

Actividad 2: Aplicación técnica de reducción de variables

Carmen Witsman García

1. Análisis factorial

1.1. Sin rotación de factores:

```
Loadings:
      Factor1 Factor2
Acido_Malico  -0.429  0.427
Alcohol       0.530  0.615
Flavonoides   0.698 -0.222
Magnesio      0.354  0.219
Matiz         0.818
Proantocianidinas 0.590 -0.661
Prolina       0.808  0.354

      Factor1 Factor2
SS loadings  2.081  1.889
Proportion Var 0.297  0.270
Cumulative Var 0.297  0.567

Test of the hypothesis that 2 factors are sufficient.
The chi square statistic is 16.81 on 8 degrees of freedom.
The p-value is 0.0322
```

Figura 1. Análisis factorial de vinos.csv sin rotación de factores

El estadístico chi-cuadrado (16.81) y el valor p (0.0322) sugieren que, a un nivel de significación de 0.05, **podemos rechazar la hipótesis nula de que solo un factor es suficiente**. En otras palabras, hay evidencia de la existencia de dos factores subyacentes en el *dataset*. El análisis identificó dos factores que explican una parte significativa de la varianza en los datos.

La **matiz** presenta **correlación positiva fuerte** con el **factor 2** y la **prolina** con el **factor 1**.

1.2. Utilizando rotación Varimax

```
Loadings:
      Factor1 Factor2
Acido_Malico      -0.604
Alcohol           0.805 -0.105
Flavonoides       0.372  0.631
Magnesio          0.409
Matiz             0.578 -0.580
Proantocianidinas      0.887
Prolina           0.838  0.275

      Factor1 Factor2
SS loadings    1.993  1.977
Proportion Var  0.285  0.282
Cumulative Var  0.285  0.567

Test of the hypothesis that 2 factors are sufficient.
The chi square statistic is 16.81 on 8 degrees of freedom.
The p-value is 0.0322
```

Figura 2. Análisis factorial de vinos.csv con rotación Varimax

Obtenemos el mismo resultado de prueba de hipótesis que haciendo análisis factorial sin rotación.

Varían las cargas de los factores: La prolina y el alcohol están **fuertemente correlacionados positivamente** con el factor 1, y las proantocianidinas están positivamente correlacionadas con el factor 2.

1.3 Sin utilizar puntuaciones factoriales

```
Loadings:
      Factor1 Factor2
Alcohol           0.805 -0.105
Acido_Malico      -0.604
Magnesio          0.409
Flavonoides       0.372  0.631
Matiz             0.578 -0.580
Proantocianidinas      0.887
Prolina           0.838  0.275

      Factor1 Factor2
SS loadings    1.993  1.977
Proportion Var  0.285  0.282
Cumulative Var  0.285  0.567

Test of the hypothesis that 2 factors are sufficient.
The chi square statistic is 16.81 on 8 degrees of freedom.
The p-value is 0.0322
```

Figura 3. Análisis factorial de vinos.csv sin puntuaciones factoriales

El análisis cuantifica la proporción de varianza en los datos que es explicada por los factores, lo que indica la efectividad general del modelo. Los resultados y pruebas de hipótesis **coinciden con los resultados de los puntos 1.1 y 1.2.**

2. Buscar el nº de factores a extraer

```
Parallel analysis suggests that the number of factors = 2 and the number of components = 2
```

Eigen Values of					
	Original factors	Resampled data	Simulated data	Original components	
1	1.96	0.67	0.67		2.46
2	1.40	0.19	0.20		2.24
Resampled components		Simulated components			
1	1.29		1.28		
2	1.16		1.17		

Figura 4. Análisis paralelo de vinos.csv

El número de factores a extraer es 2 basado en el criterio de Kaiser (autovalores mayores que 1) y el análisis de la pantalla (scree plot).

Estos dos factores representan las dimensiones subyacentes más significativas que explican la variabilidad en los datos.

3. Conclusiones

Se eligen 2 componentes porque explican la mayor parte de la varianza acumulada, lo cual es suficiente para una representación simplificada de los datos.

Los dos factores representan agrupaciones de variables que tienen comportamientos similares. El

primer factor está relacionado con el alcohol y la prolina del vino, que son dos componentes importantes del vino que influyen en sus características organolépticas y calidad. Por otro lado, **el segundo está más relacionado con las proantocianidinas y los flavanoides**. Ambas son sustancias fenólicas que se encuentran naturalmente en las uvas, contribuyen a las propiedades antioxidantes del vino e influyen su color.



Figura 5. Biplot de factores que afectan a las variables de vinos.csv