 & 3,00 \\ 4,00 & 4,00 & 4,00 & 2,00 & 3,00 \\ 4,00 & 4,00 & 4,00 & 2,00 & 3,00 \\ 2,00 & 2,00 & 3,00 & 2,00 & 1,00 \\ 3,00 & 3,00 & 2,00 & 1,00 & 3,00 \end{bmatrix}$$

Análisis descriptivo numérico I

En la Tabla se presentan las estimaciones del mínimo, máximo, media, desviación típica y varianza de las 38 variables (parcelas):

| | N | Mínimo | Máximo | Media | Desv. típ. | Varianza |
|----|----|--------|--------|-------|------------|----------|
| D3 | 44 | ,0 | 33,2 | 2,268 | 5,6887 | 32,361 |
| A3 | 44 | ,0 | 13,6 | 2,275 | 3,6767 | 13,518 |
| D8 | 44 | ,0 | 16,7 | 2,272 | 4,3941 | 19,309 |
| A8 | 44 | ,0 | 13,1 | 2,274 | 3,5467 | 12,579 |
| D7 | 44 | ,0 | 29,0 | 2,267 | 5,2419 | 27,477 |
| | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | |

- Un primer resultado “inesperado”: las medias son muy similares en todas las parcelas. PERO: si las variables representan porcentajes y la suma $\sum_{i=1}^{44} x_{ij}$ debe ser igual a 100 (salvo errores de redondeo), entonces la media de las variables será igual a $100/44 \approx 2,273$.
- Existen parcelas, por ejemplo d11_1 y d11_2 en las que fundamentalmente se ha desarrollado una especie.
- En todas las parcelas, alguna de las 44 especies no ha sido observada.

Análisis descriptivo numérico II

¿Cuáles especies tienen mayor rendimiento relativo medio?

| | N | Mínimo | Máximo | Media | Desv. típ. | Varianza |
|----------|----|--------|--------|---------|------------|----------|
| ARRHENAT | 38 | ,00 | 63,90 | 17,2145 | 21,09378 | 444,948 |
| AGROSTIS | 38 | ,00 | 84,40 | 13,5487 | 21,00719 | 441,302 |
| HOLCUS_L | 38 | ,00 | 98,10 | 13,0224 | 22,59639 | 510,597 |
| ANTHOXAN | 38 | ,00 | 76,10 | 11,8566 | 18,90876 | 357,541 |
| FESTUCA | 38 | ,00 | 55,90 | 11,7961 | 16,87219 | 284,671 |
| ALOPECUR | 38 | ,00 | 36,80 | 8,5645 | 9,98187 | 99,638 |
| POA_PRAT | 38 | ,00 | 19,50 | 3,2868 | 5,39143 | 29,068 |
| PLANTAGO | 38 | ,00 | 23,90 | 2,7658 | 5,27570 | 27,833 |
| DACTYLIS | 38 | ,00 | 11,00 | 2,1645 | 2,54609 | 6,483 |
| LATHYRUS | 38 | ,00 | 15,60 | 1,9882 | 3,97432 | 15,795 |

- Si quisiéramos concentrarnos en las especies “más frecuentes”: ¿El orden basado en las medias es satisfactorio? Si, PERO debemos contrastarlo con otros ordenes basados en medidas robustas de posición.

Análisis descriptivo numérico III

¿Cuáles especies tienen mayor rendimiento relativo medio?

- En la Tabla se presentan las estimaciones de los percentiles del 5 %, 10 %, 25 % (Q_1), 50 % ($Q_2 \equiv$ mediana), 75 % (Q_3), 90 % y 95 % que permiten decidir sobre cuáles especies son más frecuentes.

| | Percentiles | | | | | | |
|----------|-------------|-------|--------|--------|---------|---------|---------|
| | 5 | 10 | 25 | 50 | 75 | 90 | 95 |
| AGROSTIS | ,0000 | ,0000 | ,0375 | 3,7000 | 19,9500 | 37,5800 | 82,7850 |
| ALOPECUR | ,0000 | ,0000 | ,0875 | 6,3000 | 15,2250 | 28,7100 | 29,9600 |
| ANTHOXAN | ,0000 | ,0450 | ,3000 | 6,7500 | 11,7500 | 27,6000 | 71,9200 |
| ARRHENAT | ,0000 | ,0000 | ,0375 | 2,3500 | 37,1500 | 50,9700 | 60,9550 |
| DACTYLIS | ,0000 | ,0000 | ,0000 | 1,7000 | 3,0250 | 5,7600 | 7,9600 |
| FESTUCA | ,0000 | ,0000 | ,1000 | 3,0500 | 14,9250 | 41,6800 | 53,2400 |
| HOLCUS_L | ,0000 | ,0450 | 1,4500 | 4,1000 | 14,5250 | 37,4600 | 95,0600 |
| POA_PRAT | ,0000 | ,0000 | ,0375 | ,9000 | 2,6000 | 13,6400 | 16,2700 |
| RUMEX_AC | ,0000 | ,0000 | ,0000 | ,3000 | 1,4250 | 2,6100 | 2,9550 |
| TARAXACU | ,0000 | ,0000 | ,1000 | ,5000 | 1,6000 | 4,5300 | 7,9450 |

Análisis descriptivo numérico IV

pH del suelo

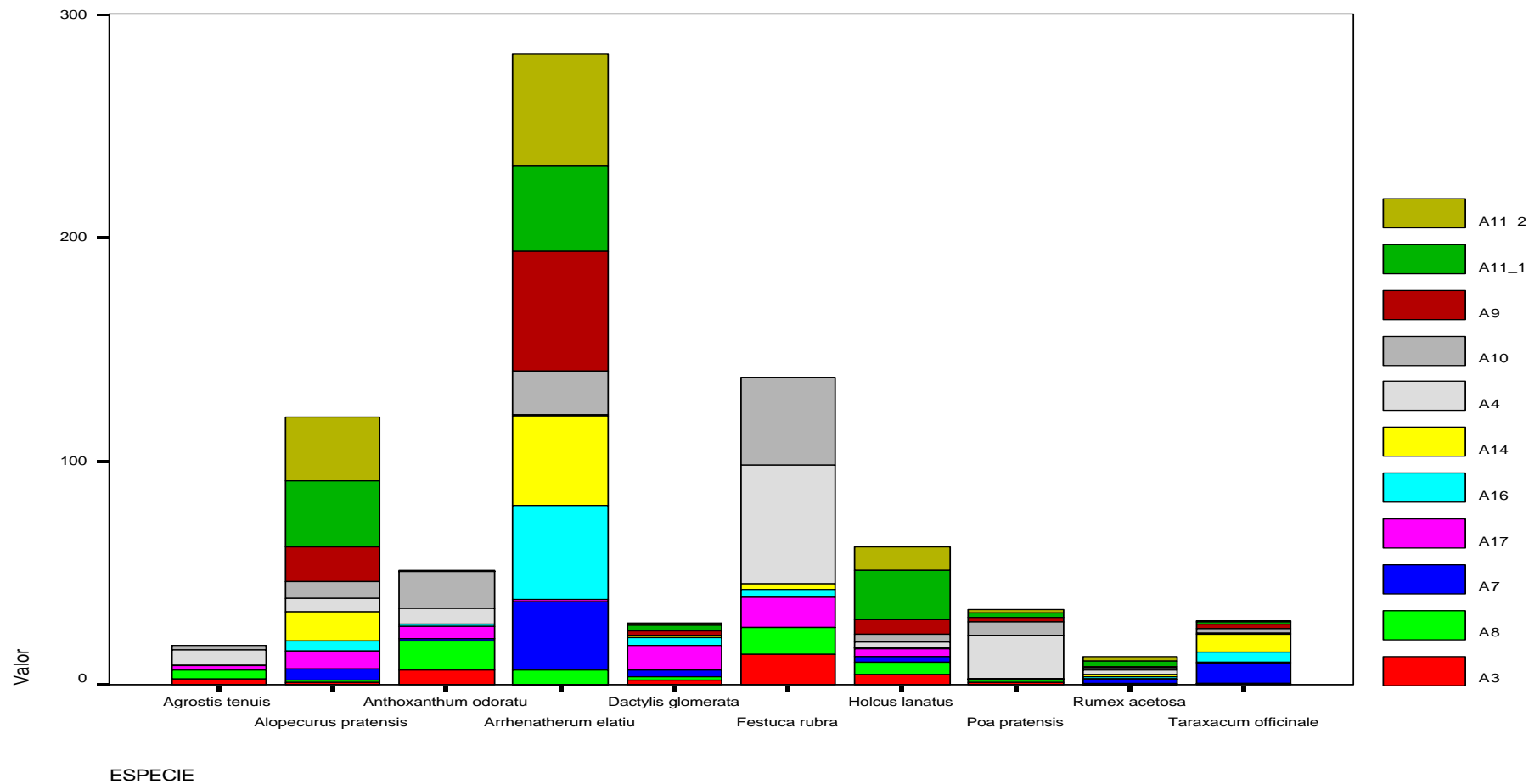
| | | |
|-------------|-------------|--------|
| Media | | 5,0632 |
| Desv. típ. | | ,99361 |
| Varianza | | ,98725 |
| Mínimo | | 3,70 |
| Máximo | | 7,10 |
| Percentiles | 25 | 4,2000 |
| | 33,33333333 | 4,3000 |
| | 50 | 5,0000 |
| | 66,66666667 | 5,5000 |
| | 75 | 5,8000 |

- En media, el pH de las parcelas es neutro, pero existen parcelas ácidas (pH mínimo igual a 3.7) y parcelas bases (pH máximo igual a 7.1).
- Hemos obtenido los percentiles del 33.33% y 66.66% para dividir la muestra en tres partes iguales teniendo en cuenta esos puntos de corte.

| | |
|---------|---|
| Ácidas | a11_1 c1 c10 c11_2 c4 c9 d1 d10 d11_1 d11_2 d18 d4 d9 |
| Neutras | a10 a11_2 a9 b11_1 b11_2 c11_1 c13 c18 d13 d16 d3 d7 d8 |
| Bases | a14 a16 a17 a3 a4 a7 a8 b10 b4 b9 d14 d17 |

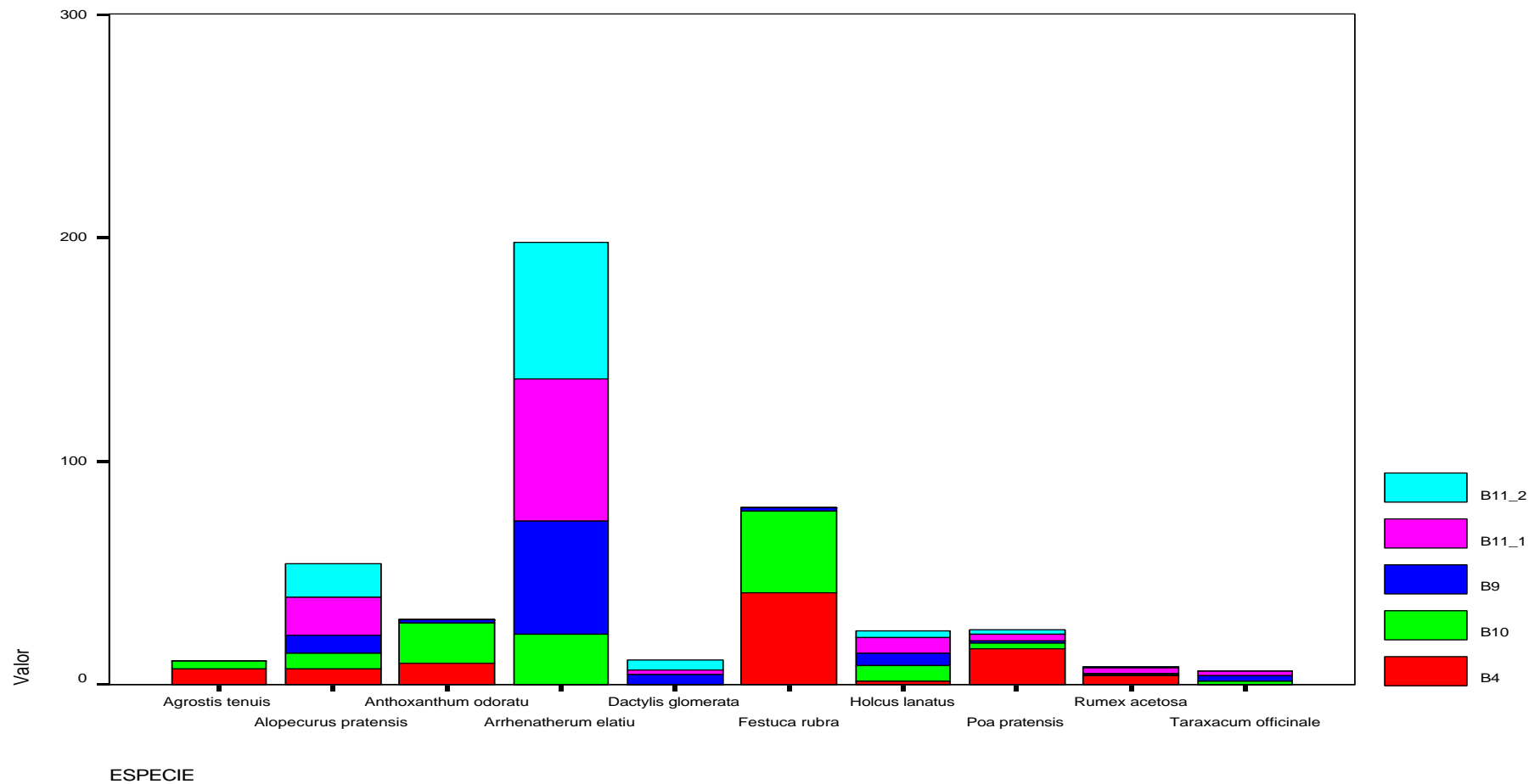
Análisis descriptivo gráfico I

- Las parcelas más frecuentes en las parcelas A son: *Arrhenatherum elatius*, y en menor cuantía *Festuca rubra* y *Alopecurus pratensis*.



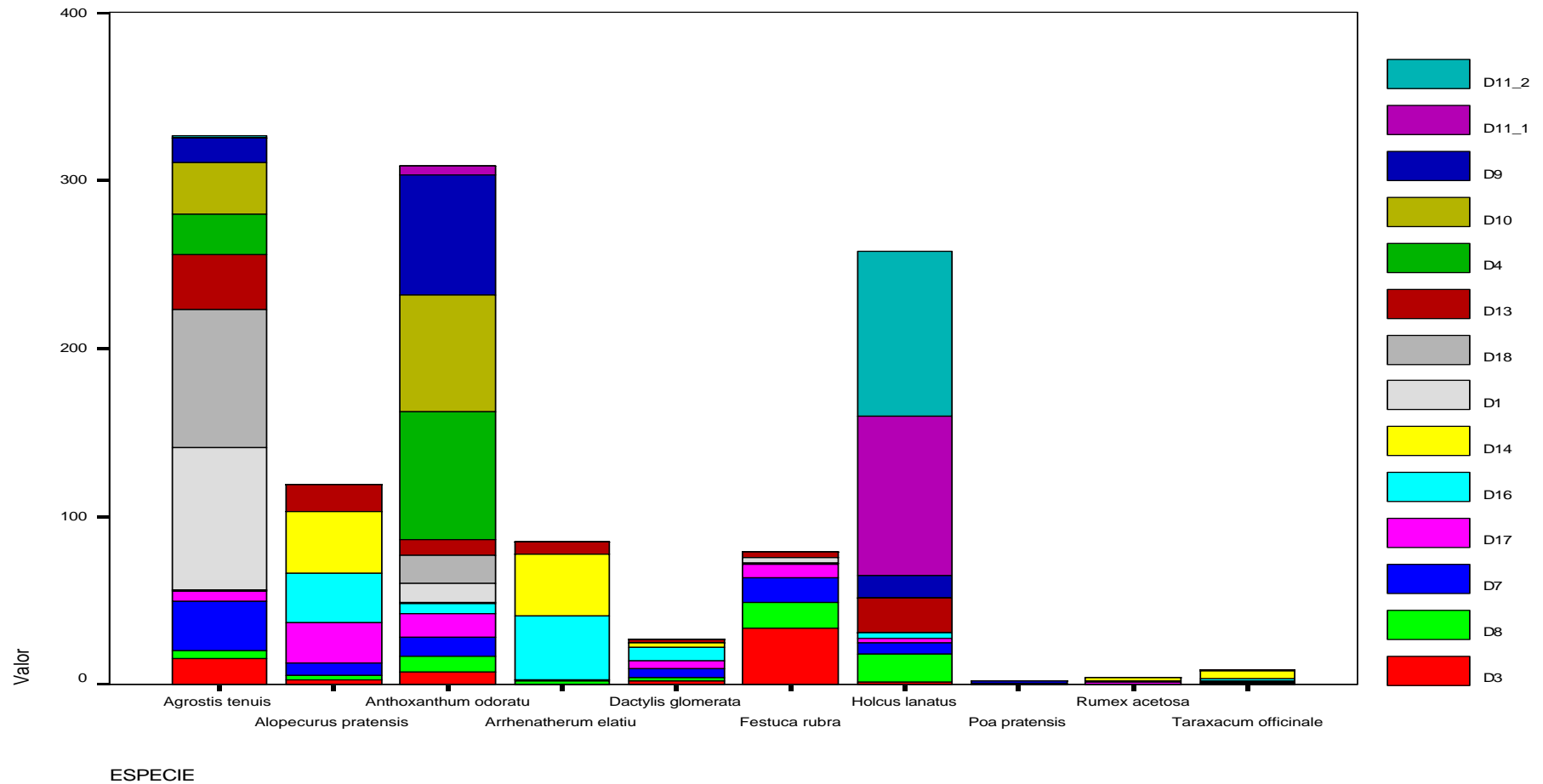
Análisis descriptivo gráfico II

- En el caso de la especie *Arrhenatherum elatius* el resultado concuerda con lo obtenido en las parcelas A.



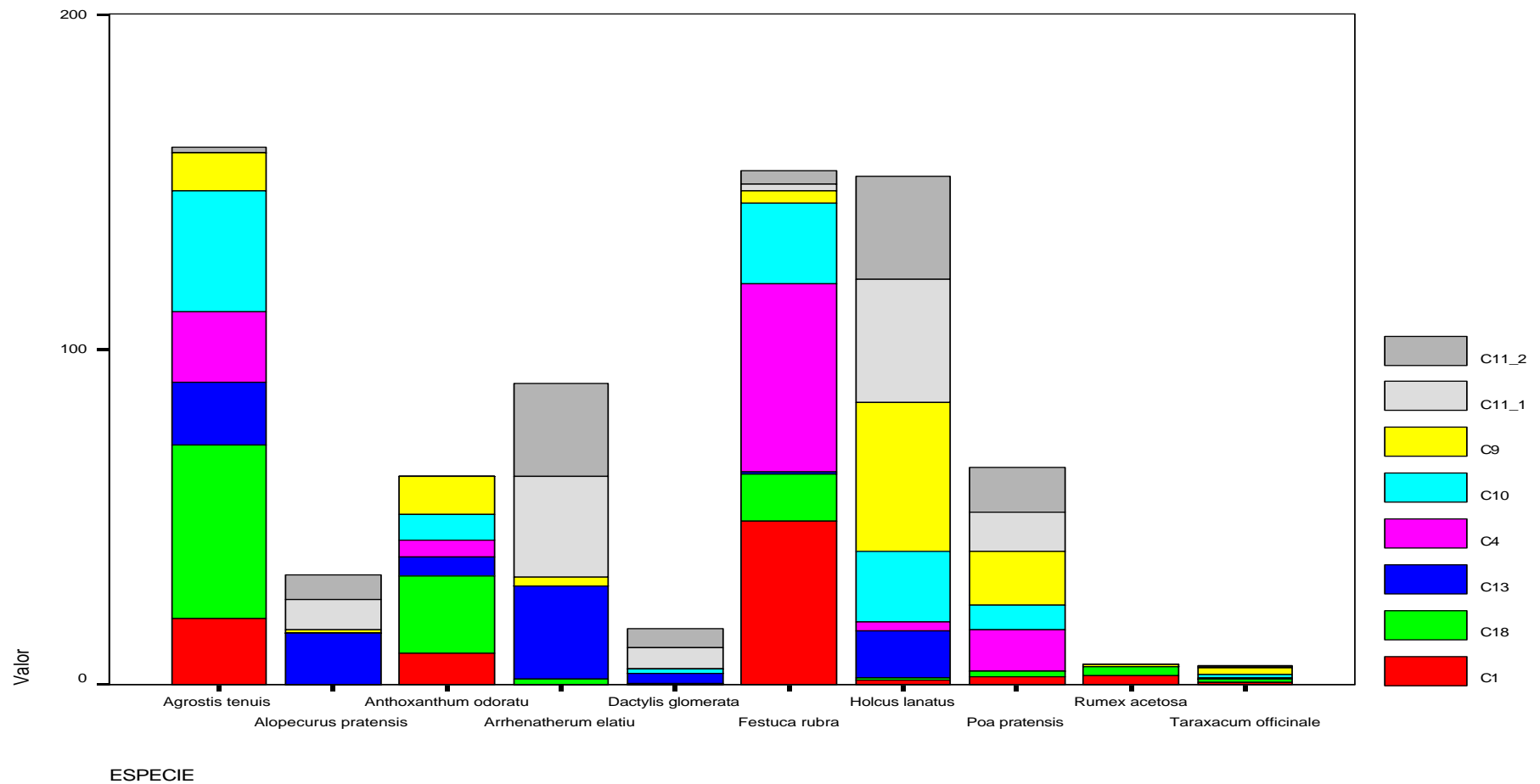
Análisis descriptivo gráfico III

- Las especies más frecuentes en las parcelas D son: *Agrostis tenuis*, *Anthoxanthum odoratum* y *Holcus lanatus*.



Análisis descriptivo gráfico IV

- En el caso de las especies *Agrostis tenuis* y *Holcus lanatus* el resultado concuerda con lo obtenido en las parcelas D.

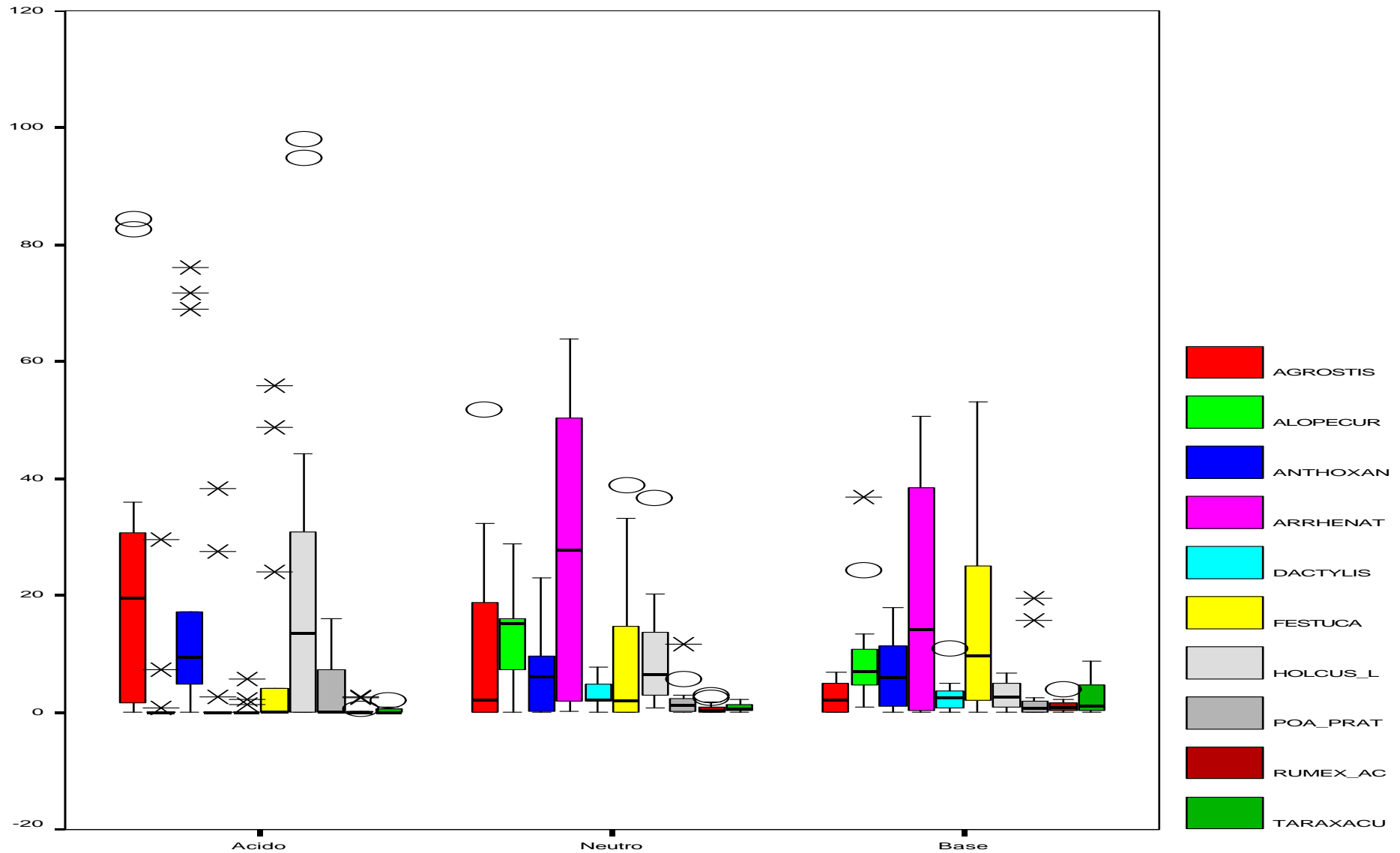


Análisis descriptivo gráfico V

Otra observación es que no en todas las parcelas los rendimientos tienen valores comparables. ¿A qué puede deberse esta diferencia? ¿pH?

- En parcelas con pH ácido son más frecuentes las especies: *Agrostis tenuis*, y *Holcus lanatus*. Las especies *Alopecurus pratensis*, *Arrhenatherum elatius*, *Dactylis glomerata* y *Rumex acetosa* presentan un pobre rendimiento en este medio.
- En parcelas con pH neutro son más frecuentes las especies: *Arrhenatherum elatius* y *Alopecurus pratensis*. Las especies *Rumex acetosa* y *Taraxacum officinale* presentan un pobre rendimiento en este medio.
- En parcelas con pH base son más frecuentes las especies: *Arrhenatherum elatius* y *Festuca rubra*.
- La especie *Anthoxanthum odoratum* se “adapta” bien a todos los medios.
- El aumento del pH es favorable a la especie *Festuca rubra* y desfavorable a la especie *Agrostis tenuis*.

Análisis descriptivo gráfico VI



Estudio de un caso - Reducción de la dimensión

Seguiremos trabajando con los datos del experimento de Park Grass. Los datos son los rendimientos relativos de 44 especies de prado en 38 parcelas.

| | A | B | C | D | E | F | G | H | I | J | K |
|----|---------------------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-----|
| 1 | ESPECIE | d3 | a3 | d8 | a8 | d7 | a7 | d17 | a17 | d16 | a16 |
| 2 | <i>Agrostis tenuis</i> | 15,50 | 2,30 | 4,60 | 4,00 | 29,00 | 0,40 | 6,10 | 1,90 | 1,00 | 0, |
| 3 | <i>Alopecurus pratensis</i> | 2,50 | 1,00 | 2,80 | 1,00 | 7,30 | 5,20 | 24,40 | 7,90 | 28,80 | 4, |
| 4 | <i>Anthoxanthum odoratum</i> | 7,20 | 6,50 | 9,60 | 13,10 | 11,30 | 1,10 | 14,20 | 5,30 | 6,10 | 1, |
| 5 | <i>Arrhenatherum elatius</i> | 0,20 | 0,20 | 2,00 | 6,10 | 0,40 | 30,80 | 0,30 | 0,80 | 37,60 | 42, |
| 6 | <i>Briza media</i> | 1,00 | 2,00 | 0,10 | 1,60 | 0,00 | 0,00 | 0,20 | 3,90 | 0,00 | 0, |
| 7 | <i>Bromus mollis</i> | 0,00 | 0,00 | 0,00 | 0,00 | 0,05 | 0,90 | 0,00 | 0,00 | 0,00 | 0, |
| 8 | <i>Cynosurus cristatus</i> | 0,00 | 0,10 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,05 | 0,00 | 0, |
| 9 | <i>Dactylis glomerata</i> | 2,20 | 2,20 | 1,90 | 1,50 | 5,00 | 3,00 | 5,00 | 11,00 | 7,80 | 3, |
| 10 | <i>Deschampsia caespitosa</i> | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0, |
| 11 | <i>Festuca rubra</i> | 33,20 | 13,60 | 15,60 | 11,70 | 14,70 | 0,30 | 7,80 | 13,40 | 1,00 | 3, |
| 12 | <i>Festuca pratensis</i> | 0,00 | 0,00 | 0,30 | 3,60 | 0,00 | 0,50 | 0,30 | 0,00 | 0,00 | 0, |
| 13 | <i>Helictotrichon pubescens</i> | 0,30 | 8,50 | 0,30 | 12,00 | 0,10 | 0,30 | 0,40 | 16,70 | 0,60 | 1, |
| 14 | <i>Holcus lanatus</i> | 1,50 | 4,30 | 16,70 | 5,60 | 6,50 | 2,50 | 2,70 | 3,80 | 3,00 | 0, |
| 15 | <i>Poa pratensis</i> | 0,20 | 0,90 | 0,20 | 0,90 | 1,30 | 0,60 | 0,00 | 0,05 | 0,10 | 0, |
| 16 | <i>Poa trivialis</i> | 0,00 | 0,40 | 0,40 | 1,20 | 0,20 | 1,10 | 0,20 | 0,10 | 0,80 | 3, |

⋮

⋮

⋮

⋮

⋮

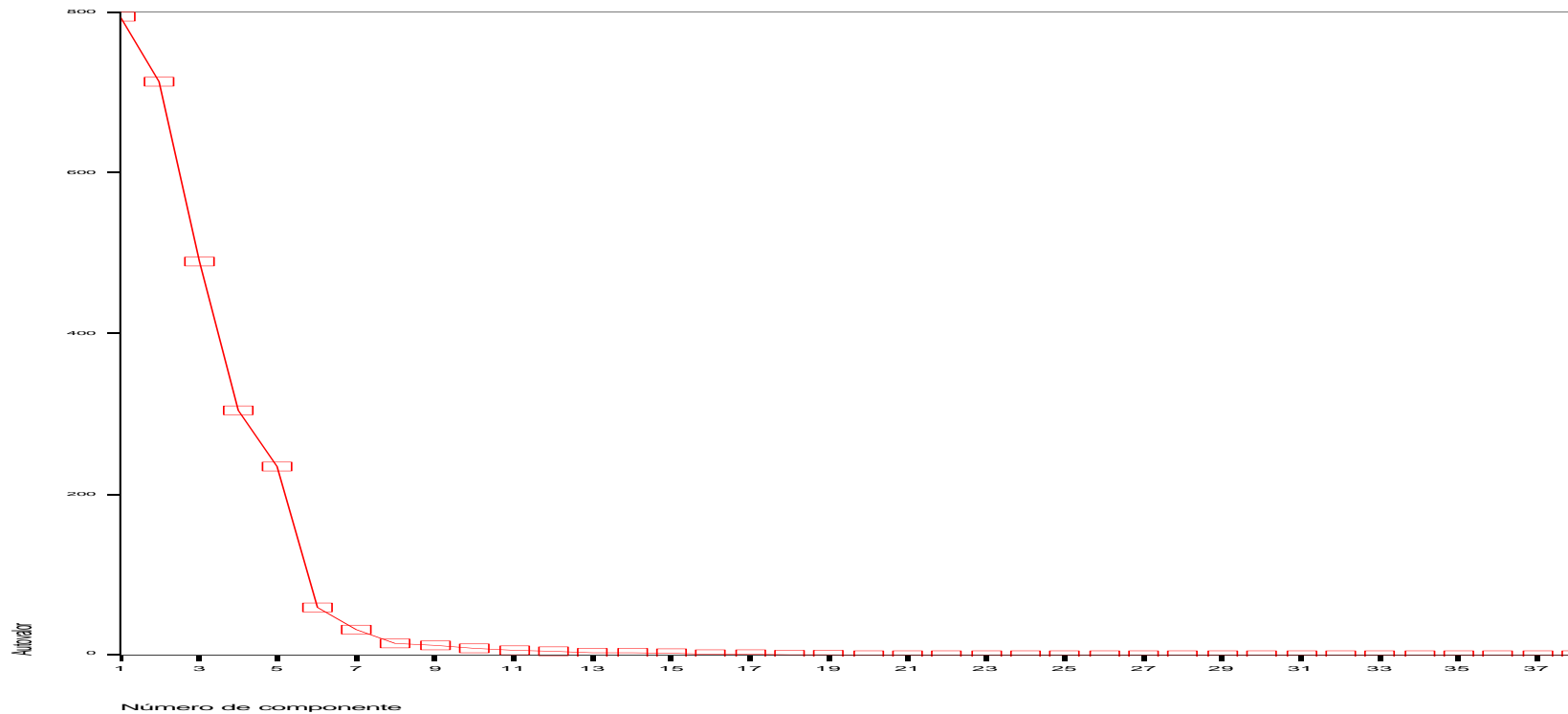
- Tenemos 38 variables si trabajamos con la matriz X .
- Tenemos 44 variables si trabajamos con la matriz X' .

Análisis de componentes principales I

| | | Sumas de las saturaciones al cuadrado de la extracción | | |
|-------|------------|--|------------------|-------------|
| | | Total | % de la varianza | % acumulado |
| Bruta | Componente | | | |
| | 1 | 794,039 | 29,557 | 29,557 |
| | 2 | 712,745 | 26,531 | 56,088 |
| | 3 | 489,272 | 18,212 | 74,300 |
| | 4 | 304,504 | 11,335 | 85,635 |
| | 5 | 234,471 | 8,728 | 94,362 |
| | 6 | 59,578 | 2,218 | 96,580 |
| | 7 | 31,846 | 1,185 | 97,766 |
| | 8 | 15,003 | ,558 | 98,324 |
| | 9 | 12,403 | ,462 | 98,786 |
| | 10 | 8,588 | ,320 | 99,105 |
| | ⋮ | ⋮ | ⋮ | ⋮ |

- El criterio de la variabilidad explicada mayor que el 80% ó que el 90% sugiere utilizar 4 ó 5 componentes.
- El criterio del valor propio mayor que la media de los valores propios ($\sum_{j=1}^{38} \lambda_j / 38 = 70,6897$), sugiere utilizar 5 componentes.

Análisis de componentes principales I



- El criterio del gráfico de sedimentación sugiere utilizar 7 componentes.

Teniendo en cuenta que 5 componentes explican el 94.3% y que los componentes 6 y 7 explicarían el 2.2% y 1.2%, respectivamente, utilizaremos 5 componentes.

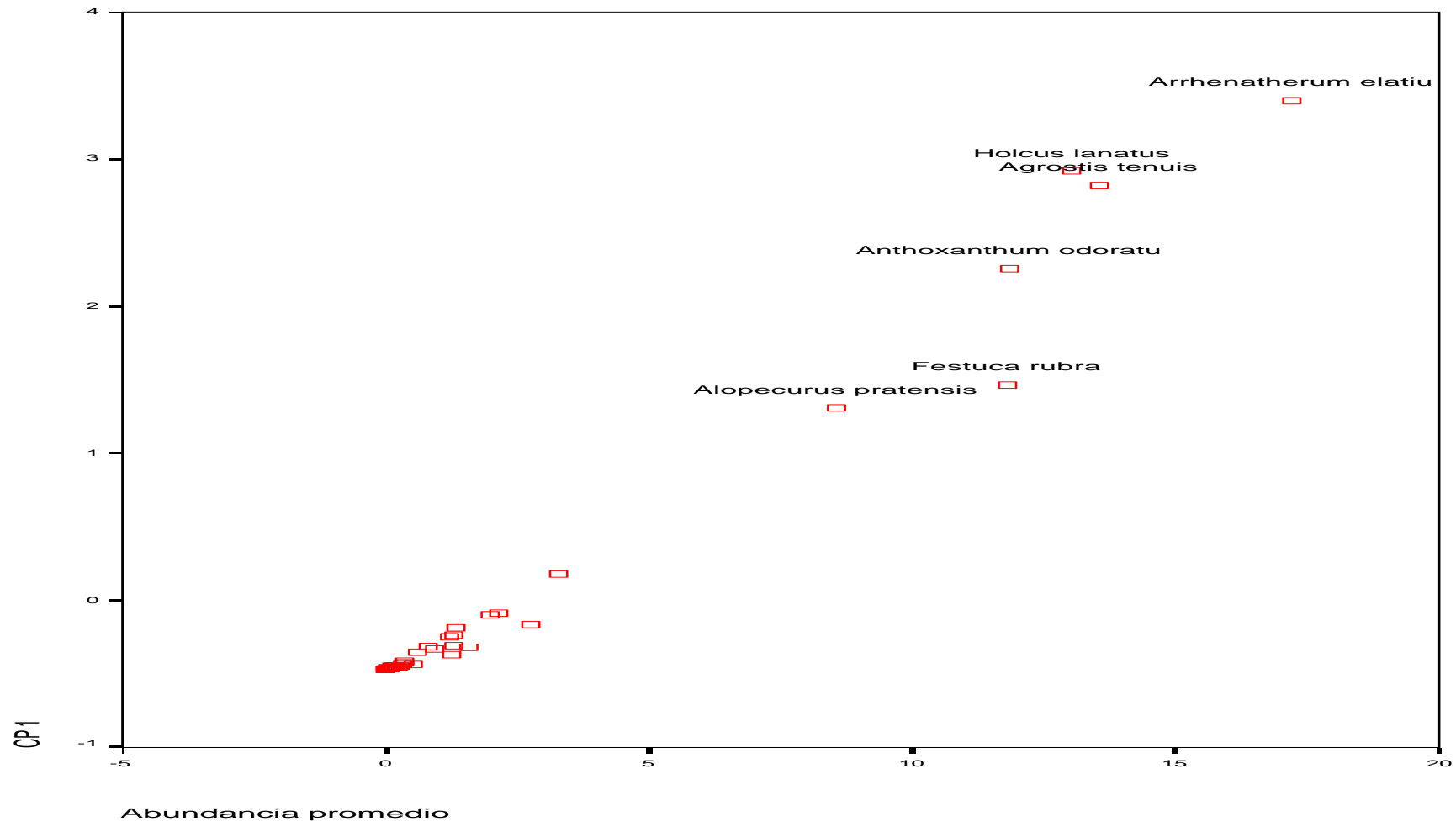
Análisis de componentes principales I - Matriz de carga

| | Bruta | | | | | Reescalada | | | | |
|-------|------------|--------|---------|--------|--------|------------|-------|-------|-------|-------|
| | Componente | | | | | Componente | | | | |
| | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| D3 | 2,409 | -2,702 | | 3,812 | | ,423 | -,475 | | ,670 | |
| A3 | ,738 | -,923 | | 1,605 | ,872 | ,201 | -,251 | | ,437 | ,237 |
| D8 | 2,387 | | -1,492 | 1,596 | ,979 | ,543 | | -,340 | ,363 | ,223 |
| A8 | 1,810 | -,820 | | ,883 | 1,396 | ,510 | -,231 | | ,249 | ,394 |
| D7 | 3,544 | -3,162 | | | -1,284 | ,676 | -,603 | | | -,245 |
| A7 | 2,544 | 3,292 | 1,868 | | | ,476 | ,616 | ,349 | | |
| D17 | 2,123 | | | | 1,283 | ,384 | | | | ,232 |
| A17 | ,973 | | | 1,652 | | ,237 | | | ,402 | |
| D16 | 4,368 | 4,203 | 2,797 | | | ,615 | ,592 | ,394 | | |
| A16 | 3,411 | 4,244 | 2,856 | | | ,497 | ,618 | ,416 | | |
| D14 | 3,935 | 4,878 | 3,163 | | | ,504 | ,625 | ,405 | | |
| A14 | 3,384 | 4,350 | 2,918 | | | ,526 | ,677 | ,454 | | |
| D1 | 6,207 | -7,993 | | | -6,972 | ,485 | -,625 | | | -,545 |
| C1 | 3,407 | -3,765 | | 5,726 | | ,432 | -,478 | | ,727 | |
| D18 | 6,337 | -8,145 | | -3,109 | -6,170 | ,500 | -,643 | | -,245 | -,487 |
| C18 | 5,220 | -6,132 | 1,993 | | -2,188 | ,605 | -,711 | ,231 | | -,254 |
| D13 | 5,074 | -1,761 | | | -2,161 | ,819 | -,284 | | | -,349 |
| C13 | 5,069 | 1,566 | | | -1,360 | ,884 | ,273 | | | -,237 |
| D4 | 5,563 | -6,870 | | -4,537 | 6,326 | ,466 | -,575 | | -,380 | ,530 |
| C4 | 3,759 | -3,768 | | 7,131 | | ,415 | -,416 | | ,787 | |
| A4 | 3,034 | -2,404 | | 7,214 | | ,357 | -,283 | | ,848 | |
| B4 | 2,595 | -2,214 | | 5,399 | | ,377 | -,321 | | ,784 | |
| D10 | 5,650 | -6,992 | | -4,349 | 4,848 | ,501 | -,620 | | -,385 | ,430 |
| C10 | 5,059 | -3,816 | | 2,072 | -2,229 | ,704 | -,531 | | ,289 | -,310 |
| A10 | 4,363 | | 1,815 | 4,352 | 2,397 | ,631 | | ,262 | ,629 | ,347 |
| B10 | 4,828 | | 1,611 | 3,798 | 2,382 | ,700 | | ,234 | ,551 | ,345 |
| D9 | 5,645 | -5,498 | | -4,090 | 6,667 | ,508 | -,495 | | -,368 | ,600 |
| C9 | 4,709 | | -5,155 | | | ,645 | | -,706 | | |
| A9 | 5,066 | 5,889 | 2,880 | | | ,601 | ,699 | ,342 | | |
| B9 | 4,645 | 5,290 | 2,738 | | | ,582 | ,663 | ,343 | | |
| D11_1 | 6,707 | | -12,475 | | | ,469 | | -,872 | | |
| C11_1 | 5,227 | 3,977 | -2,777 | | | ,716 | ,544 | -,380 | | |
| A11_1 | 5,371 | 5,180 | | | | ,689 | ,665 | | | |
| B11_1 | 6,014 | 6,866 | 3,416 | | | ,607 | ,693 | ,345 | | |
| D11_2 | 6,778 | | -12,936 | | | ,458 | | -,875 | | |
| C11_2 | 4,739 | 3,379 | -2,127 | | | ,726 | ,518 | -,326 | | |
| A11_2 | 5,539 | 6,035 | 2,403 | | | ,636 | ,693 | ,276 | | |
| B11_2 | 5,390 | 6,467 | 3,777 | | | ,570 | ,684 | ,399 | | |

Método de extracción: Análisis de componentes principales.

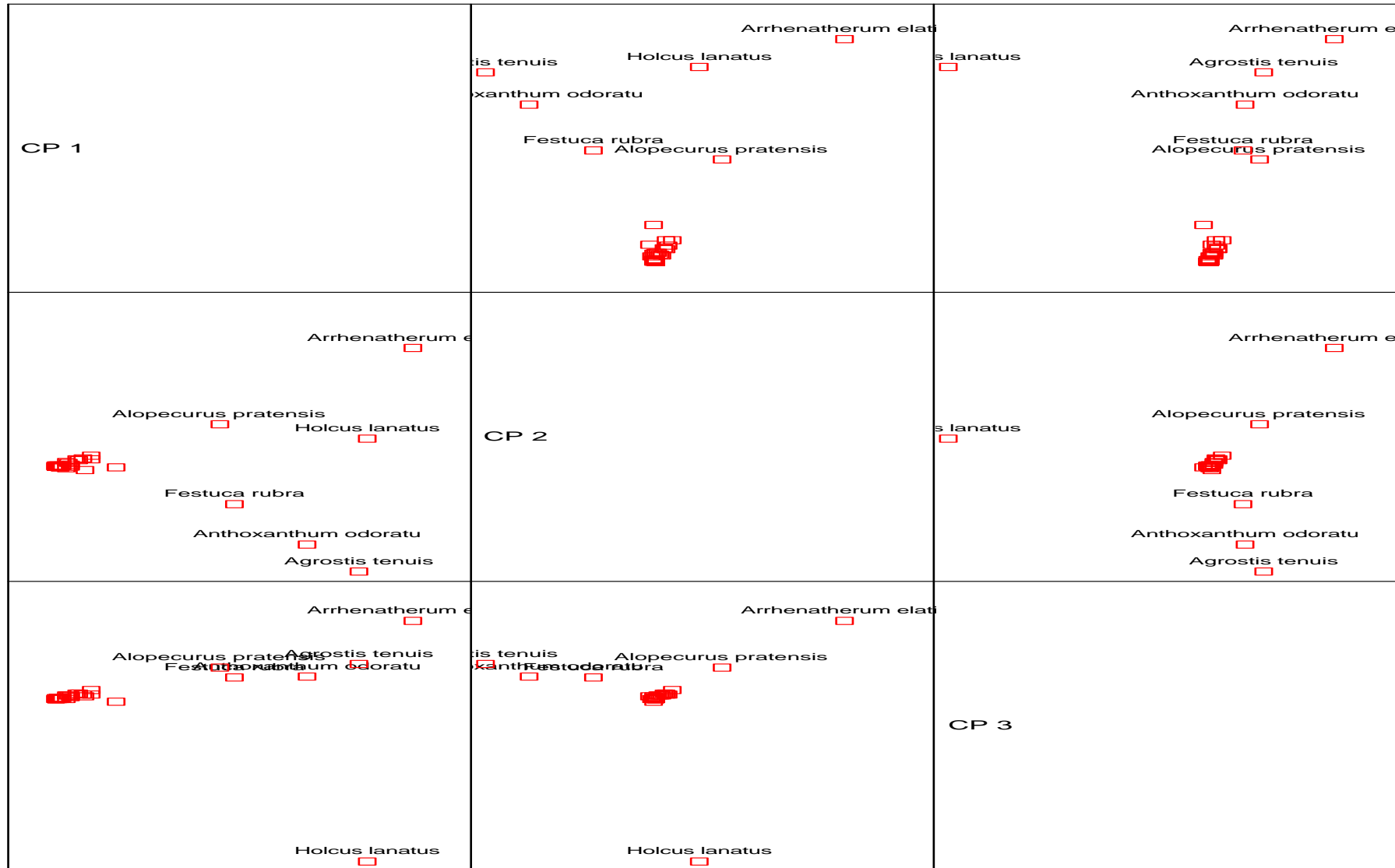
a. 5 componentes extraídos

Componentes principales I - Diagramas de dispersión

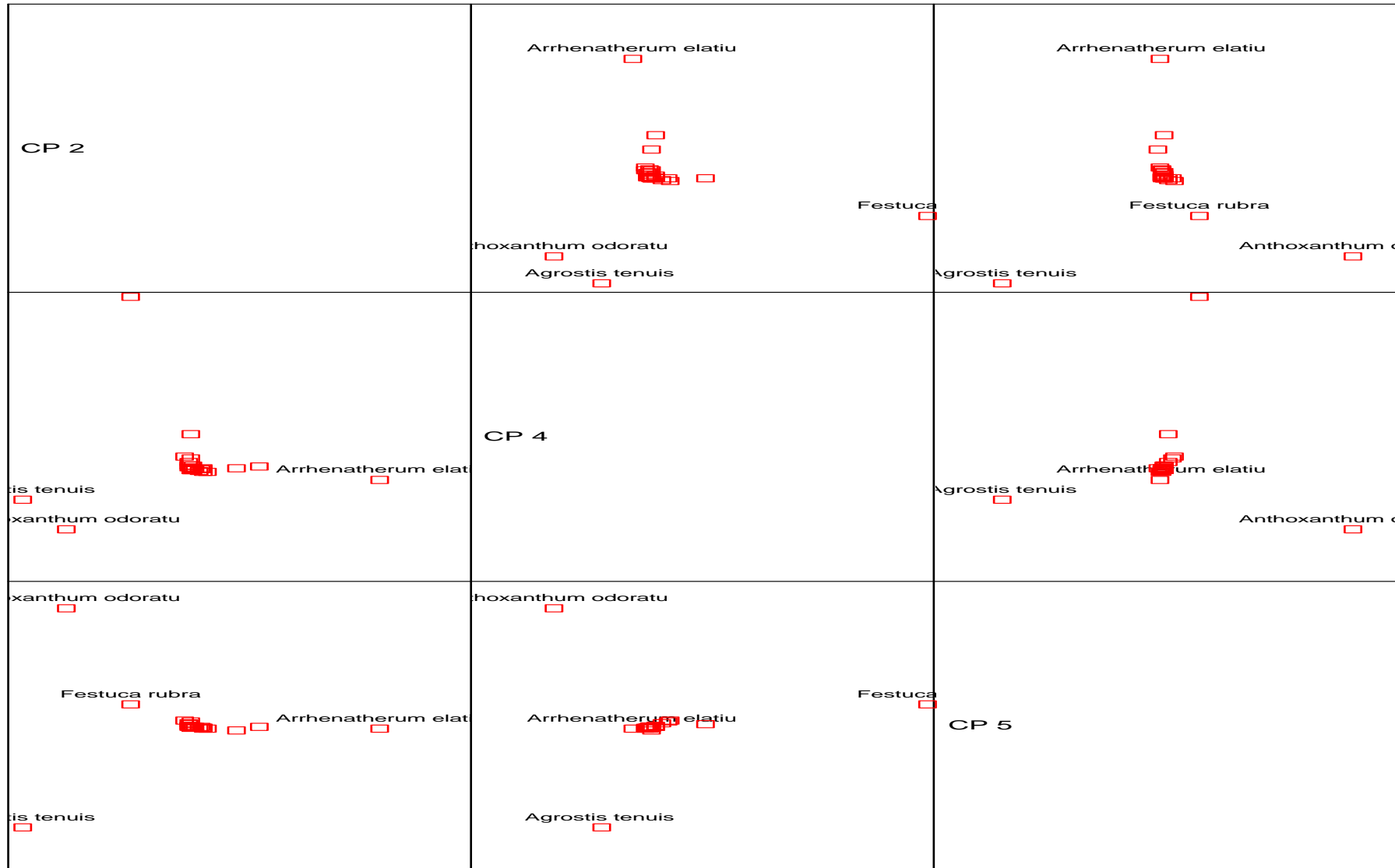


- La primera componente es una componente de tamaño.

Componentes principales I - Diagramas de dispersión



Componentes principales I - Diagramas de dispersión



Componentes principales I - Diagramas de dispersión

- La componente principal 2 separa las especies más frecuentes en dos grupos: (i) *Arrhenatherum elatius*, *Alopecurus pratensis* y *Holcus lanatus*, y (ii) *Festuca rubra*, *Anthoxanthum odoratum* y *Agrostis tenuis*.
- La componente principal 3 separa la especie *Holcus lanatus* del resto.
- La componente principal 4 separa la especie *Arrhenatherum elatius* de la especie *Festuca rubra* y éstas del grupo formado por *Anthoxanthum odoratum* y *Agrostis tenuis*.
- La componente principal 5 separa la especie *Agrostis tenuis* de *Anthoxanthum odoratum*.

Análisis de componentes principales II

Problema: Puesto que el análisis de componentes principales se centra en explicar la variabilidad global de los datos, el énfasis recae en los datos (especies) con mayor variación en la abundancia relativa. ¿Cómo podemos modificar este defecto?

Solución: Para disminuir la influencia de las especies más frecuentes podemos utilizar $z_{ij} = \log(x_{ij} + 1)$ y para eliminar la componente de tamaño:

$$y_{ij} = z_{ij} - z_{i\bullet} - z_{\bullet j} + z_{\bullet\bullet},$$

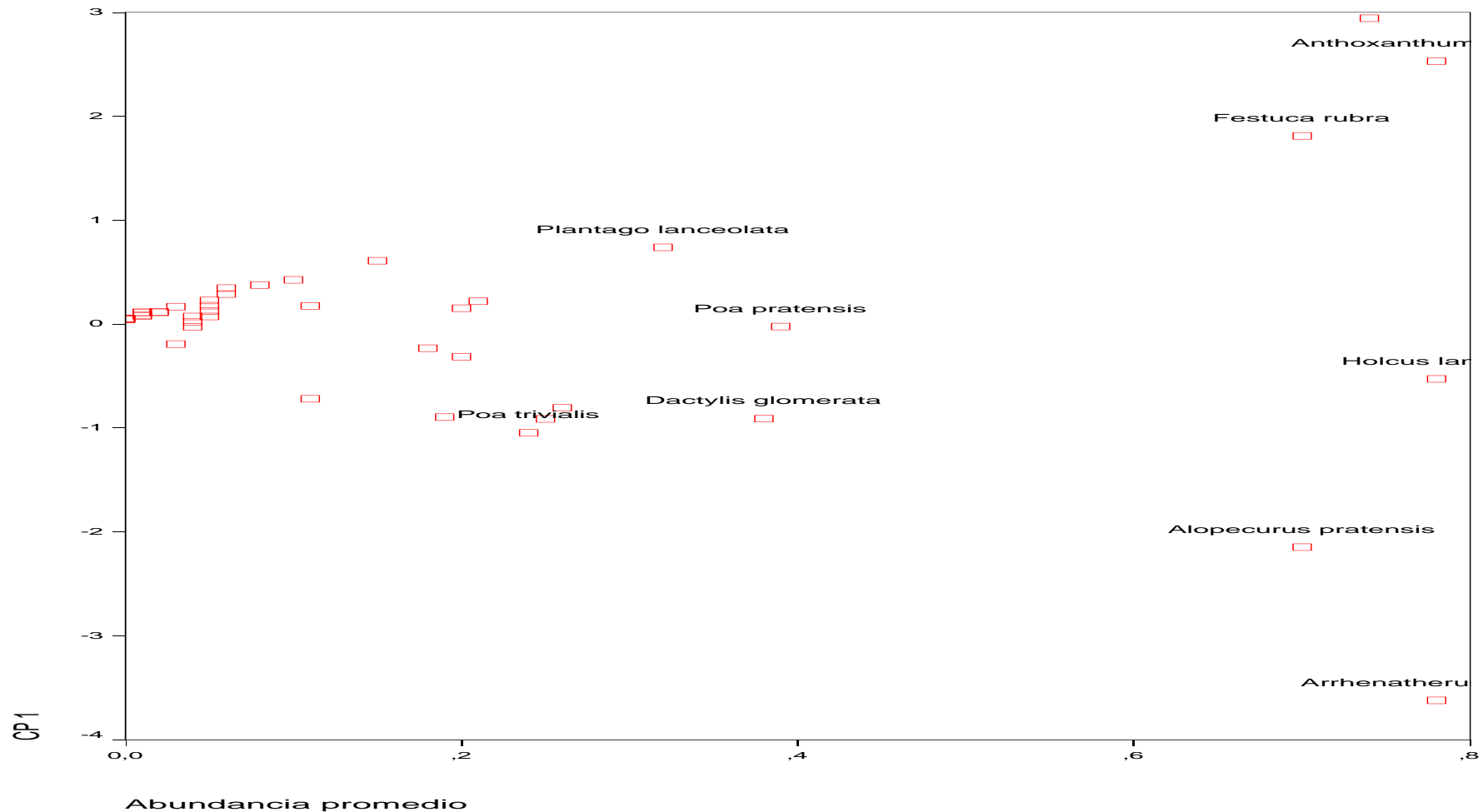
$$\text{donde } \begin{cases} z_{i\bullet} = p^{-1} \sum_{j=1}^p z_{ij}, \\ z_{\bullet j} = n^{-1} \sum_{i=1}^n z_{ij}, \\ z_{\bullet\bullet} = (np)^{-1} \sum_{i=1}^n \sum_{j=1}^p z_{ij}. \end{cases}$$

Análisis de componentes principales II

| | | Autovalores iniciales | | |
|-------|----|-----------------------|---------------------|-------------|
| | | Total | % de la varianza | % acumulado |
| Bruta | 1 | 1,292 | 40,386 | 40,386 |
| | 2 | ,404 | 12,630 | 53,016 |
| | 3 | ,354 | 11,064 | 64,080 |
| | 4 | ,308 | 9,636 | 73,716 |
| | 5 | ,169 | 5,296 | 79,011 |
| | 6 | ,136 | 4,245 | 83,257 |
| | 7 | ,101 | 3,166 | 86,423 |
| | 8 | 7,696E-02 | 2,405 | 88,828 |
| | 9 | 6,482E-02 | 2,026 | 90,854 |
| | 10 | 5,241E-02 | 1,638 | 92,491 |
| | ⋮ | ⋮ | ⋮ | ⋮ |

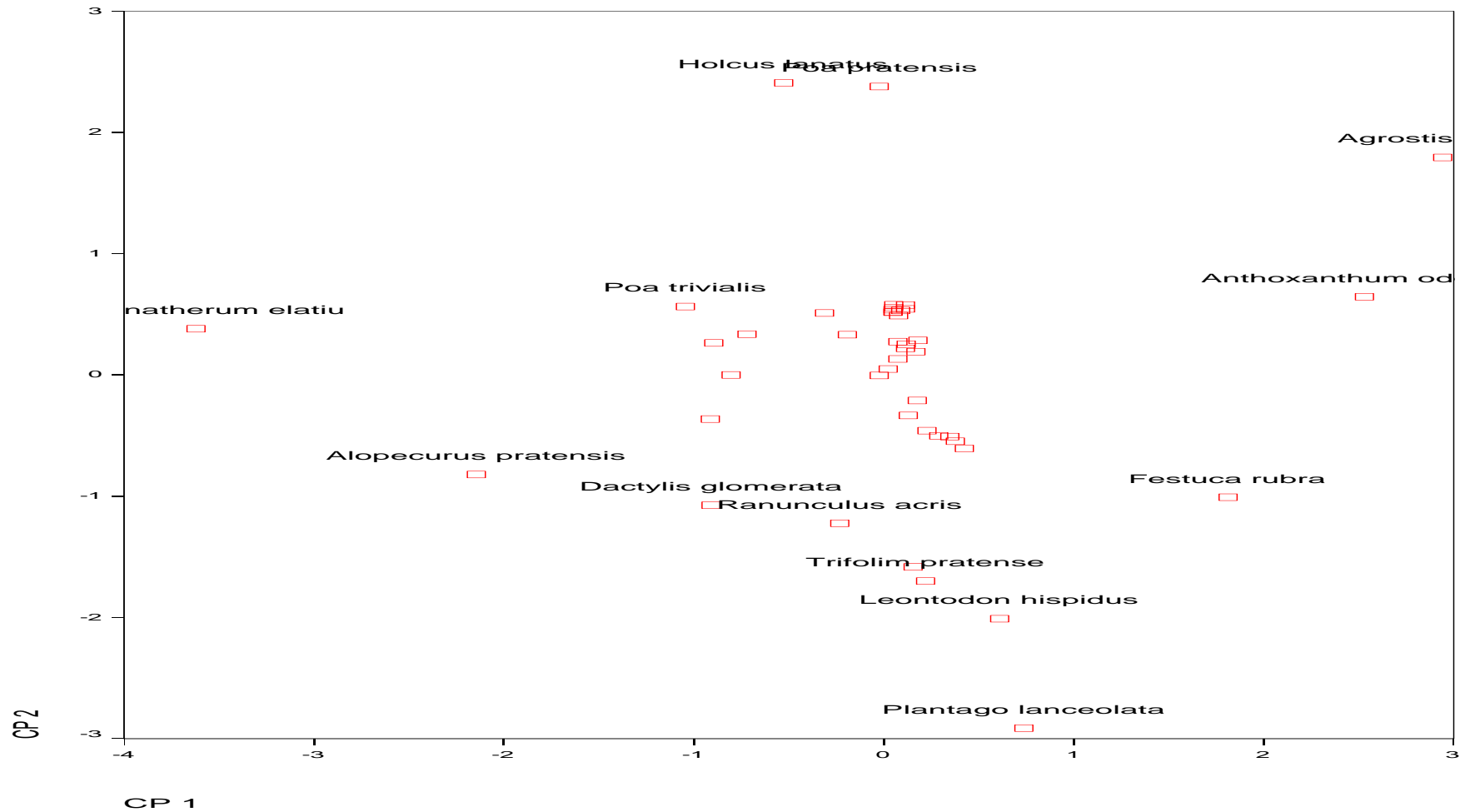
- El criterio de la variabilidad explicada mayor que el 80% sugiere utilizar 5 ó 6 componentes.
- El criterio del valor propio sugiere utilizar 7 componentes.
 - La varianza explicada en este caso es menor que la explicada por las cinco componentes anteriores. Esto manifiesta que más especies están contribuyendo al patrón de variación explicado por las nuevas componentes.

Componentes principales II - Diagramas de dispersión



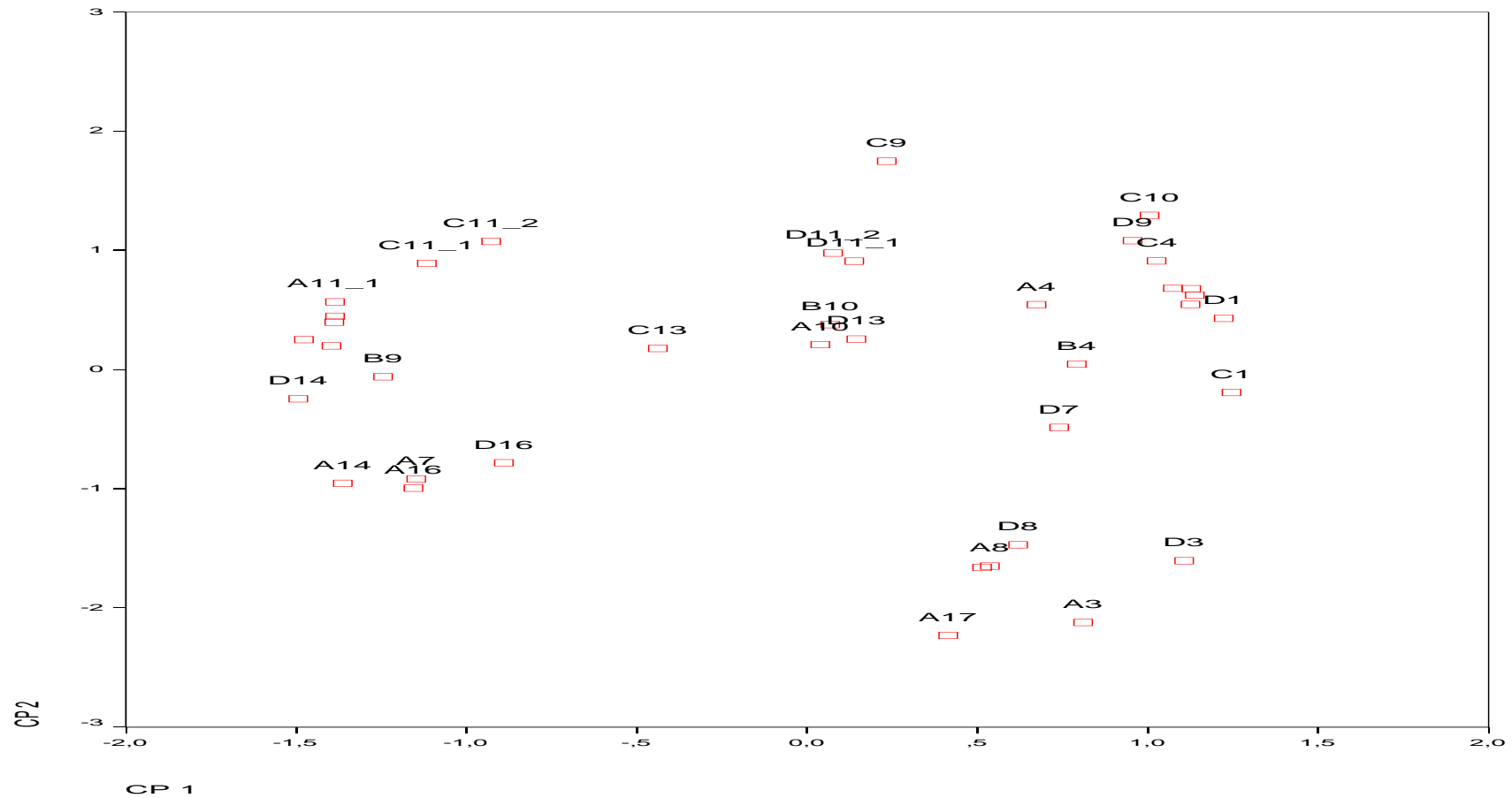
- La primera componente NO es una componente de tamaño.

Componentes principales II - Diagramas de dispersión



Ejercicio 1

Componentes principales III - Matriz X'



- El pH de las parcelas aumenta en la dirección definida por el vector $(1,1)$, y la productividad aumenta en la dirección definida por el vector $(-1,1)$.