

Análisis de Correspondencias Múltiples

Ejemplo 1 - TE

Biblioteca

Para desarrollar el Análisis de Correspondencias Múltiples (ACM) puede utilizarse la biblioteca **FactoMineR** como anteriormente.

```
> library(FactoMineR)
> #library(help=FactoMineR)
> #?MCA
```

Datos - Ejemplo de FactoMineR

Se cargan los datos **tea** que están en la misma librería.

```
> data(tea)
> # str(tea); head(tea)
```

Los datos provienen de un cuestionario sobre el TÉ, aplicado a 300 encuestados, con 18 preguntas acerca de como lo toman, 12 preguntas acerca de su percepción y 4 con preguntas personales.

Se dispone de un *data.frame* con 300 filas una por cada encuestados y 36 columnas por cada una de las diferentes preguntas:

breakfast Factor w/ 2 levels: breakfast, Not.breakfast

tea.time Factor w/ 2 levels: Not.tea time, tea time

evening Factor w/ 2 levels: evening, Not.evening

lunch Factor w/ 2 levels: lunch, Not.lunch

dinner Factor w/ 2 levels: dinner, Not.dinner

always Factor w/ 2 levels: always, Not.always

home Factor w/ 2 levels: home, Not.home
work Factor w/ 2 levels: Not.work, work
tearoom Factor w/ 2 levels: Not.tearoom , tearoom
friends Factor w/ 2 levels: friends, Not.friends
resto Factor w/ 2 levels: Not.resto, resto
pub Factor w/ 2 levels: Not.pub, pub
Tea Factor w/ 3 levels: black, Earl Grey , green
How Factor w/ 4 levels: alone, lemon, milk , other
sugar Factor w/ 2 levels: No.sugar, sugar
how Factor w/ 3 levels: tea bag, tea bag+unpackaged, unpackaged
where Factor w/ 3 levels: chain store, chain store+tea shop , tea shop
price Factor w/ 6 levels: p_branded, p_cheap, p_private label, p_unknown, p_upscale, p_variable

La variable continua **age** es considerada suplementaria y en el ejemplo se plantean las siguientes variables como suplementarias cualitativas:

sex Factor w/ 2 levels F,M
SPC Factor w/ 7 levels employee, middle, non-worker , other worker, senior,student, workman
Sport Factor w/ 2 levels Not.sportsman, sportsman
age_Q Factor w/ 5 levels 15-24,25-34, 35-44, 45-59, mas60
frequency Factor w/ 4 levels 1/day,1 to 2/week, +2/day, 3 to 6/week
escape.exoticism Factor w/ 2 levels escape-exoticism, Not.escape-exoticism
spirituality Factor w/ 2 levels Not.spirituality, spirituality
healthy Factor w/ 2 levels healthy, Not.healthy
diuretic Factor w/ 2 levels diuretic, Not.diuretic
friendliness Factor w/ 2 levels friendliness, Not.friendliness

[iron.absorption](#) Factor w/ 2 levels iron absorption,Not.iron absorption
[feminine](#) Factor w/ 2 levels feminine, Not.feminine
[sophisticated](#) Factor w/ 2 levels Not.sophisticated, sophisticated
[slimming](#) Factor w/ 2 levels No.slimming, slimming
[exciting](#) Factor w/ 2 levels exciting, No.exciting
[relaxing](#) Factor w/ 2 levels No.relaxing, relaxing
[effect.on.health](#) Factor w/ 2 levels effect on health, No.effect on health

breakfast	tea.time	evening	lunch
breakfast :144	Not.tea time:131	evening :103	lunch : 44
Not.breakfast:156	tea time :169	Not.evening:197	Not.lunch:256

dinner	always	home	work
dinner : 21	always :103	home :291	Not.work:213
Not.dinner:279	Not.always:197	Not.home: 9	work : 87

tearoom	friends	resto	pub
Not.tearoom:242	friends :196	Not.resto:221	Not.pub:237
tearoom : 58	Not.friends:104	resto : 79	pub : 63

Tea	How	sugar	how
black : 74	alone:195	No.sugar:155	tea bag :170
Earl Grey:193	lemon: 33	sugar :145	tea bag+unpackaged: 94
green : 33	milk : 63		unpackaged : 36
	other: 9		

	where		price
chain store	:192	p_branded	: 95
chain store+tea shop:	78	p_cheap	: 7
tea shop	: 30	p_private label:	21
		p_unknown	: 12
		p_upscale	: 53
		p_variable	:112

age

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
15.00	23.00	32.00	37.05	48.00	90.00

sex		SPC	Sport	age_Q	frequency
F:178	employee	:59	Not.sportsman:121	15-24:92	1/day : 95
M:122	middle	:40	sportsman :179	25-34:69	1 to 2/week: 44
	non-worker	:64		35-44:40	+2/day :127
	other worker:	20		45-59:61	3 to 6/week: 34
	senior	:35		+60 :38	
	student	:70			
	workman	:12			
	escape.exoticism		spirituality		healthy
	escape-exoticism	:142	Not.spirituality:206	healthy	:210
	Not.escape-exoticism:	158	spirituality : 94	Not.healthy:	90

	diuretic		friendliness		iron.absorption
diuretic	:174	friendliness	:242	iron absorption	: 31
Not.diuretic:	126	Not.friendliness:	58	Not.iron absorption:	269

feminine	sophisticated	slimming	exciting
feminine :129	Not.sophisticated: 85	No.slimming:255	exciting :116
Not.feminine:171	sophisticated :215	slimming : 45	No.exciting:184

relaxing	effect.on.health
No.relaxing:113	effect on health : 66
relaxing :187	No.effect on health:234

ACM

Se utilizan las primeras 18 preguntas como variables acivas, la siguiente (age) como variable suplementarias cualitativa y las últimas como variables suplementarias categóricas.

La función que realiza el ACM en FactoMineR es:

```
MCA(X, ncp = 5, ind.sup = NULL, quanti.sup = NULL, quali.sup = NULL,
graph = TRUE, level.ventil = 0, axes = c(1,2), row.w = NULL,
method="Indicator", na.method="NA", tab.disj=NULL)
```

Los principales argumentos son:

- **X** data frame de $I \times J$ individuos por variables categóricas
- **ncp** numero de dimensiones que se mantienen en los resultados
- **ind.sup ; quanti.sup ; quali.sup** vectores son los indices de individuos y variables suplementarias
- **graph** parámetro lógico.

```
res.mca <- MCA(tea[,1:18],graph=FALSE)
```

Call:

```
.ess_weave(Sweave, "/home/andres/Desktop/multivar2015/ACM/acm_lab/ejemplo1.Rnw")
```

Eigenvalues

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5	Dim.6	Dim.7
Variance	0.148	0.122	0.090	0.078	0.074	0.071	0.068
% of var.	9.885	8.103	6.001	5.204	4.917	4.759	4.522
Cumulative % of var.	9.885	17.988	23.989	29.192	34.109	38.868	43.390
	Dim.8	Dim.9	Dim.10	Dim.11	Dim.12	Dim.13	Dim.14
Variance	0.065	0.062	0.059	0.057	0.054	0.052	0.049
% of var.	4.355	4.123	3.902	3.805	3.628	3.462	3.250
Cumulative % of var.	47.745	51.867	55.769	59.574	63.202	66.664	69.914
	Dim.15	Dim.16	Dim.17	Dim.18	Dim.19	Dim.20	Dim.21
Variance	0.048	0.047	0.046	0.040	0.038	0.037	0.036
% of var.	3.221	3.127	3.037	2.683	2.541	2.438	2.378
Cumulative % of var.	73.135	76.262	79.298	81.982	84.523	86.961	89.339
	Dim.22	Dim.23	Dim.24	Dim.25	Dim.26	Dim.27	

Variance	0.035	0.031	0.029	0.027	0.021	0.017
% of var.	2.323	2.055	1.915	1.821	1.407	1.139
Cumulative % of var.	91.662	93.717	95.633	97.454	98.861	100.000

Categories (the 10 first)

	Dim.1	ctr	cos2	v.test	Dim.2	ctr	cos2	v.test
breakfast	0.166	0.495	0.025	2.756	-0.166	0.607	0.026	-2.764
Not.breakfast	-0.153	0.457	0.025	-2.756	0.154	0.560	0.026	2.764
Not.tea time	-0.498	4.053	0.192	-7.578	0.093	0.174	0.007	1.423
tea time	0.386	3.142	0.192	7.578	-0.072	0.135	0.007	-1.423
evening	0.319	1.307	0.053	3.985	-0.058	0.053	0.002	-0.728
Not.evening	-0.167	0.683	0.053	-3.985	0.030	0.028	0.002	0.728
lunch	0.659	2.385	0.075	4.722	-0.390	1.018	0.026	-2.793
Not.lunch	-0.113	0.410	0.075	-4.722	0.067	0.175	0.026	2.793
dinner	-0.661	1.146	0.033	-3.136	0.796	2.025	0.048	3.774
Not.dinner	0.050	0.086	0.033	3.136	-0.060	0.152	0.048	-3.774

	Dim.3	ctr	cos2	v.test
breakfast	-0.483	6.900	0.215	-8.017
Not.breakfast	0.445	6.369	0.215	8.017
Not.tea time	0.265	1.886	0.054	4.027
tea time	-0.205	1.462	0.054	-4.027
evening	0.451	4.312	0.106	5.640
Not.evening	-0.236	2.254	0.106	-5.640
lunch	0.301	0.822	0.016	2.160
Not.lunch	-0.052	0.141	0.016	-2.160
dinner	0.535	1.235	0.022	2.537
Not.dinner	-0.040	0.093	0.022	-2.537

Categorical variables (eta2)

	Dim.1	Dim.2	Dim.3
breakfast	0.025	0.026	0.215
tea.time	0.192	0.007	0.054
evening	0.053	0.002	0.106
lunch	0.075	0.026	0.016
dinner	0.033	0.048	0.022
always	0.045	0.001	0.101
home	0.005	0.000	0.134
work	0.112	0.043	0.005
tearoom	0.372	0.022	0.008
friends	0.243	0.015	0.103

La función devuelve una lista:

- `eig` - eigenvalues, %de variancia y acumulada
- `var` - resultados para las variables activas
- `var$coord` - coordenadas de las categorías
- `var$cos2` - cos2 para las categorías
- `var$contrib` - contribuciones de las categorías
- `ind` - resultados para los individuos
- `...sup` resultados para las variables suplementarias

```
plot(x, axes = c(1, 2), choix=c("ind","var","quanti.sup"), xlim = NULL, ylim = NULL,  
invisible = c("none","ind","var","ind.sup","quali.sup","quanti.sup"),col.ind = "black",
```

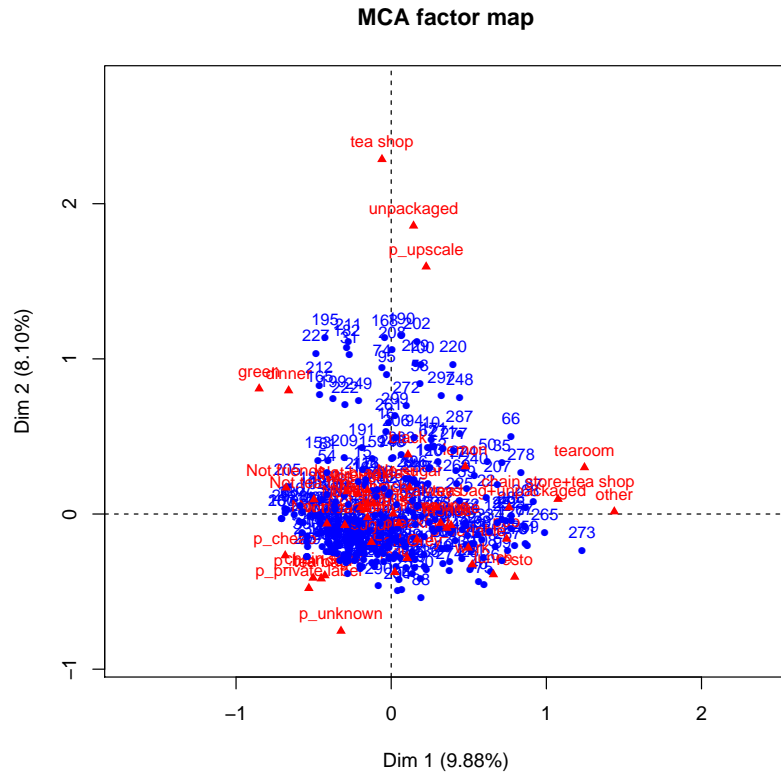
Los argumentos mas importantes son:

- `x` objeto de clase MCA
- `axes` vector de los 2 componentes a graficar
- `choix` “ind”, “var”, “quanti.sup” ...
- `xlim ; ylim ; invisible ; col.ind; col.var ;col.quali.sup ...`

Los gráficos disponibles de forma inmediata son:

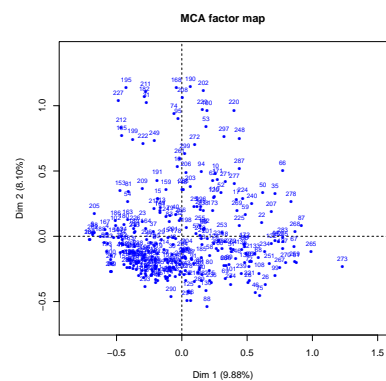
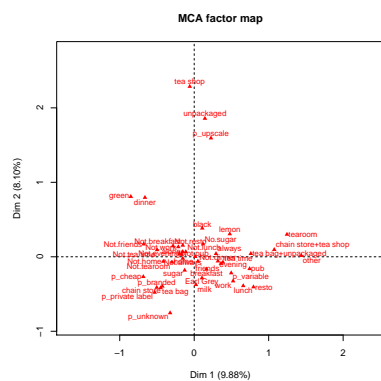
- El del plano principal con la representación simultánea de individuos y modalidades de las variables.

```
> pdf('ej1_varind.pdf');plot(res.mca,cex=0.8);graphics.off()
```

- Alternativamente pueden graficarse por separado, usando el argumento `invisible=` para no desplegar algunos de los elementos

```
> pdf('ej1_ind.pdf');plot(res.mca,invisible=c("var"),cex=0.7)
> pdf('ej1_var.pdf');plot(res.mca,invisible=c("ind"),cex=0.8)
> graphics.off()
```



- Si la ejecución de la función **MCA** fué realizada con elementos suplementarios, estos también pueden graficarse.
- Por defecto se grafica el plano principal, pero con el argumento **axes** pueden graficarse otros planos: `plot(res.mca,axes=c(1,3))`
- Se pueden controlar muchos parámetros gráficos, por ejemplo, límites, colores, etiquetas, títulos, etc: `?plot.MCA`

Adicionalmente, como ayudas a la interpretación pueden utilizarse:

- Al igual que con otras técnicas/objetos del package: `dimdesc(res.mca)`
- `plotellipses(res.mca,keepvar=1:4)`