

ACP de “Whisky”

Sebastian Aristizabal, Esteban Avendaño Forero,
Juan David Sarmiento.

October 16, 2020

Objetivo

El objetivo es estudiar la relación calidad-precio de 35 marcas de whisky utilizando las variables precio (francos franceses), proporción de malta (vejez (añejamiento en años) y apreciación (nota promedio de un panel de catadores redondeada a entera). Se dispone además de una variable nominal categoría, que clasifica las marcas según su contenido de malta (1 = Bajo, 2 = Estándar, 3 = Puro malta) (Fine, 1996).

Preguntas

Realice primero un ACP no normado y luego un ACP normado y responda a las preguntas.

1. En el ACP no normado, analice la contribución de las variables a la inercia. ¿Realmente se puede considerar un análisis de las cuatro variables?
2. Analice la matriz de varianzas y covarianzas con la ayuda del primer plano factorial de las variables. Haga un resumen (interpretación del primer plano factorial de las variables).
3. Realice el ACP normado y justifique por qué es el que conviene para los objetivos de este taller.

A continuacion, realizamos el ACP normado con los datos del paquete Whisky de FactoClass.

```
> library(FactoClass)
> data(Whisky)
> names(Whisky)

[1] "price" "malt"  "type"  "aging" "taste"

> Y <- Whisky[,c(1,2,4,5)]
> names(Y)

[1] "price" "malt"  "aging" "taste"
```

```
> acp <- dudi.pca(Y, scannf = FALSE , nf =3)
```

```
> acp
```

Duality diagramm

class: pca dudi

\$call: dudi.pca(df = Y, scannf = FALSE, nf = 3)

\$nf: 3 axis-components saved

\$rank: 4

eigen values: 2.233 0.8065 0.6295 0.3307

	vector	length	mode	content
1	\$cw	4		numeric column weights
2	\$lw	35		numeric row weights
3	\$eig	4		numeric eigen values

	data.frame	nrow	ncol	content
1	\$tab	35	4	modified array
2	\$li	35	3	row coordinates
3	\$l1	35	3	row normed scores
4	\$co	4	3	column coordinates
5	\$c1	4	3	column normed scores

other elements: cent norm

```
> barplot(acp$eig)
```

```
> valp <- t(inertia(acp)$tot.inertia)
```

```
> valp
```

	Ax1	Ax2	Ax3	Ax4
inertia	2.233269	0.8064826	0.6295103	0.3307381
cum	2.233269	3.0397516	3.6692619	4.0000000
cum(%)	55.831725	75.9937904	91.7315485	100.0000000

4. ¿Cuántos ejes retiene para el análisis? ¿Por qué?

De acuerdo al analisis realizado, decidimos retener 3 ejes para el analisis. Esta decision la tomamos al observar la forma del barplot de los valores propios; aqui podemos ver que hay una valor mayor que dos, mientras que los dos siguientes son muy parecidos entre si, mientras que el ultimo valor es muy pequeño.

5. ¿Cuál es la variable que más contribuye al primer eje? ¿Cuál es la que menos? (indique los porcentajes).
6. Según el círculo de correlaciones, ¿cuáles son las variables más correlacionadas? ¿Cuánto es el valor de la correlación? ¿Sí corresponden a lo que se observa en la matriz de correlaciones?

```
> round(cor(Y), 3)
```

	price	malt	aging	taste
price	1.000	0.657	0.483	0.316
malt	0.657	1.000	0.388	0.263
aging	0.483	0.388	1.000	0.297
taste	0.316	0.263	0.297	1.000

```

> s.corcircle (acp$co)
> Tipo <- cor(as.numeric(Whisky$type), acp$li)
> rownames (Tipo) <- "tipo"
> Tipo

              Axis1      Axis2      Axis3
tipo 0.8069072 -0.1429604 0.2768258

> s.arrow(Tipo, add.plot=TRUE, boxes=FALSE)

```

Segun el circulo de correlaciones, las variables mas correlacionadas entre si son la variable "malt" (Malta) y "price" (Precio). El valor de la correlacion es de 0.657. De acuerdo a la matriz de correlaciones, estas si corresponden a lo observado en la grafica ya que entre mas correlacionadas esten las variables, mas juntas estan las flechas, mientras que por otro lado, entre menos se correlaciones mas separadas estan las flechas que representan a las variables.

7. ¿Cuál es la variable mejor representada en el primer plano factorial?
¿Cuál la peor? (Escriba los porcentajes).
8. ¿Qué representa el primer eje? ¿Qué nombre le asignaría? ¿Qué representa el segundo eje?
9. ¿Cuál es el individuo mejor representado en el primer plano factorial?
Ubique sobre el gráfico de individuos al peor representado sobre el primer plano factorial (indique los porcentajes).

10. Supongamos que usted tiene una gráfica de individuos, donde no se muestran los antiguos ejes de las variables. ¿Cómo dibuja los ejes de apreciación y de precio? (Responda concretamente, es decir, con números).
11. ¿Qué características tienen las marcas de whisky según sus ubicaciones en el plano (a la derecha, a la izquierda, arriba, abajo)?
12. ¿Qué significa el círculo del primer plano factorial de variables? ¿Cómo lo dibujaría en una gráfica impresa donde no está? (Suponga que las escalas de los dos ejes son iguales).
13. A partir de la posición en el plano deduzca las características de las tres categorías de whisky (bajo, estándar y pura malta).
14. Supongamos que usted desea comprar una botella de whisky con buena apreciación y que no sea tan cara. Dé dos números de marcas que compraría. ¿Por qué? ¿Cuáles son las características de las dos marcas?
15. Seleccione dos marcas que definitivamente no compraría. ¿Por qué? ¿Qué características tienen?
16. Realice un resumen práctico del análisis suponiendo que lo va a entregar a una compañía que contrató el estudio. Debe dar respuesta al objetivo y apoyarse en las tablas y gráficas que crea necesarias.