

Ejercitación

Componentes principales

1

El decatlón es una prueba olímpica que comprende diez pruebas (cuatro carreras, tres lanzamientos y tres saltos), se disputa en dos días consecutivos, siguiendo un orden establecido.

Los resultados de un grupo de deportistas hombres de alto rendimiento en las diferentes pruebas se encuentran en la base de datos decatlón

Variables

100m: Carrera de 100 metros

saltL: Salto largo

lanzB: Lanzamiento de la bala

saltoA: Salto alto

400m: Carrera de 400 metros

100mV: Carrera de 100 metros con vallas

disco: Lanzamiento del disco

saltoP: salto con garrocha

jav: Lanzamiento de la javalina

1500m: Carrera de 1500 metros

Un entrenador deportivo desea analizar los datos y determinar:

- (1) Cuáles son las pruebas que han sido más importantes en el desempeño de los competidores
- (2) Según su desempeño si se pueden encontrar deportistas que se asemejen o diferencia
- (3) Caracterizar los elementos más relevantes en el desempeño de cada deportista

Para su análisis decide utilizar el método de componentes principales

Veamos algunos componentes de este análisis

2

Pregunta 1

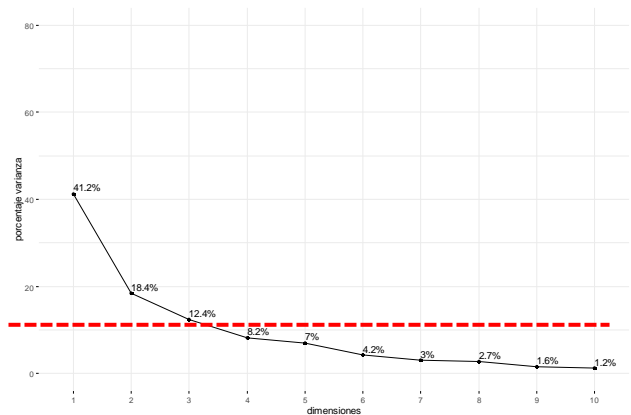
- (a) Utilizando variables estandarizadas construya las componentes principales del problema, y obtenga las varianzas de cada componente y las varianzas acumuladas. Utilice esta información junto con el diagrama de pendiente para valorar si se podría realizar el análisis con 3 dimensiones.
- (b) Una vez decidida la cantidad de componentes(3) escriba las ecuaciones que definen las componentes principales respectivas (recuerde que las variables en la ecuación representan variables estandarizadas), en cada caso acompañe la ecuación con su varianza y la varianza que se acumula al considerarla
- (c) En el análisis de 3 dimensiones ¿cuáles serían las coordenadas del competidor **BOURGUIGNON** que se encuentra en la posición 11 ¿Estas coordenadas qué significan?

(d) Si hubiera utilizando dos dimensiones, es decir, el plano principal formado por las dos primeras componentes [Y1,Y2], ¿cuál el porcentaje de la inercia total explicada por este plano?

3

```
> var_res$coord
```

	Dim.1	Dim.2	Dim.3
100m	-0.850625692	-0.17939806	0.30155643
saltL	0.794180641	0.28085695	-0.19054653
lanzB	0.733912733	0.08540412	0.51759781
saltoA	0.610083985	-0.46521415	0.33008517
400m	-0.701603377	0.29017826	0.28353292
100mV	-0.764125197	-0.02474081	0.44888733
disco	0.743209016	0.04966086	0.17652518
saltoP	-0.217268042	0.80745110	0.09405773
jav	0.428226639	0.38610928	0.60412432
1500m	0.004278487	0.78448019	-0.21947068



```
> vp
```

	eigenvalue	variance.percent	cumulative.variance.percent
Dim.1	4.1242133	41.242133	41.24213
Dim.2	1.8385309	18.385309	59.62744
Dim.3	1.2391403	12.391403	72.01885
Dim.4	0.8194402	8.194402	80.21325
Dim.5	0.7015528	7.015528	87.22878
Dim.6	0.4228828	4.228828	91.45760
Dim.7	0.3025817	3.025817	94.48342
Dim.8	0.2744700	2.744700	97.22812
Dim.9	0.1552169	1.552169	98.78029
Dim.10	0.1219710	1.219710	100.00000

4

(d) Calcule el coseno cuadrado de los ángulos entre las variables y los ejes, así como la contribución de cada variable a los ejes principales.

(d.1) ¿Cuál es la correlación entre la variable carrera de 100m y el eje 1?

(d.2) En el caso anterior, utilizando el cuadrado del coseno del ángulo entre los vectores, ¿qué puede decir acerca de la relación entre la variable y el eje?

(d.3) ¿Cuál es la contribución de la variable 100m al eje 1?

(d.4) ¿Será importante la variable 100m para la explicación del eje 1? O ¿hay algún otro eje para el que sea más importante?

5

```
> round(var_res$cor[,1:3],2)
```

	Dim.1	Dim.2	Dim.3
100m	-0.85	-0.18	0.30
saltL	0.79	0.28	-0.19
lanzB	0.73	0.09	0.52
saltoA	0.61	-0.47	0.33
400m	-0.70	0.29	0.28
100mV	-0.76	-0.02	0.45
disco	0.74	0.05	0.18
saltoP	-0.22	0.81	0.09
jav	0.43	0.39	0.60
1500m	0.00	0.78	-0.22

```
> round(var_res$contrib[,1:3])
```

	Dim.1	Dim.2	Dim.3
100m	17.5	1.8	7.3
saltL	15.3	4.3	2.9
lanzB	13.1	0.4	21.6
saltoA	9.0	11.8	8.8
400m	11.9	4.6	6.5
100mV	14.2	0.0	16.3
disco	13.4	0.1	2.5
saltoP	1.1	35.5	0.7
jav	4.4	8.1	29.5
1500m	0.0	33.5	3.9

```
> cbind(round(var_res$cos2
```

	Dim.1	Dim.2	Dim.3
100m	0.724	0.032	0.091
saltL	0.631	0.079	0.036
lanzB	0.539	0.007	0.268
saltoA	0.372	0.216	0.109
400m	0.492	0.084	0.080
100mV	0.584	0.001	0.201
disco	0.552	0.002	0.031
saltoP	0.047	0.652	0.009
jav	0.183	0.149	0.365
1500m	0.000	0.615	0.048

6

(e) En el plano principal
(1,2), específicamente en el eje 1

(e.1) ¿cuáles son las variables
con mayor influencia?

(e.2) Caracterice el significado de
los valores numéricos del primer
eje principal

7

correlación

```
> round(var_res$cor[,1:3],2)
      Dim.1 Dim.2 Dim.3
100m  -0.85 -0.18  0.30
saltL  0.79  0.28 -0.19
lanzB  0.73  0.09  0.52
saltoA  0.61 -0.47  0.33
400m   -0.70  0.29  0.28
100mV  -0.76 -0.02  0.45
disco   0.74  0.05  0.18
saltoP -0.22  0.81  0.09
jav     0.43  0.39  0.60
1500m   0.00  0.78 -0.22
```

cos2

```
> cbind(round(var_res$cos2
      Dim.1 Dim.2 Dim.3
100m  0.724 0.032 0.091
saltL 0.631 0.079 0.036
lanzB 0.539 0.007 0.268
saltoA 0.372 0.216 0.109
400m  0.492 0.084 0.080
100mV 0.584 0.001 0.201
disco 0.552 0.002 0.031
saltoP 0.047 0.652 0.009
jav   0.183 0.149 0.365
1500m 0.000 0.615 0.048
```

```
> cos(pi/4)^2
[1] 0.5
```

contribución

```
> round(var_res$contrib[,1:3])
      Dim.1 Dim.2 Dim.3
100m    17.5   1.8   7.3
saltL   15.3   4.3   2.9
lanzB   13.1   0.4  21.6
saltoA    9.0  11.8   8.8
400m    11.9   4.6   6.5
100mV   14.2   0.0  16.3
disco    13.4   0.1   2.5
saltoP    1.1  35.5   0.7
jav       4.4   8.1  29.5
1500m    0.0  33.5   3.9
```

100mV no está muy claro,
pero tiene mayor correlación
con eje 1 y eje 3 poco imp

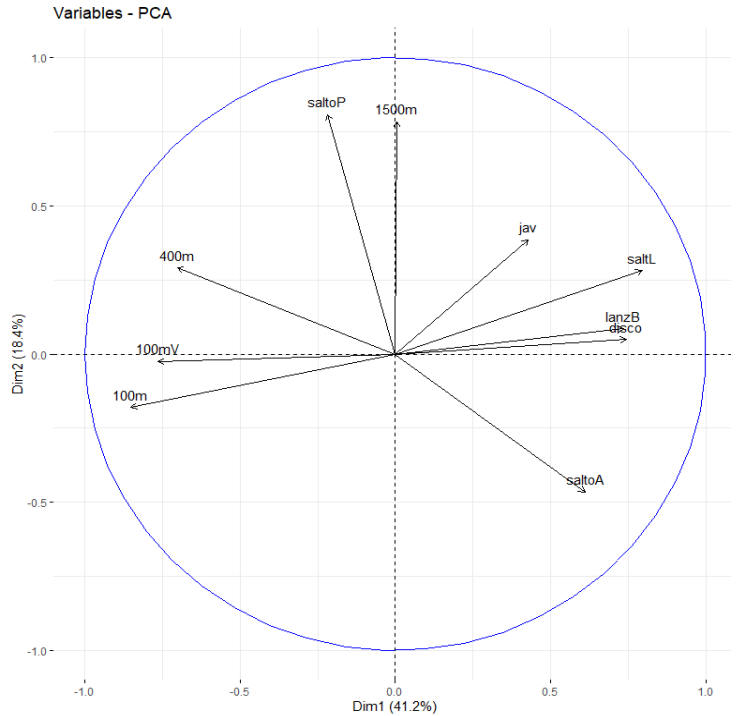
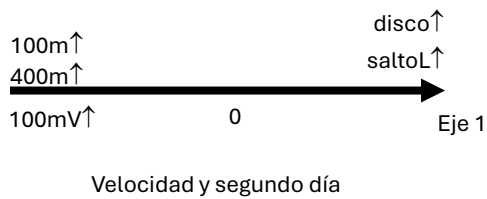
Veamos los diagramas circulares

8

contribución

```
> round(var_res$contrib[,1:3])
```

	Dim.1	Dim.2	Dim.3
100m	17.5	1.8	7.3
saltL	15.3	4.3	2.9
lanzB	13.1	0.4	21.6
saltoA	9.0	11.8	8.8
400m	11.9	4.6	6.5
100mV	14.2	0.0	16.3
disco	13.4	0.1	2.5
saltoP	1.1	35.5	0.7
jav	4.4	8.1	29.5
1500m	0.0	33.5	3.9



9

(e) En el plano principal
(1,2), específicamente en el eje 2

(e.1) ¿cuáles son las variables
con mayor influencia?

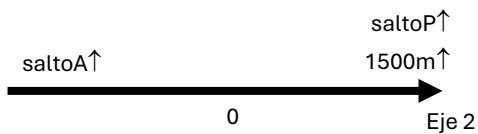
(e.2) Caracterice el significado de
los valores numéricos del primer
eje principal

10

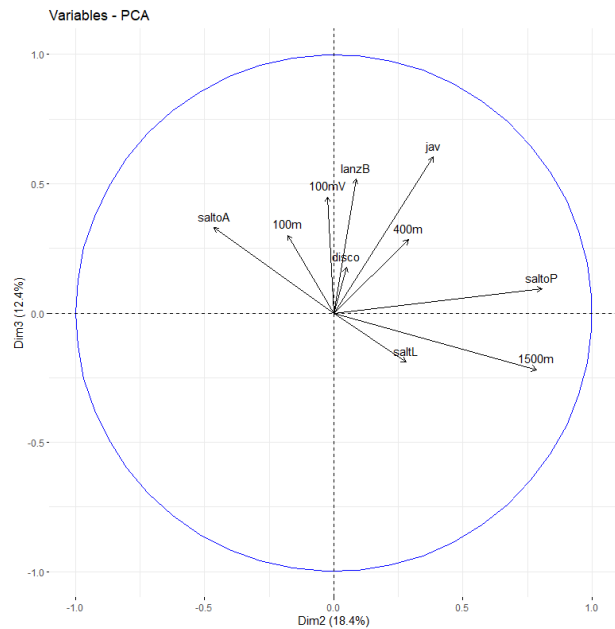
contribución

```
> round(var_res$contrib[,1:3])
```

	Dim.1	Dim.2	Dim.3
100m	17.5	1.8	7.3
saltoL	15.3	4.3	2.9
lanzB	13.1	0.4	21.6
saltoA	9.0	11.8	8.8
400m	11.9	4.6	6.5
100mV	14.2	0.0	16.3
disco	13.4	0.1	2.5
saltoP	1.1	35.5	0.7
jav	4.4	8.1	29.5
1500m	0.0	33.5	3.9



Salto y carrera de fondo



Las correlaciones son más pequeñas

11

(f.1) Utilice la representación de los individuos en los ejes [1,2] y comente el desempeño de **BOURGUIGNON** (posición 11 base datos) ¿estará BOURGUIGNON bien representado en el plano [1,2]?

12

Los individuos en los ejes principales

```
> round(ind$contrib[,1:3],2)
```

	Dim.1	Dim.2	Dim.3
SEBRLE	0.04	5.97	1.45
CLAY	0.69	14.48	6.75
BERNARD	1.95	6.42	0.14
YURKOV	0.83	0.46	22.45
ZSIVOCZKY	0.01	10.12	6.25
McMULLEN	0.02	2.43	2.61
MARTINEAU	6.31	4.08	2.96
HERNU	3.94	3.43	0.09
BARRAS	3.47	0.42	1.59
NOOL	8.50	6.11	1.35
BOURGUIGNON	21.47	3.95	0.00
sebrle	13.13	3.66	10.33
clay	12.12	6.51	1.40
Karpov	18.26	1.58	3.78
Macey	3.77	10.53	3.52
Warners	2.12	0.20	14.67
Zsivoczky	0.25	2.12	10.96
Hernu	0.08	3.36	0.11
Bernard	2.06	5.40	2.54
Schwarzl	0.56	1.07	3.97
Pogorelov	0.05	1.83	0.32
Schoenbeck	0.27	4.18	0.04
Barras	0.11	1.66	2.72

Analizar ángulo de forma complementaria

```
> round(ind$cos2[,1:3],2)
```

	Dim.1	Dim.2	Dim.3
SEBRLE	0.01	0.50	0.08
CLAY	0.05	0.46	0.14
BERNARD	0.20	0.29	0.00
YURKOV	0.10	0.02	0.78
ZSIVOCZKY	0.00	0.58	0.24
McMULLEN	0.00	0.15	0.11
MARTINEAU	0.40	0.12	0.06
HERNU	0.40	0.16	0.00
BARRAS	0.62	0.03	0.08
NOOL	0.49	0.16	0.02
BOURGUIGNON	0.86	0.07	0.00
sebrle	0.68	0.08	0.16
clay	0.69	0.16	0.02
Karpov	0.78	0.03	0.05
Macey	0.36	0.45	0.10
Warners	0.26	0.01	0.53
Zsivoczky	0.05	0.17	0.61
Hernu	0.02	0.44	0.01
Bernard	0.29	0.34	0.11
Schwarzl	0.12	0.10	0.25
Pogorelov	0.01	0.13	0.02
Schoenbeck	0.07	0.46	0.00
Barras	0.02	0.13	0.14

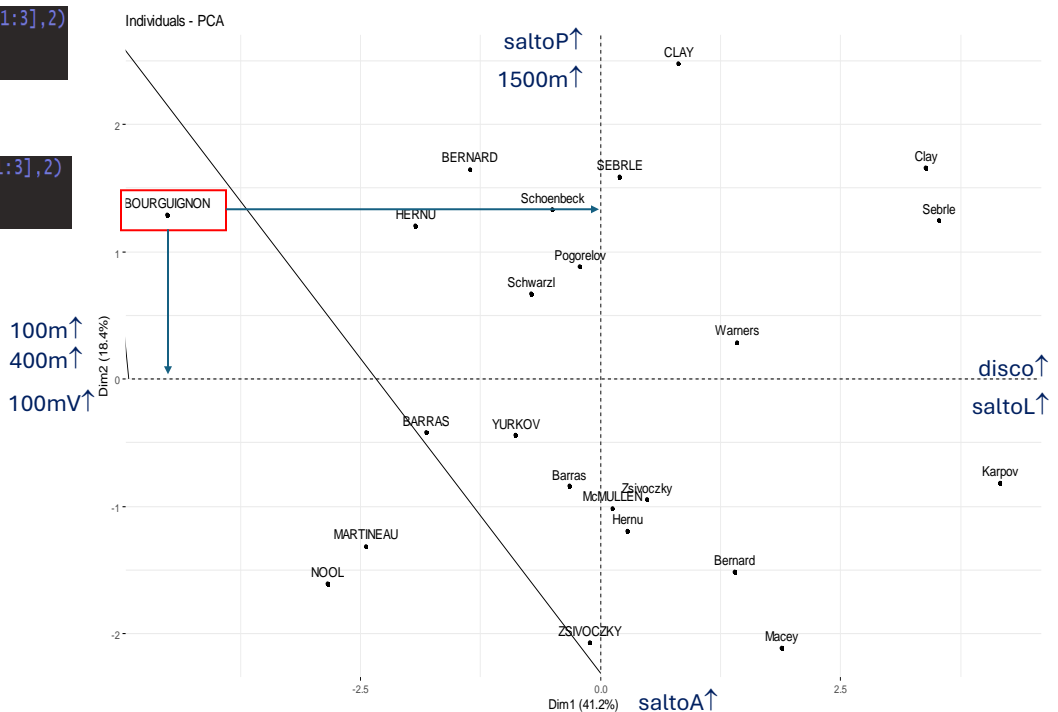
13

```
> round(ind$contrib[11,1:3],2)
```

Dim.1	Dim.2	Dim.3
21.47	3.95	0.00

```
> round(ind$cos2[11,1:3],2)
```

Dim.1	Dim.2	Dim.3
0.86	0.07	0.00



El eje 3 no es importante en este caso

14

(f.2) Utilice la representación de los individuos en los ejes [1,2] y comente el desempeño de **Sebrle** (con minúsculas) (posición 12 base datos)
¿estará Sebrle bien representado en el plano [1,2]?

15

Los individuos en los ejes principales

```
> round(ind$contrib[,1:3],2)
```

	Dim.1	Dim.2	Dim.3
SEBRLE	0.04	5.97	1.45
CLAY	0.69	14.48	6.75
BERNARD	1.95	6.42	0.14
YURKOV	0.83	0.46	22.45
ZSIVOCZKY	0.01	10.12	6.25
McMULLEN	0.02	2.43	2.61
MARTINEAU	6.31	4.08	2.96
HERNU	3.94	3.43	0.09
BARRAS	3.47	0.42	1.59
NOOL	8.50	6.11	1.35
BOURGUIGNON	21.47	3.95	0.00
sebrle	13.13	3.66	10.33
clay	12.12	6.51	1.40
Karpov	18.26	1.58	3.78
Macey	3.77	10.53	3.52
Warners	2.12	0.20	14.67
Zsivoczky	0.25	2.12	10.96
Hernu	0.08	3.36	0.11
Bernard	2.06	5.40	2.54
Schwarzl	0.56	1.07	3.97
Pogorelov	0.05	1.83	0.32
Schoenbeck	0.27	4.18	0.04
Barras	0.11	1.66	2.72

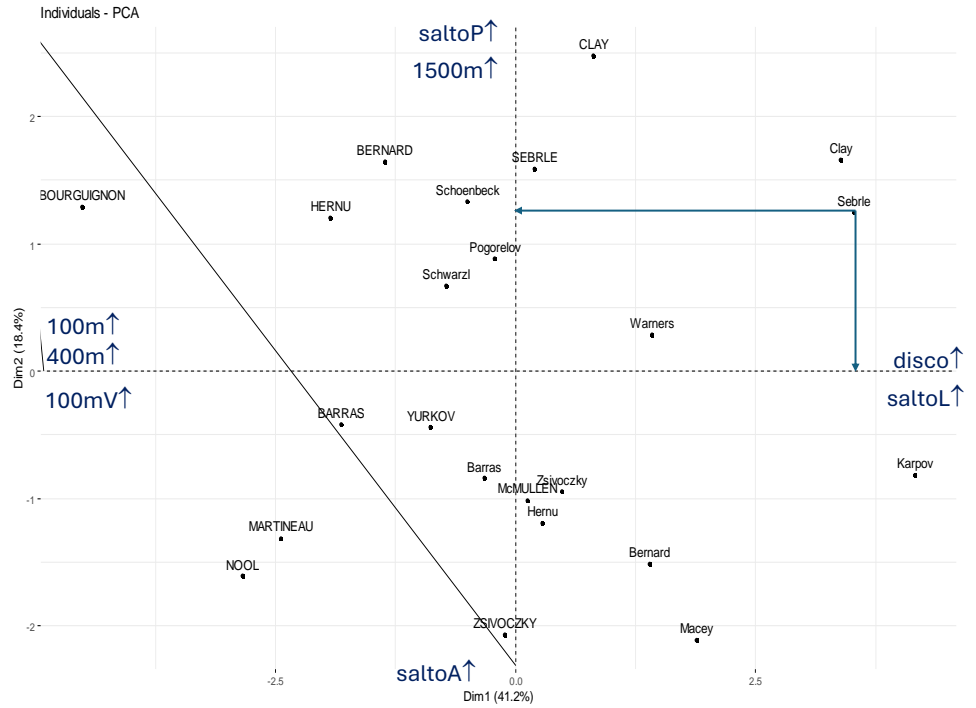
16

Analizar ángulo de forma complementaria

```
> round(ind$cos2[,1:3],2)
```

	Dim.1	Dim.2	Dim.3
SEBRLE	0.01	0.50	0.08
CLAY	0.05	0.46	0.14
BERNARD	0.20	0.29	0.00
YURKOV	0.10	0.02	0.78
ZSIVOCZKY	0.00	0.58	0.24
McMULLEN	0.00	0.15	0.11
MARTINEAU	0.40	0.12	0.06
HERNU	0.40	0.16	0.00
BARRAS	0.62	0.03	0.08
NOOL	0.49	0.16	0.02
BOURGUIGNON	0.86	0.07	0.00
Sebrle	0.68	0.08	0.16
Clay	0.69	0.16	0.02
Karpov	0.78	0.03	0.05
Macey	0.36	0.45	0.10
Warners	0.26	0.01	0.53
Zsivoczky	0.05	0.17	0.61
Hernu	0.02	0.44	0.01
Bernard	0.29	0.34	0.11
Schwarzl	0.12	0.10	0.25
Pogorelov	0.01	0.13	0.02
Schoenbeck	0.07	0.46	0.00
Barras	0.02	0.13	0.14

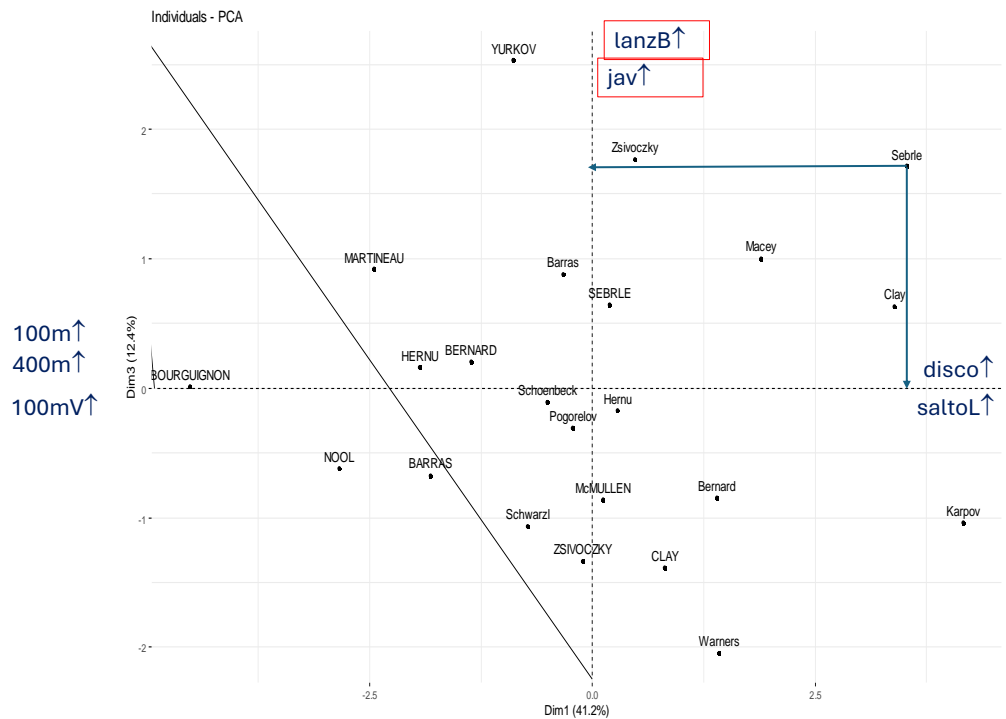

```
> #Sebrle
> round(ind$contrib
Dim.1 Dim.2 Dim.3
13.13  3.66 10.33
> round(ind$cos2[12
Dim.1 Dim.2 Dim.3
0.68  0.08 0.16
```



Pero también el eje
1,3

17

```
> #Sebrle
> round(ind$contrib
Dim.1 Dim.2 Dim.3
13.13  3.66 10.33
> round(ind$cos2[12
Dim.1 Dim.2 Dim.3
0.68  0.08 0.16
```



18

Ejercicio

19

El archivo `alim` contiene información sobre datos del consumo per cápita de ciertos alimentos en 29 países europeos alrededor de 1970. Adicionalmente se dispone de información sobre el régimen político de cada país. Las variables consideradas son:

```
1 Red_Meat
2 White_Meat
3 Eggs
4 Milk
5 Fish
6 Cereal
7 Starch
8 Nuts
9 Fruits_Vegetables
```

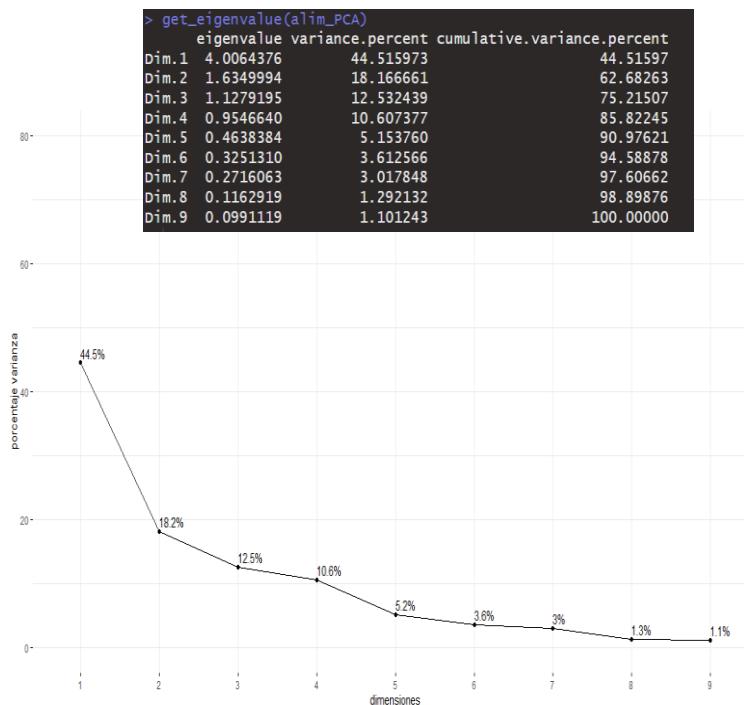
20

Pregunta 1

- (a) Utilizando variables estandarizadas construya las componentes principales del problema, y obtenga las varianzas de cada componente y las varianzas acumuladas. Utilice esta información junto con el diagrama de pendiente para valorar si se podría realizar el análisis con 3 dimensiones.
- (b) Una vez decidida la cantidad de componentes(3) escriba las ecuaciones que definen las componentes principales respectivas (recuerde que las variables en la ecuación representan variables estandarizadas), en cada caso acompañe la ecuación con su varianza y la varianza que se acumula al considerarla
- (c) En el análisis de 3 dimensiones ¿cuáles serían las coordenadas de **Portugal** que se encuentra en la posición 17 ¿Estas coordenadas qué significan?
- (d) Si hubiera utilizando dos dimensiones, específicamente decir, el plano principal formado por la primera y la tercera componente [Y1,Y3], ¿cuál el porcentaje de la inercia total explicada por este plano?

21

- (d) Si hubiera utilizando dos dimensiones, específicamente decir, el plano principal formado por la primera y la tercera componente [Y1,Y3], ¿cuál el porcentaje de la inercia total explicada por este plano?



22

(c) En el análisis de 3 dimensiones ¿cuáles serían las coordenadas de **Portugal** que se encuentra en la posición 17 ¿Estas coordenadas qué significan?

```
> #individuos
> ind=get_pca_ind(alim_PCA)
> #coordenadas
> ind$coord[17,1:3]
```

	Dim.1	Dim.2	Dim.3
	-1.74102993	4.37737493	0.04453254

23

(d) Calcule el coseno cuadrado de los ángulos entre las variables y los ejes, así como la contribución de cada variable a los ejes principales.

(d.1) ¿Cuál es la correlación entre la variable Red_Meat y el eje 3?

(d.2) En el caso anterior, utilizando el cuadrado del coseno del ángulo entre los vectores, ¿qué puede decir acerca de la relación entre la variable y el eje?

(d.3) ¿Cuál es la contribución de la variable Red_Meat al eje 1?

(d.4) ¿Será importante la variable Red_Meat para la explicación del eje 1, para el 3? O ¿hay algún otro eje para el que sea más importante?

24

```
> round(var_alim$contrib[,1:3],2)
```

	Dim.1	Dim.2	Dim.3
Red_Meat	9.16	0.32	8.86
White_Meat	9.64	5.61	38.92
Eggs	18.21	0.12	3.30
Milk	14.27	3.41	14.87
Fish	1.84	41.84	10.32
Cereal	19.16	5.45	0.92
Starch	8.84	12.45	5.90
Nuts	17.67	2.05	0.30
Fruits_Vegetables	1.22	28.75	16.61

```
> round(var_alim$cor[,1:3],2)
```

	Dim.1	Dim.2	Dim.3
Red_Meat	0.61	-0.07	-0.32
White_Meat	0.62	-0.30	0.66
Eggs	0.85	-0.05	0.19
Milk	0.76	-0.24	-0.41
Fish	0.27	0.83	-0.34
Cereal	-0.88	-0.30	0.10
Starch	0.59	0.45	0.26
Nuts	-0.84	0.18	-0.06
Fruits_Vegetables	-0.22	0.69	0.43

```
> round(var_alim$cos2[,1:3],2)
```

	Dim.1	Dim.2	Dim.3
Red_Meat	0.37	0.01	0.10
White_Meat	0.39	0.09	0.44
Eggs	0.73	0.00	0.04
Milk	0.57	0.06	0.17
Fish	0.07	0.68	0.12
Cereal	0.77	0.09	0.01
Starch	0.35	0.20	0.07
Nuts	0.71	0.03	0.00
Fruits_Vegetables	0.05	0.47	0.19

(d) Calcule el coseno cuadrado de los ángulos entre las variables y los ejes, así como la contribución de cada variable a los ejes principales.

(d.1) ¿Cuál es la correlación entre la variable Red_Meat y el eje 3?

(d.2) En el caso anterior, utilizando el cuadrado del coseno del ángulo entre los vectores, ¿qué puede decir acerca de la relación entre la variable y el eje?

25

```
> round(var_alim$contrib[,1:3],2)
```

	Dim.1	Dim.2	Dim.3
Red_Meat	9.16	0.32	8.86
White_Meat	9.64	5.61	38.92
Eggs	18.21	0.12	3.30
Milk	14.27	3.41	14.87
Fish	1.84	41.84	10.32
Cereal	19.16	5.45	0.92
Starch	8.84	12.45	5.90
Nuts	17.67	2.05	0.30
Fruits_Vegetables	1.22	28.75	16.61

```
> round(var_alim$cor[,1:3],2)
```

	Dim.1	Dim.2	Dim.3
Red_Meat	0.61	-0.07	-0.32
White_Meat	0.62	-0.30	0.66
Eggs	0.85	-0.05	0.19
Milk	0.76	-0.24	-0.41
Fish	0.27	0.83	-0.34
Cereal	-0.88	-0.30	0.10
Starch	0.59	0.45	0.26
Nuts	-0.84	0.18	-0.06
Fruits_Vegetables	-0.22	0.69	0.43

```
> round(var_alim$cos2[,1:3],2)
```

	Dim.1	Dim.2	Dim.3
Red_Meat	0.37	0.01	0.10
White_Meat	0.39	0.09	0.44
Eggs	0.73	0.00	0.04
Milk	0.57	0.06	0.17
Fish	0.07	0.68	0.12
Cereal	0.77	0.09	0.01
Starch	0.35	0.20	0.07
Nuts	0.71	0.03	0.00
Fruits_Vegetables	0.05	0.47	0.19

(d.3) ¿Cuál es la contribución de la variable Red_Meat al eje 1?

(d.4) ¿Será importante la variable Red_Meat para la explicación del eje 1, para el 3? O ¿hay algún otro eje para el que sea más importante?

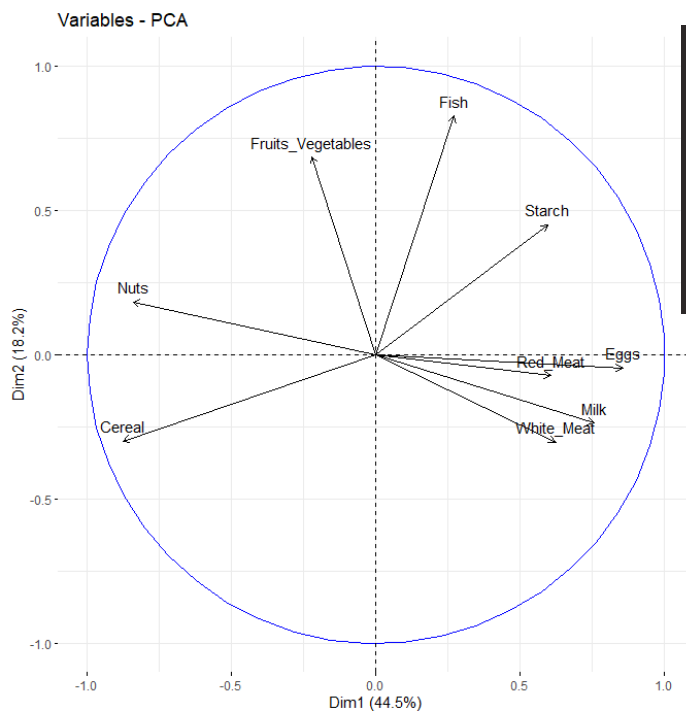
26

(e) En los 3 planos principales
que se van a considerar

(e.1) ¿cuáles son las variables
con mayor influencia?

(e.2) Caracterice el significado de
los valores numéricos de los ejes

27



```
> round(var_alim$contrib[,1:3],2)
```

	Dim.1	Dim.2	Dim.3
Red_Meat	9.16	0.32	8.86
White_Meat	9.64	5.61	38.92
Eggs	18.21	0.12	3.30
Milk	14.27	3.41	14.87
Fish	1.84	41.84	10.32
Cereal	19.16	5.45	0.92
Starch	8.84	12.45	5.90
Nuts	17.67	2.05	0.30
Fruits_Vegetables	1.22	28.75	16.61

```
> round(var_alim$cor[,1:3],2)
```

	Dim.1	Dim.2	Dim.3
Red_Meat	0.61	-0.07	-0.32
White_Meat	0.62	-0.30	0.66
Eggs	0.85	-0.05	0.19
Milk	0.76	-0.24	-0.41
Fish	0.27	0.83	-0.34
Cereal	-0.88	-0.30	0.10
Starch	0.59	0.45	0.26
Nuts	-0.84	0.18	-0.06
Fruits_Vegetables	-0.22	0.69	0.43

28

(f) Utilice la representación de los individuos y caracterice la alimentación de Albania, Francia y Portugal

Paso 1: Valore estos países a qué ejes contribuyen más y a cuáles no

29

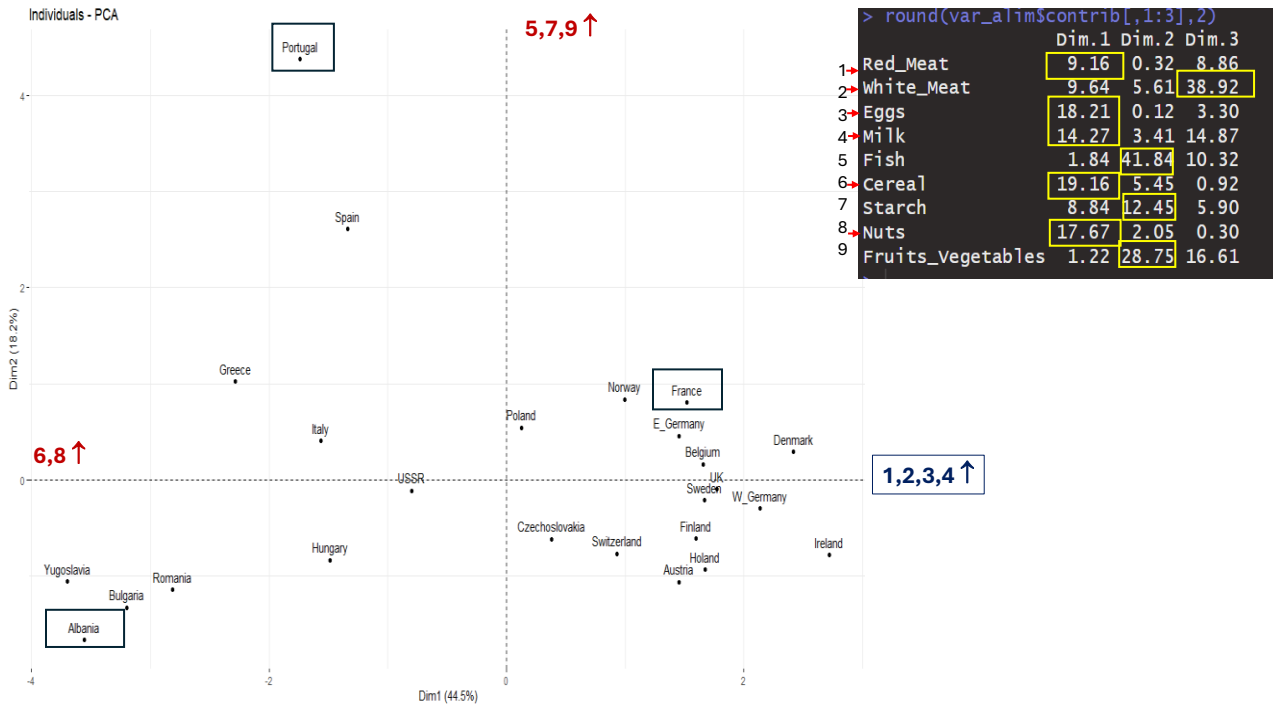
```
> round(ind$contrib[,1:3],2)
```

	Dim.1	Dim.2	Dim.3
Albania	12.63	6.77	11.46
Austria	2.10	2.76	6.61
Belgium	2.74	0.06	0.17
Bulgaria	10.22	4.31	0.08
Czechoslovakia	0.14	0.93	5.28
Denmark	5.82	0.21	2.09
E_Germany	2.10	0.52	6.27
Finland	2.54	0.91	15.52
France	2.30	1.57	0.00
Greece	5.22	2.55	2.88
Hungary	2.21	1.70	13.54
Ireland	7.38	1.49	0.00
Italy	2.45	0.41	0.06
Holand	2.80	2.12	2.17
Norway	0.99	1.72	10.73
Poland	0.02	0.72	8.03
Portugal	3.03	46.88	0.01
Romania	7.90	3.19	0.02
Spain	1.79	16.62	0.98
Sweden	2.78	0.11	6.06
Switzerland	0.87	1.44	0.09
UK	3.13	0.02	4.91
USSR	0.64	0.03	0.50
W_Germany	4.56	0.22	2.39
Yugoslavia	13.65	2.75	0.16

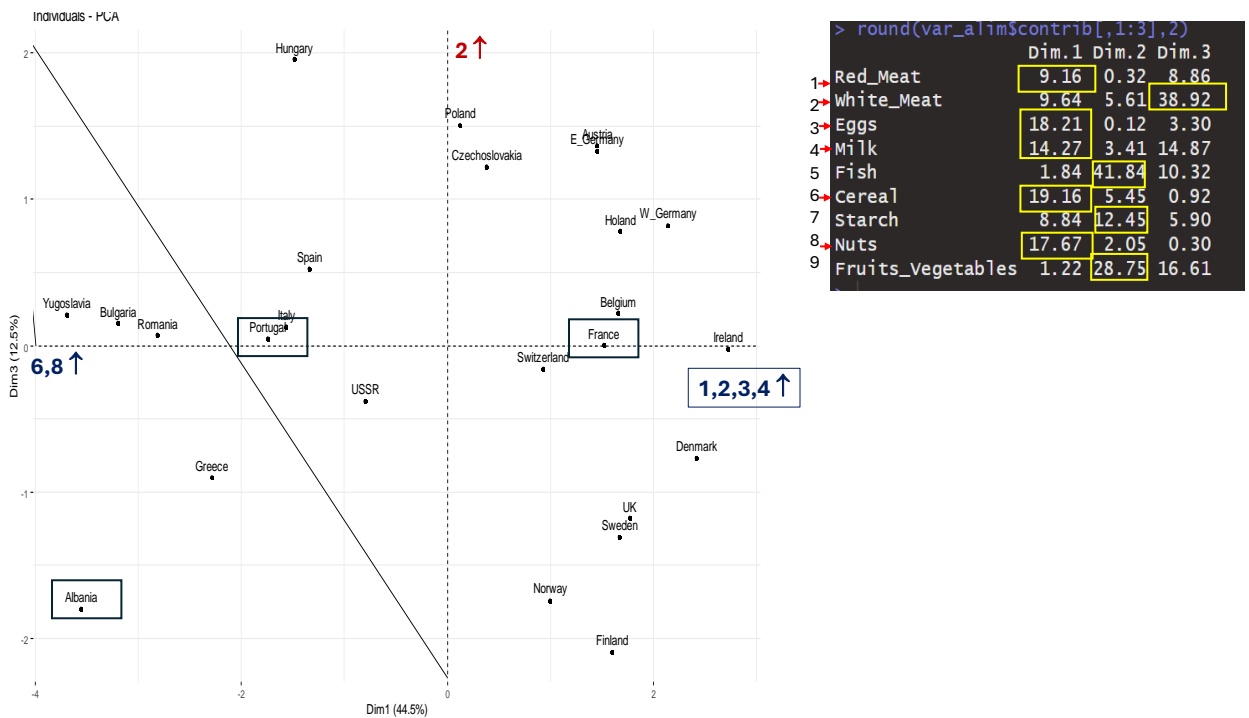
```
> round(ind$cos2[,1:3],2)
```

	Dim.1	Dim.2	Dim.3
Albania	0.61	0.13	0.16
Austria	0.34	0.18	0.30
Belgium	0.71	0.01	0.01
Bulgaria	0.74	0.13	0.00
Czechoslovakia	0.05	0.12	0.47
Denmark	0.66	0.01	0.07
E_Germany	0.33	0.03	0.27
Finland	0.24	0.03	0.41
France	0.25	0.07	0.00
Greece	0.43	0.09	0.07
Hungary	0.28	0.09	0.48
Ireland	0.78	0.06	0.00
Italy	0.45	0.03	0.00
Holand	0.53	0.16	0.12
Norway	0.16	0.11	0.48
Poland	0.00	0.06	0.46
Portugal	0.13	0.80	0.00
Romania	0.80	0.13	0.00
Spain	0.17	0.65	0.03
Sweden	0.45	0.01	0.28
Switzerland	0.21	0.14	0.01
UK	0.33	0.00	0.15
USSR	0.13	0.00	0.03
W_Germany	0.80	0.02	0.12
Yugoslavia	0.86	0.07	0.00

30



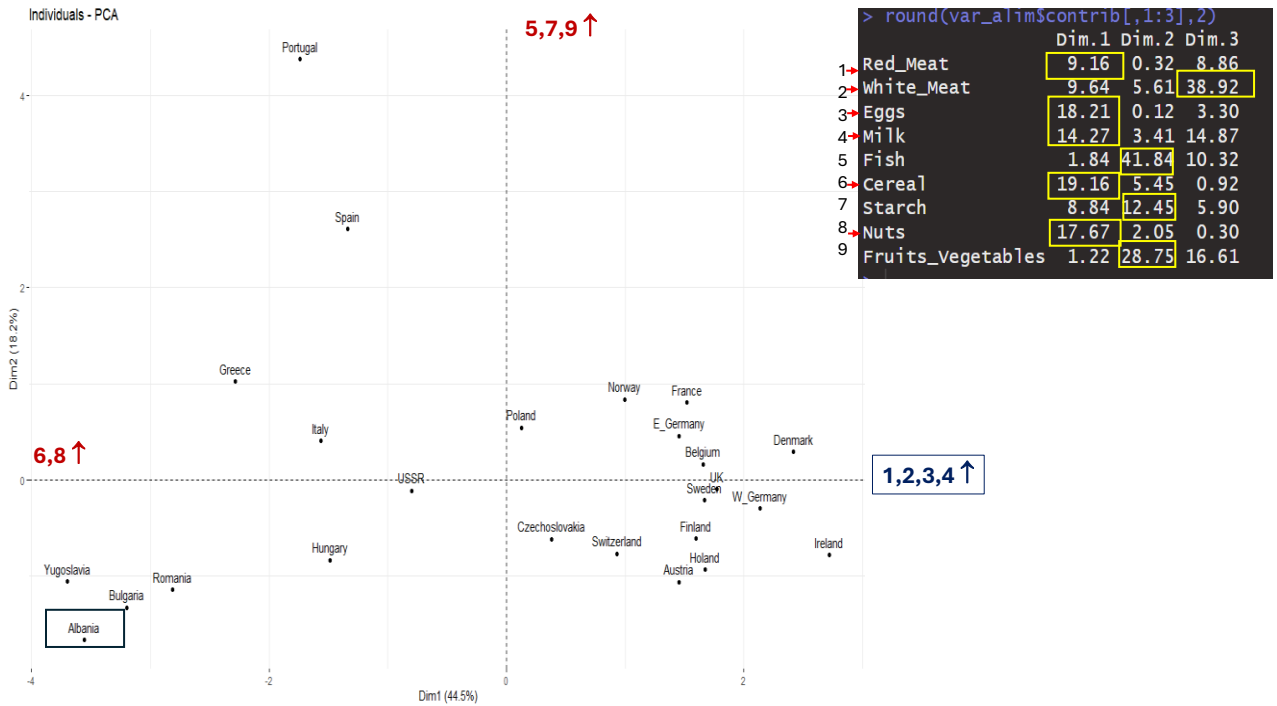
31



32



17



35