

Introducción a la Análisis Multivariante

Datos de la actividad

Número actividad	3
Docente	Mireia Martorell Colom
Alumno	JOSÉ DAVID VENTURA ROLDÁN KATHERINE CARBONELL MONDRAGÓN MATÍAS DÁVILA WINDER SEBASTIÁN SOTO VALDÉS
Lengua de docencia	Castellano
Agrupación	GRUPO 4
Fecha de entrega	12 Diciembre 2023

SOLUCIÓN

1.1. Carga de la Tabla de Contingencia

El código comienza creando una matriz que representa una tabla de contingencia. Las filas representan el estado de la montaña ("Bueno", "Aceptable", "Deficiente") y las columnas representan la cantidad de agua ("A", "B", "C", "D"). Esta tabla es la base para el análisis posterior.

```
data <- matrix(c(8, 4, 4, 2, 11, 7, 8, 5, 5, 3, 6, 6), nrow = 3, byrow = TRUE)
```

```
rownames(data) <- c("Bueno", "Aceptable", "Deficiente")
```

```
colnames(data) <- c("A", "B", "C", "D")
```

Como resultado obtendremos:

Una matriz de 3x4 representando la tabla de contingencia con estados de la montaña y cantidad de agua.

```
> print(data)
      A B C D
Bueno  8 4 4 2
Aceptable 11 7 8 5
Deficiente 5 3 6 6
```

1.2. Prueba de Chi-Cuadrado para Independencia

Se realiza una prueba de chi-cuadrado para determinar si hay una asociación significativa entre el estado de la montaña y la cantidad de agua. Los resultados muestran una estadística chi-cuadrado de 3.661995 con 6 grados de libertad y un valor p de 0.722307. Este alto valor p sugiere que no hay evidencia suficiente para rechazar la hipótesis nula de independencia entre las dos variables.

```
chi_sq_test <- chisq.test(data)

cat("Estadística chi-cuadrado:", chi_sq_test$statistic, "\n")

cat("Grados de libertad:", chi_sq_test$parameter, "\n")

cat("Valor p:", chi_sq_test$p.value, "\n")
```

Como resultado obtendremos:

Estadística chi-cuadrado: 3.661995

Grados de libertad: 6

Valor p: 0.722307

1.3. Cálculo de la Matriz de Correspondencia

Se utiliza el paquete ca para calcular la matriz de correspondencia, una técnica utilizada para analizar las relaciones entre categorías de variables categóricas. Los resultados incluyen las inercias (valores propios), que muestran la cantidad de variabilidad explicada por cada dimensión. En este caso, la primera dimensión explica el 97.12% de la variabilidad.

```
library(ca)

correspondencia <- ca(data)

print(correspondencia)
```

Como resultado obtendremos:

Muestra las inercias (valores propios) y la distribución de las categorías en las dimensiones del análisis de correspondencia.



```
> print(correspondencia)

Principal inertias (eigenvalues):
      1      2
Value  0.051544 0.001528
Percentage 97.12%  2.88%

Rows:
      Bueno Aceptable Deficiente
Mass    0.260870  0.449275  0.289855
ChiDist 0.257128  0.081611  0.336560
Inertia 0.017247  0.002992  0.032833
Dim. 1  1.111443  0.309370 -1.479823
Dim. 2  1.264131 -1.063060  0.510025

Columns:
      A      B      C      D
Mass    0.347826 0.202899 0.260870 0.188406
ChiDist 0.207941 0.166719 0.110902 0.393576
Inertia 0.015040 0.005640 0.003208 0.029184
Dim. 1  0.898456 0.683316 -0.479176 -1.731091
Dim. 2  1.033333 -1.561867 -0.551111 0.537395
```

Determinación de Categorías Más y Menos Frecuentes

El código identifica "Aceptable" como la categoría más frecuente y "Bueno" como la menos frecuente, basándose en la suma de las filas de la tabla de contingencia.

```
mas_frecuente <- rownames(data)[which.max(rowSums(data))]
menos_frecuente <- rownames(data)[which.min(rowSums(data))]
cat("Categoría más frecuente:", mas_frecuente, "\n")
cat("Categoría menos frecuente:", menos_frecuente, "\n")
```

Como resultado obtendremos:

Categoría más frecuente: Aceptable

Categoría menos frecuente: Bueno

1.4. Conteos Esperados Bajo Independencia

Se calculan los conteos esperados asumiendo independencia entre las variables. La suma de estos conteos es 69, lo que corresponde al total de observaciones.

```
conteos_esperados <- chisq.test(data)$expected
suma_conteos_esperados <- sum(conteos_esperados)
```

```
cat("Suma de todos los conteos esperados:", suma_conteos_esperados, "\n")
```

Como resultado obtendremos:

Suma de todos los conteos esperados: 69

1.5. Perfiles de Fila y Promedio Ponderado

Se calculan los perfiles de fila, que son proporciones de cada categoría dentro de cada fila. Luego, se calcula un promedio ponderado de estos perfiles, utilizando las masas de las filas como pesos.

```
perfiles_fila <- prop.table(data, margin = 1)
pesos_fila <- apply(data, 1, sum) / sum(data)
promedio_ponderado <- colSums(perfiles_fila * pesos_fila)
print(promedio_ponderado)
```

Como resultado obtendremos:

Muestra los perfiles de fila y el promedio ponderado de estos perfiles.

```
> print(perfiles_fila)
```

	A	B	C	D
Bueno	0.4444444	0.2222222	0.2222222	0.1111111
Aceptable	0.3548387	0.2258065	0.2580645	0.1612903
Deficiente	0.2500000	0.1500000	0.3000000	0.3000000

1.6. Inercia Total de la Tabla

La inercia total, que es la suma de las inercias (valores propios), se calcula como 0.053072. Esto proporciona una medida de la variabilidad total en los datos.

```
inercia_total <- sum(correspondencia$eig)
cat("Inercia total:", inercia_total, "\n")
```

Como resultado obtendremos:

Inercia total: 0.053072



1.7. Análisis de Correspondencia con Resumen

Se muestra un resumen del análisis de correspondencia, incluyendo detalles sobre las inercias, las masas, las calidades (qlt), y las contribuciones (ctr) de las filas y las columnas.

summary(correspondencia)

Como resultado obtendremos:

Un resumen detallado del análisis de correspondencia, incluyendo inercias, masas y contribuciones de cada categoría.

```
> summary(correspondencia)
```

Principal inertias (eigenvalues):

dim	value	%	cum%	scree plot
1	0.051544	97.1	97.1	*****
2	0.001528	2.9	100.0	*

Total: 0.053072 100.0

Rows :

	name	mass	qlt	inr	k=1	cor	ctr	k=2	cor	ctr
1	Buen	261	1000	325	252	963	322	49	37	417
2	Acpt	449	1000	56	70	741	43	-42	259	508
3	Dfcn	290	1000	619	-336	996	635	20	4	75

Columns :

	name	mass	qlt	inr	k=1	cor	ctr	k=2	cor	ctr
1	A	348	1000	283	204	962	281	40	38	371
2	B	203	1000	106	155	866	95	-61	134	495
3	C	261	1000	60	-109	962	60	-22	38	79
4	D	188	1000	550	-393	997	565	21	3	54

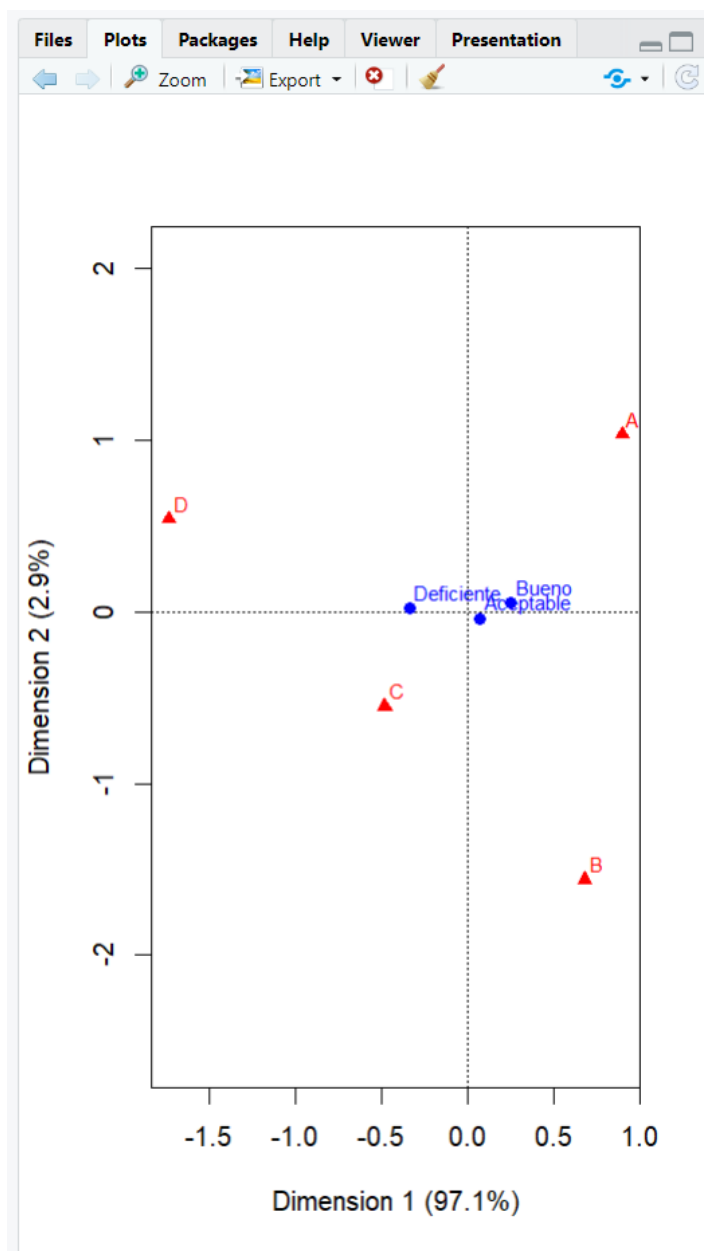
1.8. Gráfico de Perfiles de Fila

Se crea un gráfico que visualiza los perfiles de fila en el espacio de correspondencia. Este gráfico ayuda a interpretar visualmente las relaciones entre las categorías.

plot(correspondencia, map = "rowprincipal")

Como resultado obtendremos:

Un gráfico visual que muestra la distribución de los perfiles de fila en el espacio de correspondencia.



Para los Pasos 1.9, 1.10 y 1.11

Estos pasos involucran la interpretación visual del gráfico y la determinación de los principales contribuyentes a la primera dimensión del análisis. La opción de escala "rowgreen" no se utiliza, ya que puede no estar disponible en el paquete ca.

COMO QUEREIS QUE HAGAMOS ESTA PARTE

1.1 BIBLIOGRAFÍA

- Peña D (2002). Análisis de datos multivariantes. McGraw-Hill Interamericana. Apartados: 1, 2, 3, 4 y 5. <https://elibro.net/es/lc/universitatcarlemany/titulos/50267>
- Redondo Figuero (2017), Carlos G. El programa R, herramienta clave en investigación. Editorial de la Universidad de Cantabria.