

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 $[Y_1, Y_2]$, ¿cuál el porcentaje de la inercia total explicada por este plano?

3

(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?

4

(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

5

(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

6

(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]?

(f.2) Utilice la representación de los individuos en los ejes [1,2] y comente el desempeño de **BARRAS** (posición 9 base datos)
¿estará Barras bien representado en el plano [1,2]?

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

```

9

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 $[Y_1, Y_3]$, ¿cuál el porcentaje de la inercia total explicada por este plano?

10

(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 100m para la explicación del eje 1, para el 3? O ¿hay algún otro eje para el que sea más importante?

11

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

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

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

12

(f) Valore qué
países pudieran
ser
representados (y
analizados) en
el plano 1 - 2

13

(g) Utilice la
representación de los
individuos en los ejes
[1,2] y compare las
características de la
alimentación de
Albania, Francia y
Portugal

(h) ¿Cuáles serían
los países más
parecidos a Portugal
en su alimentación?

(i) ¿Cuáles serían los
países más
parecidos a Albania
en su alimentación?

14

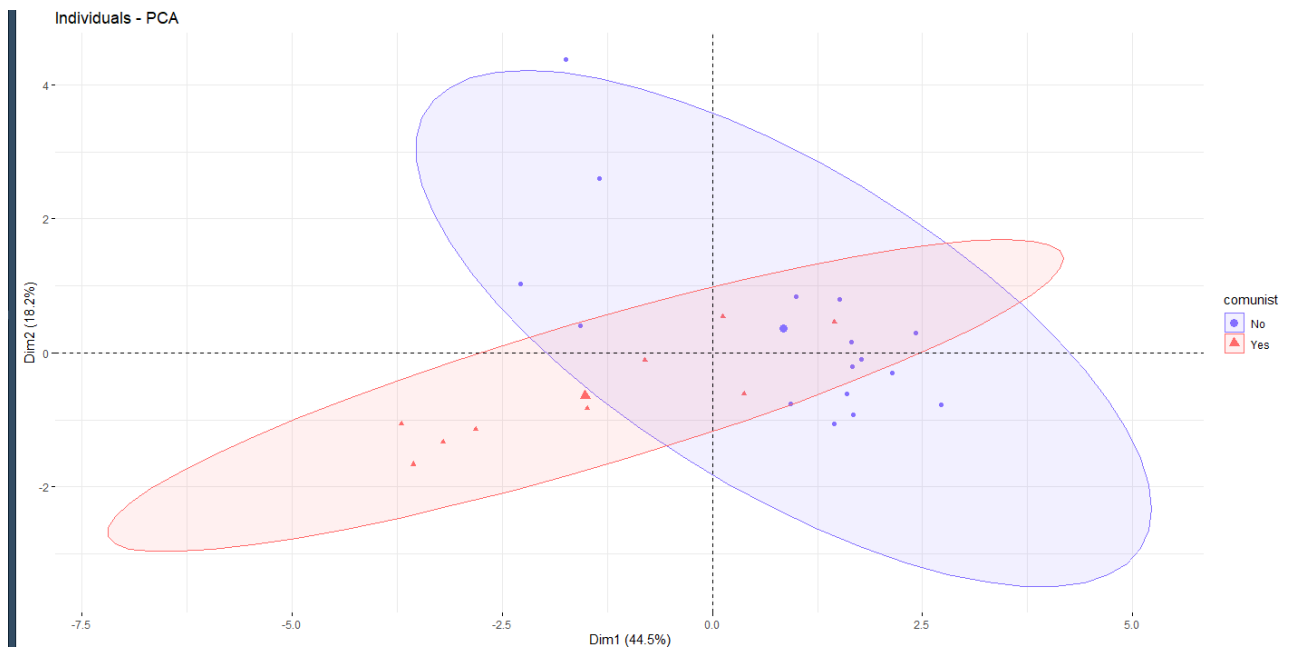
¿Cómo podríamos haber valorado
la pregunta del taller utilizando
componentes principales?

La comparación de grupos puede mejorarse usando

....

Ver macro:
grafico por grupos

15



16