Tema 4: Análisis Factorial

En numerosas áreas de Psicología y de Ciencias del Comportamiento no es posible medir directamente las variables que interesan; por ejemplo, los conceptos de *inteligencia* y de *clase social*. En estos casos es necesario recoger medidas indirectas que estén relacionadas con los conceptos que interesan. Las variables que interesan reciben el nombre de *variables latentes* y la metodología que las relaciona con variables observadas recibe el nombre de Análisis Factorial.

El modelo de Análisis Factorial es un modelo de regresión múltiple que relaciona variables latentes con variables observadas.

El Análisis Factorial tiene muchos puntos en común con el análisis de componentes principales, y busca esencialmente nuevas variables o *factores* que expliquen los datos. En el análisis de componentes principales, en realidad, sólo se hacen transformaciones ortogonales de las variables originales, haciendo hincapié en la varianza de las nuevas variables. En el análisis factorial, por el contrario, interesa más explicar la estructura de las covarianzas entre las variables.

Al igual que en el método de los componentes principales, para efectuar el análisis factorial, es necesario que las variables originales no estén incorreladas porque si lo estuvieran no habría nada que explicar de las variables.

Consideramos un conjunto de p variables observadas $\mathbf{x}' = (x_1, x_2, \ldots, x_p)$ que se asume relacionadas con un número dado de variables latentes $f_1, f_2, \ldots, f_k$, donde $k < p$, mediante una relación del tipo

$$\begin{aligned} x_1 &= \lambda_{11} f_1 + \cdots + \lambda_{1k} f_k + u_1 \\ &\vdots \\ x_p &= \lambda_{p1} f_1 + \cdots + \lambda_{pk} f_k + u_p \end{aligned}$$

o de modo más conciso

$$\mathbf{x} = \Lambda \mathbf{f} + \mathbf{u}.$$

donde

$$\Lambda = \begin{pmatrix} \lambda_{11} & \cdots & \lambda_{1k} \\ \vdots & \ddots & \vdots \\ \lambda_{p1} & \cdots & \lambda_{pk} \end{pmatrix}, \quad \mathbf{f} = \begin{pmatrix} f_1 \\ \vdots \\ f_k \end{pmatrix}, \quad \mathbf{u} = \begin{pmatrix} u_1 \\ \vdots \\ u_p \end{pmatrix}.$$

Los λ_{ij} son los *pesos factoriales* que muestran como cada x_i depende de factores comunes y se usan para interpretar los factores. Por ejemplo, valores altos relacionan un factor con la correspondiente variable observada y así se puede caracterizar cada factor.

Se asume que los términos residuales $u_1, \ldots, u_p$ están incorrelados entre sí y con los factores $f_1, \ldots, f_k$. Cada variable u_i es particular para cada x_i y se denomina *variable específica*.

Dado que los factores no son observables, se puede fijar arbitrariamente su media en 0 y su varianza en 1, esto es, se consideran variables estandarizadas que están incorreladas entre sí, de modo que los pesos factoriales resultan ser las correlaciones entre las variables y los factores.

Así, con las suposiciones previas, la varianza de la variable x_i es

$$\sigma_i^2 = \sum_{j=1}^{k} \lambda_{ij}^2 + \psi_i$$

donde ψ_i es la varianza de u_i.

De este modo, la varianza de cada variable observada se puede descomponer en dos partes. La primera h_i^2, denominada *comunalidad*, es

$$h_i^2 = \sum_{j=1}^{k} \lambda_{ij}^2$$

y representa la varianza compartida con las otras variables por medio de los factores comunes. La segunda parte, ψ_i, se denomina varianza específica y recoge la variabilidad no compartida con las otras variables.

La definición del modelo implica que la covarianza entre las variables x_i y x_j es

$$\sigma_{ij} = \sum_{l=1}^{k} \lambda_{il}\lambda_{lj}.$$

Las covarianzas no dependen en absoluto de las variables específicas, de hecho, basta con los factores comunes. De este modo, la matriz de covarianzas Σ de las variables observadas es

$$\Sigma = \Lambda\Lambda' + \Psi$$

donde Ψ es una matriz diagonal cuyos componentes son las varianzas específicas: $\Psi = diag(\psi_i)$.

Lo contrario también se verifica: dada la descomposición de la varianza anterior, se puede encontrar un modelo factorial para las variables originales, $\mathbf{x}$, con k factores.

En la práctica se tienen que estimar los parámetros del modelo a partir de una muestra, de modo que el problema se centra en encontrar los valores $\hat{\Lambda}$ y $\hat{\Psi}$ tales que la matriz de covarianzas muestral S es aproximadamente

$$S \approx \hat{\Lambda}\hat{\Lambda}' + \hat{\Psi}$$

Se tienen dos métodos de estimación de los términos anteriores: el método de los factores principales y el método de máxima verosimilitud.

Método de los factores principales

Es una técnica basada en autovalores y autovectores pero en lugar de operar sobre la matriz de covarianzas se opera sobre la llamada matriz de covarianzas *reducida*,

$$S^* = S - \hat{\Psi}$$

donde $\hat{\Psi}$ es una matriz diagonal que contiene las estimas de ψ_i.

Los elementos diagonales de S^* contiene las comunalidades estimadas (las partes de las varianzas de cada variable explicada por los factores comunes). Al contrario que el análisis de componentes principales, el análisis factorial no pretende recoger toda la varianza observada de los datos, sino la que comparten los factores comunes. De hecho, el análisis factorial se centra más en recoger las covarianzas o correlaciones que aparecen entre las variables originales.

El procedimiento es iterativo: se parte de unas comunalidades estimadas a partir de las correlaciones entre las variables observadas y luego se efectua un análisis de componentes principales sobre la matriz S^*.

Método de la máxima verosimilitud

Este método es el habitualmente preferido por los estadísticos. Asumiendo normalidad en los datos se define una distancia F, entre la matriz de covarianzas observada y los valores predichos de esta matriz por el modelo del análisis factorial. La expresión de dicha distancia es

$$F = \ln |\Lambda\Lambda' + \Psi| + traza\left(S\,|\Lambda\Lambda' + \Psi|^{-1}\right) - \ln |S| - p$$

Las estimaciones de los pesos factoriales se obtienen minimizando esta función, y esto es equivalente a maximizar la función de verosimilitud del modelo k factorial asumiendo normalidad.

Estimación del número de factores

El hecho de tomar un número adecuado de factores k para representar las covarianzas observadas es muy importante: entre una solución con k ó con $k+1$ factores se pueden encontrar pesos factoriales muy diferentes, al contrario que en el método de componentes principales, donde los primeros k componentes son siempre iguales.

Una ventaja del método de máxima verosimilitud es que lleva asociado un test estadístico para estimar el número de factores.

Rotación de los factores

En el Análisis Factorial no existe una solución única para determinar la matriz de pesos, de hecho, se puede multiplicar por una matriz ortogonal M de orden $k \times k$ de modo que

$$\begin{aligned} \mathbf{x} &= \Lambda \mathbf{f} + \mathbf{u} = \\ &= (\Lambda M)(M'\mathbf{f}) + \mathbf{u}, \end{aligned}$$

y este nuevo *modelo* verifica las mismas propiedades que el anterior: tiene como factores $\mathbf{f}^* = M'\mathbf{f}$ y como matriz de pesos ΛM. En este caso, la matriz de covarianzas de las variables originales es

$$\Sigma = (\Lambda M)(\Lambda M)' + \Psi,$$

que como $MM' = I$, se reduce a que $\Sigma = \Lambda\Lambda' + \Psi$ como antes; de este modo se explica de manera equivalente.la matriz de covarianzas de las variables originales.

Puede ser que la solución sea más interpretable mediante el uso de alguna matriz ortogonal, lo que lleva al concepto de rotación de los factores. Según Thurstone, la intención fundamental al realizar una rotación es encontrar una *estructura simple.* Las propiedades que debe cumplir son

- Cada fila de la matriz factorial de pesos debe contener, al menos, un cero.
- Cada columna de la matriz factorial de pesos debe contener, al menos, k ceros.
- Cada par de columnas de la matriz factorial de pesos debe contener varias variables cuyos pesos sean nulos en una columna pero no en la otra.
- Si hay más de cuatro factores cada par de columnas de la matriz factorial de pesos debe contener un número elevado de variables con pesos nulos en ambas columnas.
- De manera recíproca, si hay más de cuatro factores, en cada par de columnas de la matriz factorial de pesos sólo un número pequeño de variables debe contener pesos no nulos.

Cuando se consigue una estructura simple, las variables observadas se encuentran en grupos mutuamente excluyentes de modo que los pesos son altos en unos pocos factores y bajos en el resto.

Tipos de rotaciones

Hay dos posible tipos de rotaciones: *ortogonales* y *oblicuas.*

La ventaja principal de las rotaciones ortogonales es su simplicidad, ya que los pesos representan las correlaciones entre los factores y las variables, sin embargo esto no se cumple en el caso de las rotaciones oblicuas. Entre las rotaciones ortogonales se encuentran dos tipos principales:

Rotación Varimax: Fue propuesta por Kaiser (1958), y trata de que los factores tengan unas pocas saturaciones altas y muchas casi nulas en las variables. Esto hace que haya factores con correlaciones altas con un número pequeño de variables y correlaciones nulas en el resto, quedando así redistribuida la varianza de los factores.

Rotación Cuartimax: Trata que una variable dada esté muy correlacionada con un factor y muy poco correlacionada con el resto de factores. Se usa menos frecuentemente que la anterior.

Entre las rotaciones oblicuas, la más empleada es:

Rotación Oblimín: Trata de encontrar una estructura simple sin que importe el hecho de que las rotaciones sean ortogonales, esto es, las saturaciones no representan ya la correlaciones entre los factores y las variables. Se considera un parámetro que controla el grado de correlación entre los factores, con valores preferentemente entre $-0{,}5$ y $0{,}5$.

En cualquier caso, el hecho de rotar los factores siempre es controvertido ya que se pueden elegir los ejes que resulten de mayor conveniencia. Sin embargo, se puede considerar que una rotación es sólo un medio para conseguir unos ejes que permitan describir los puntos de la muestra de la manera más simple posible.

Ejemplos

Ejemplo 1

Se considera una muestra de los años de vida esperados por país, edad y sexo procedentes de Keyfitz y Flieger (1971).

	m0	m25	m50	m75	w0	w25	w50	w75
Algeria	63.00	51.00	30.00	13.00	67.00	54.00	34.00	15.00
Cameroon	34.00	29.00	13.00	5.00	38.00	32.00	17.00	6.00
Madagascar	38.00	30.00	17.00	7.00	38.00	34.00	20.00	7.00
Mauritius	59.00	42.00	20.00	6.00	64.00	46.00	25.00	8.00
Reunion	56.00	38.00	18.00	7.00	62.00	46.00	25.00	10.00
Seychelles	62.00	44.00	24.00	7.00	69.00	50.00	28.00	14.00
South Africa(C)	50.00	39.00	20.00	7.00	55.00	43.00	23.00	8.00
South Africa(W)	65.00	44.00	22.00	7.00	72.00	50.00	27.00	9.00
Tunisia	56.00	46.00	24.00	11.00	63.00	54.00	33.00	19.00
Canada	69.00	47.00	24.00	8.00	75.00	53.00	29.00	10.00
Costa Rica	65.00	48.00	26.00	9.00	68.00	50.00	27.00	10.00
Dominican Rep	64.00	50.00	28.00	11.00	66.00	51.00	29.00	11.00
El Salvador	56.00	44.00	25.00	10.00	61.00	48.00	27.00	12.00
Greenland	60.00	44.00	22.00	6.00	65.00	45.00	25.00	9.00
Grenada	61.00	45.00	22.00	8.00	65.00	49.00	27.00	10.00
Guatemala	49.00	40.00	22.00	9.00	51.00	41.00	23.00	8.00
Honduras	59.00	42.00	22.00	6.00	61.00	43.00	22.00	7.00
Jamaica	63.00	44.00	23.00	8.00	67.00	48.00	26.00	9.00
Mexico	59.00	44.00	24.00	8.00	63.00	46.00	25.00	8.00
Nicaragua	65.00	48.00	28.00	14.00	68.00	51.00	29.00	13.00
Panama	65.00	48.00	26.00	9.00	67.00	49.00	27.00	10.00
Trinidad(62)	64.00	63.00	21.00	7.00	68.00	47.00	25.00	9.00
Trinidad (67)	64.00	43.00	21.00	6.00	68.00	47.00	24.00	8.00
United States (66)	67.00	45.00	23.00	8.00	74.00	51.00	28.00	10.00
United States (NW66)	61.00	40.00	21.00	10.00	67.00	46.00	25.00	11.00
United States (W66)	68.00	46.00	23.00	8.00	75.00	52.00	29.00	10.00
United States (67)	67.00	45.00	23.00	8.00	74.00	51.00	28.00	10.00
Argentina	65.00	46.00	24.00	9.00	71.00	51.00	28.00	10.00
Chile	59.00	43.00	23.00	10.00	66.00	49.00	27.00	12.00
Columbia	58.00	44.00	24.00	9.00	62.00	47.00	25.00	10.00
Ecuador	57.00	46.00	28.00	9.00	60.00	49.00	28.00	11.00

Se usa un análisis factorial por máxima verosimilitud. Primero se prueban tres soluciones con 1, 2 o 3 factores, observándose que la solución con tres factores es la adecuada, al observar el test con la hipótesis nula de que con tres factores es suficiente.

Se obtiene la solución rotada (varimax por defecto) y se observa:

(i) primer factor: está muy relacionado con la esperanza de vida en el nacimiento para mujeres y hombres;

(*ii*) segundo factor: refleja la esperanza de vida para edades más avanzadas;

(*iii*) tercer factor: tiene los pesos factoriales más altos en las esperanzas de vida de hombres entre 50 y 75 años.

En el primer eje se observa que en un extremo se sitúan Camerún y Madagascar frente al otro extremo donde está USA.

En el tercer eje se sitúa en el valor más alto Argelia (que tiene alta esperanza de vida para hombres de edad avanzada) frente a Camerún.

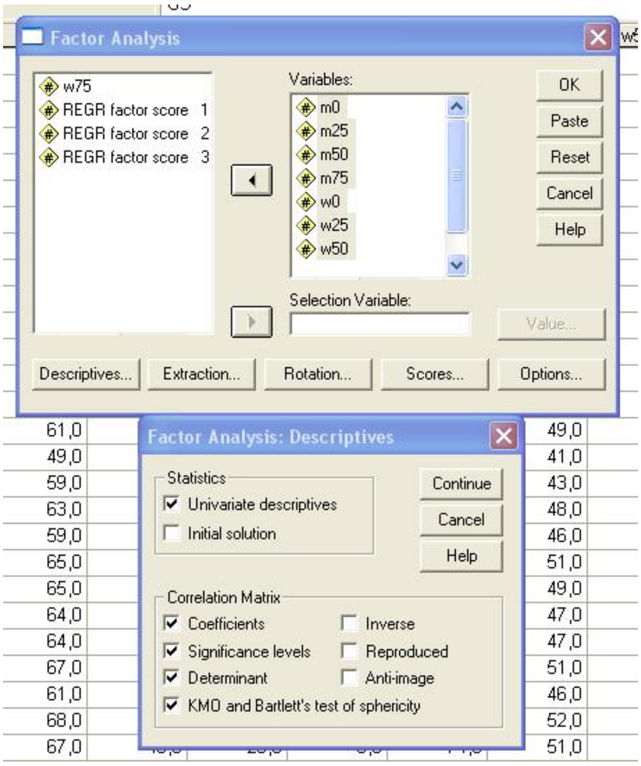
Ejemplo 2

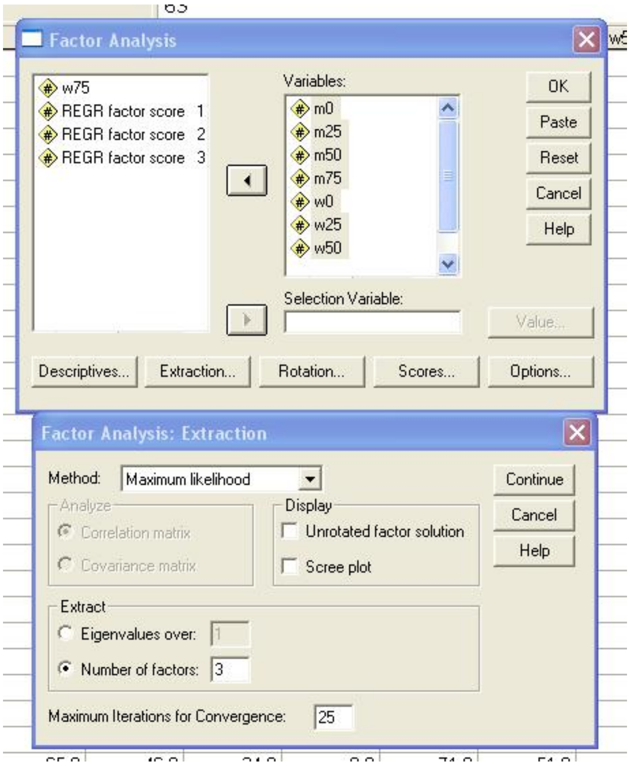
En el siguiente ejemplo, se estudia una muestra de consumo de drogas entre 1634 estudiantes de Los Angeles. Se consideraron 13 tipos de sustancias y, así, 13 variables con 5 niveles de respuesta (desde consumo nulo hasta consumo habitual). Se obtiene la matriz de correlaciones.

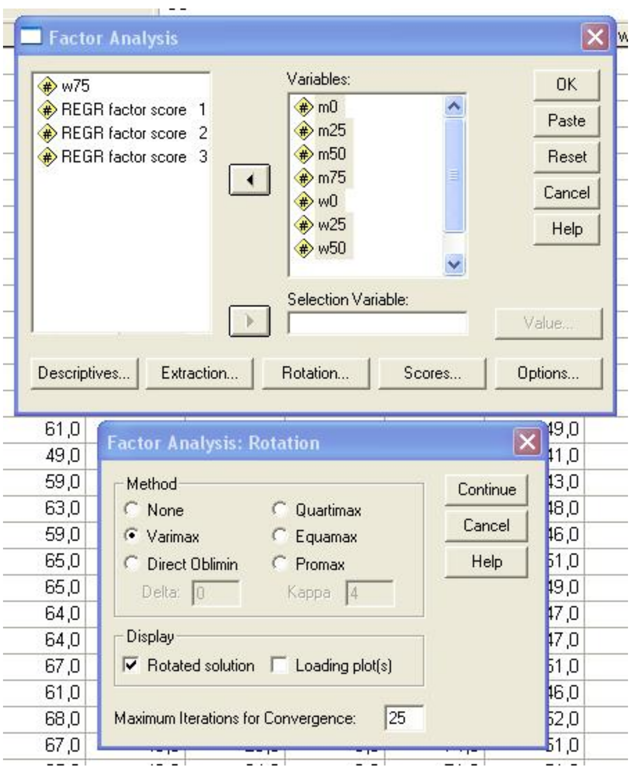
Se obtiene que el número más razonable de factores es de 6. El primero recoge drogas *socialmente aceptadas* y *blandas*, el segundo factor se refiere a drogas *duras*, el tercer factor es simplemente anfetaminas y el cuarto, hachís. Los dos últimos factores resultan difíciles de interpretar.

Aunque el número de factores matemáticamente más coherente es 6, se puede considerar una solución con 3 ó 4 factores sólo dado que los residuos, obtenidos al restar la matriz de correlaciones original y la reproducida, son pequeños.

ANALISIS FACTORIAL (con SPSS)







Estadísticos descriptivos			
	Media	**Desviación típica**	**N del análisis**
m0	59,613	7,9191	31
m25	44,129	5,9033	31
m50	22,935	3,4052	31
m75	8,387	2,0278	31
w0	64,194	8,8220	31
w25	47,516	4,9858	31
w50	26,290	3,3386	31
w75	10,129	2,5787	31

Matriz de correlaciones(a)									
		m0	**m25**	**m50**	**m75**	**w0**	**w25**	**w50**	**w75**
Correlación	**m0**	1,000	,748	,636	,290	,980	,874	,697	,318
	m25	,748	1,000	,667	,391	,693	,725	,647	,393
	m50	,636	,667	1,000	,752	,557	,772	,802	,593
	m75	,290	,391	,752	1,000	,247	,547	,687	,710
	w0	,980	,693	,557	,247	1,000	,887	,710	,365
	w25	,874	,725	,772	,547	,887	1,000	,940	,684
	w50	,697	,647	,802	,687	,710	,940	1,000	,828
	w75	,318	,393	,593	,710	,365	,684	,828	1,000
Sig. (Unilateral)	**m0**		,000	,000	,057	,000	,000	,000	,041
	m25	,000		,000	,015	,000	,000	,000	,014
	m50	,000	,000		,000	,001	,000	,000	,000
	m75	,057	,015	,000		,090	,001	,000	,000
	w0	,000	,000	,001	,090		,000	,000	,022
	w25	,000	,000	,000	,001	,000		,000	,000
	w50	,000	,000	,000	,000	,000	,000		,000
	w75	,041	,014	,000	,000	,022	,000	,000	
a Determinante = 7,91E-007									

KMO y prueba de Bartlett		
Medida de adecuación muestral de Kaiser-Meyer-Olkin.		,794
Prueba de esfericidad de Bartlett	**Chi-cuadrado aproximado**	372,323
	gl	28
	Sig.	,000

Comunalidades(a)
Método de extracción: Máxima verosimilitud.

Matriz factorial(a)
a 3 factores extraídos. Requeridas 7 iteraciones.

Prueba de la bondad de ajuste		
Chi-cuadrado	**gl**	**Sig.**
6,275	7	,508

Correlaciones reproducidas									
		m0	**m25**	**m50**	**m75**	**w0**	**w25**	**w50**	**w75**
Correlación reproducida	**m0**	,999(b)	,748	,636	,290	,980	,874	,696	,318
	m25	,748	,649(b)	,684	,431	,697	,723	,647	,369
	M50	,636	,684	,905(b)	,725	,557	,769	,807	,600
	M75	,290	,431	,725	,707(b)	,244	,556	,690	,653
	W0	,980	,697	,557	,244	,996(b)	,887	,711	,363
	W25	,874	,723	,769	,556	,887	,989(b)	,939	,688
	W50	,696	,647	,807	,690	,711	,939	,980(b)	,827
	W75	,318	,369	,600	,653	,363	,688	,827	,852(b)
Residual(a)	**M0**		,001	,000	-7,91E-5	1,04E-5	-7,65E-5	,000	,000
	m25	,001		-,017	-,040	-,004	,002	,000	,024
	m50	,000	-,017		,027	,000	,003	-,004	-,006
	m75	-7,91E-5	-,040	,027		,003	-,009	-,003	,058
	w0	1,04E-5	-,004	,000	,003		,000	-,001	,002
	w25	-7,65E-5	,002	,003	-,009	,000		,001	-,004
	w50	,000	,000	-,004	-,003	-,001	,001		,001
	w75	,000	,024	-,006	,058	,002	-,004	,001	
Método de extracción: Máxima verosimilitud.									
a Los residuos se calculan entre las correlaciones observadas y reproducidas.									
Hay 1 (3,0%) residuales no redundantes con valores absolutos mayores que 0,05.									
b Comunalidades reproducidas									

Matriz de factores rotados(a)			
	Factor		
	1	2	3
m0	,964	,120	,233
m25	,645	,168	,453
m50	,428	,376	,762
m75	,078	,537	,642
w0	,970	,220	,078
w25	,763	,561	,303
w50	,535	,732	,397
w75	,156	,869	,271
Método de extracción: Máxima verosimilitud. Método de rotación: Normalización Varimax con Kaiser.			
a La rotación ha convergido en 6 iteraciones.			

Varianza total explicada			
Factor	Suma de las saturaciones al cuadrado de la rotación		
	Total	% de la varianza	% acumulado
1	3,369	42,107	42,107
2	2,127	26,589	68,696
3	1,580	19,751	88,447
Método de extracción: Máxima verosimilitud.			

Matriz de transformación de los factores			
Factor	**1**	**2**	**3**
1	,956	,187	,225
2	-,258	,902	,347
3	-,138	-,390	,910
Método de extracción: Máxima verosimilitud. Método de rotación: Normalización Varimax con Kaiser.			

Análisis Factorial (con R)

```
# Leo los datos
vida <- read.table("c:\\...\\datFA.txt",header=T)
# dimnames(vida)[[1]] <- c("Algeria", "Cameroon", "Madagascar", "Mauritius",
"Reunion", "Seychelles", "South Africa(C)", "South Africa(W)", "Tunisia",
"Canada", "Costa Rica", "Dominican Rep", "El Salvador", "Greenland",
"Grenada", "Guatemala", "Honduras", "Jamaica", "Mexico", "Nicaragua",
"Panama", "Trinidad(62)", "Trinidad (67)", "United States (66)", "United
States (NW66)", "United States (W66)", "United States (67)", "Argentina",
"Chile", "Columbia", "Ecuador")
```

```
vida
                     m0 m25 m50 m75 w0 w25 w50 w75
Algeria              63  51  30  13 67  54  34  15
Cameroon             34  29  13   5 38  32  17   6
Madagascar           38  30  17   7 38  34  20   7
Mauritius            59  42  20   6 64  46  25   8
Reunion              56  38  18   7 62  46  25  10
Seychelles           62  44  24   7 69  50  28  14
South Africa(C)      50  39  20   7 55  43  23   8
South Africa(W)      65  44  22   7 72  50  27   9
Tunisia              56  46  24  11 63  54  33  19
Canada               69  47  24   8 75  53  29  10
Costa Rica           65  48  26   9 68  50  27  10
Dominican Rep        64  50  28  11 66  51  29  11
El Salvador          56  44  25  10 61  48  27  12
Greenland            60  44  22   6 65  45  25   9
Grenada              61  45  22   8 65  49  27  10
Guatemala            49  40  22   9 51  41  23   8
Honduras             59  42  22   6 61  43  22   7
Jamaica              63  44  23   8 67  48  26   9
Mexico               59  44  24   8 63  46  25   8
Nicaragua            65  48  28  14 68  51  29  13
Panama               65  48  26   9 67  49  27  10
Trinidad(62)         64  63  21   7 68  47  25   9
Trinidad (67)        64  43  21   6 68  47  24   8
United States (66)   67  45  23   8 74  51  28  10
United States (NW66) 61  40  21  10 67  46  25  11
United States (W66)  68  46  23   8 75  52  29  10
United States (67)   67  45  23   8 74  51  28  10
Argentina            65  46  24   9 71  51  28  10
Chile                59  43  23  10 66  49  27  12
Columbia             58  44  24   9 62  47  25  10
Ecuador              57  46  28   9 60  49  28  11
```

```
# Se prueba una solución factorial con 1 factor
vida.fa1 <- factanal(vida, factors=1, method="mle")
vida.fa1

Call:
factanal(x = vida, factors = 1, method = "mle")

Uniquenesses:
   m0   m25   m50   m75    w0   w25   w50   w75
0.238 0.470 0.399 0.696 0.217 0.005 0.117 0.532

Loadings:
    Factor1
m0  0.873
m25 0.728
m50 0.776
m75 0.552
w0  0.885
w25 0.998
w50 0.940
w75 0.684
```

```
               Factor1
SS loadings      5.329
Proportion Var   0.666

Test of the hypothesis that 1 factor is sufficient.
The chi square statistic is 163.11 on 20 degrees of freedom.
The p-value is 1.88e-24
```

 # Se prueba una solución factorial con 2 factores

vida.fa2 <- factanal(vida, factors=2, method="mle")
vida.fa2

```
Call:
factanal(x = vida, factors = 2, method = "mle")

Uniquenesses:
   m0   m25   m50   m75    w0   w25   w50   w75
0.024 0.442 0.346 0.408 0.015 0.011 0.015 0.178

Loadings:
    Factor1 Factor2
m0  0.972   0.179
m25 0.670   0.329
m50 0.480   0.651
m75 0.122   0.760
w0  0.973   0.194
w25 0.790   0.603
w50 0.567   0.815
w75 0.185   0.888

               Factor1 Factor2
SS loadings      3.567   2.994
Proportion Var   0.446   0.374
Cumulative Var   0.446   0.820

Test of the hypothesis that 2 factors are sufficient.
The chi square statistic is 45.24 on 13 degrees of freedom.
The p-value is 1.91e-05
```

 # Se prueba una solución factorial con 3 factores

vida.fa3 <- factanal(vida, factors=3, method="mle")
vida.fa3

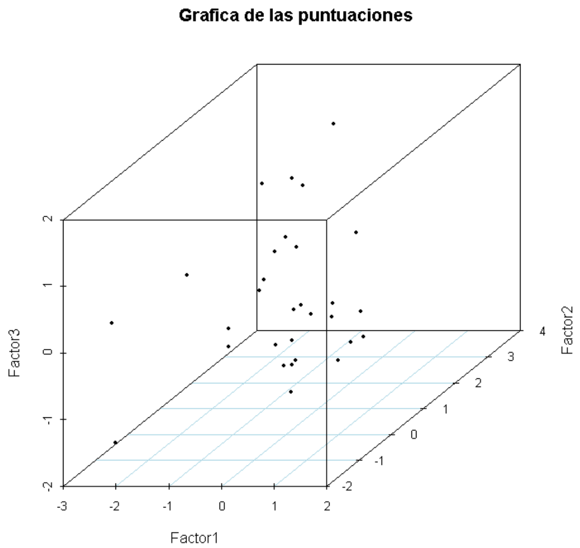
```
Call:
factanal(x = vida, factors = 3, method = "mle")

Uniquenesses:
   m0   m25   m50   m75    w0   w25   w50   w75
0.005 0.362 0.066 0.288 0.005 0.011 0.020 0.146

Loadings:
    Factor1 Factor2 Factor3
m0  0.964   0.122   0.226
m25 0.646   0.169   0.438
m50 0.430   0.354   0.790
m75         0.525   0.656
w0  0.970   0.217
w25 0.764   0.556   0.310
w50 0.536   0.729   0.401
w75 0.156   0.867   0.280

               Factor1 Factor2 Factor3
SS loadings      3.375   2.082   1.640
Proportion Var   0.422   0.260   0.205
Cumulative Var   0.422   0.682   0.887

Test of the hypothesis that 3 factors are sufficient.
The chi square statistic is 6.73 on 7 degrees of freedom.
The p-value is 0.458
```

```
# La solución con tres factores resulta la más apropiada
# Obtengo los pesos factoriales
```

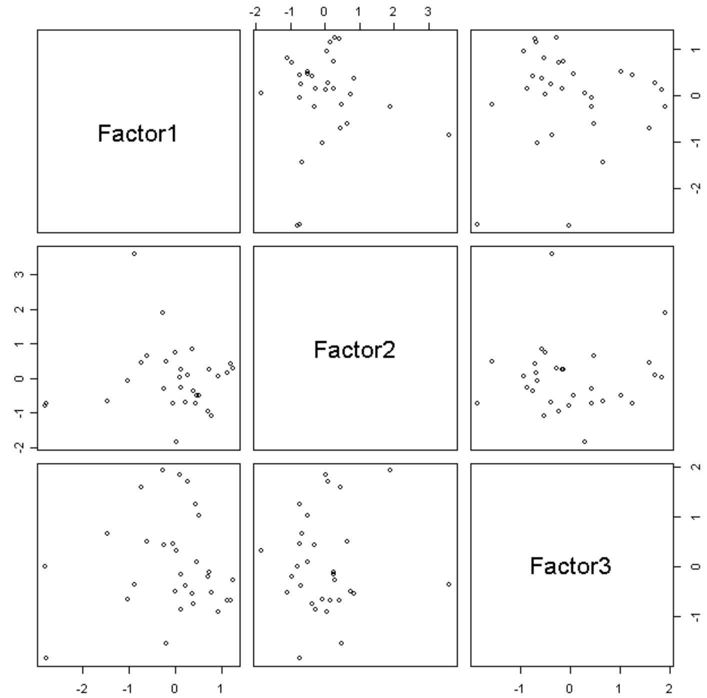
```
scores <- factanal(vida, factors=3, method="mle", scores="regression")$scores
scores
```

```
                        Factor1     Factor2     Factor3
Algeria             -0.258062561  1.90095771  1.91581631
Cameroon            -2.782495791 -0.72340014 -1.84772224
Madagascar          -2.806428187 -0.81158820 -0.01210318
Mauritius            0.141004934 -0.29028454 -0.85862443
Reunion             -0.196352142  0.47429917 -1.55046466
Seychelles           0.367371307  0.82902375 -0.55214085
South Africa(C)     -1.028567629 -0.08065792 -0.65421971
South Africa(W)      0.946193522  0.06400408 -0.91995289
Tunisia             -0.862493550  3.59177195 -0.36442148
Canada               1.245304248  0.29564122 -0.27342781
Costa Rica           0.508736247 -0.50500435  1.01328707
Dominican Rep        0.106044085  0.01111171  1.83871599
El Salvador         -0.608155779  0.65100820  0.48836431
Greenland            0.235114220 -0.69123901 -0.38558654
Grenada              0.132008172  0.25241049 -0.15220645
Guatemala           -1.450336359 -0.67765804  0.65911906
Honduras             0.043253249 -1.85175707  0.30633182
Jamaica              0.462124701 -0.51918493  0.08032855
Mexico              -0.052332675 -0.72020002  0.44417800
Nicaragua            0.268974443  0.08407227  1.70568388
Panama               0.442333434 -0.73778272  1.25218728
Trinidad(62)         0.711367053 -0.95989475 -0.21545329
Trinidad (67)        0.787286051 -1.10729029 -0.51958264
United States (66)   1.128331259  0.16389896 -0.68177046
United States (NW66) 0.400058903 -0.36230253 -0.74299137
United States (W66)  1.214345385  0.40877239 -0.69225320
United States (67)   1.128331259  0.16389896 -0.68177046
Argentina            0.731344988  0.24811968 -0.12817725
Chile                0.009751528  0.75222637 -0.49198911
Columbia            -0.240602517 -0.29543613  0.42919600
Ecuador             -0.723451797  0.44246371  1.59164974
```

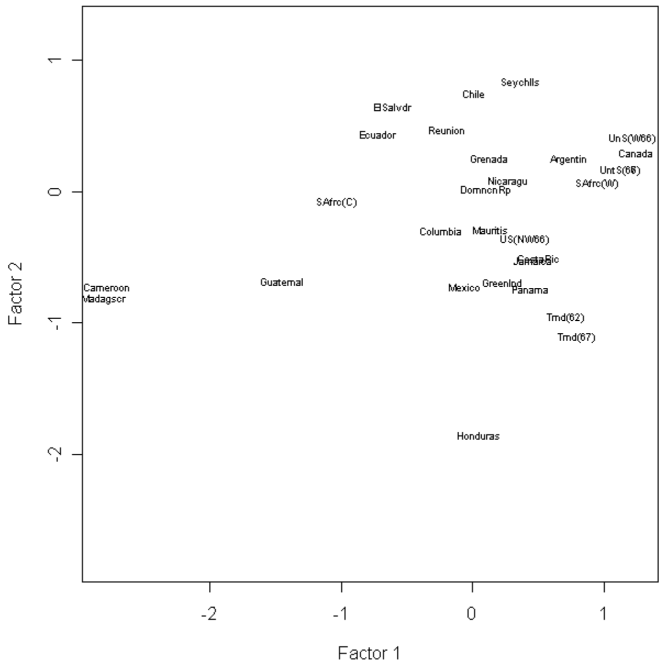
```
library(scatterplot3d)
scatterplot3d(scores, angle=35, col.grid="lightblue", main="Grafica de las
puntuaciones", pch=20)
```

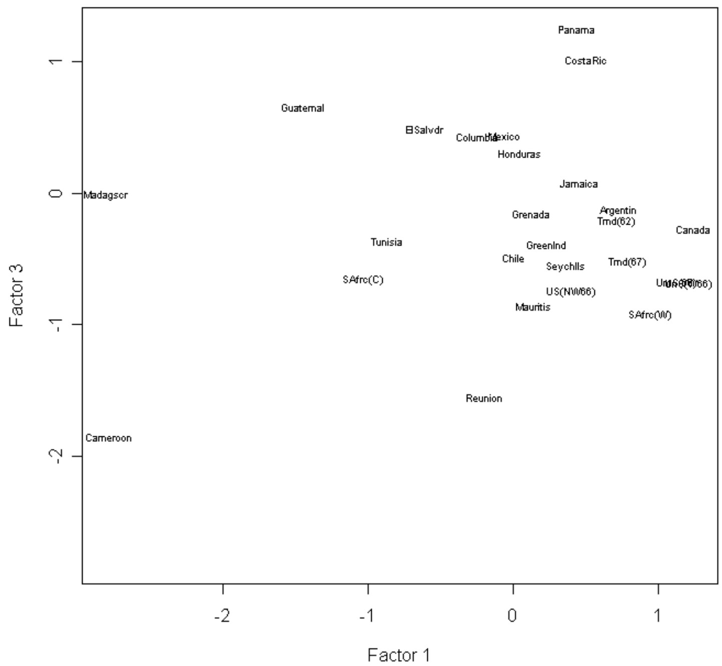


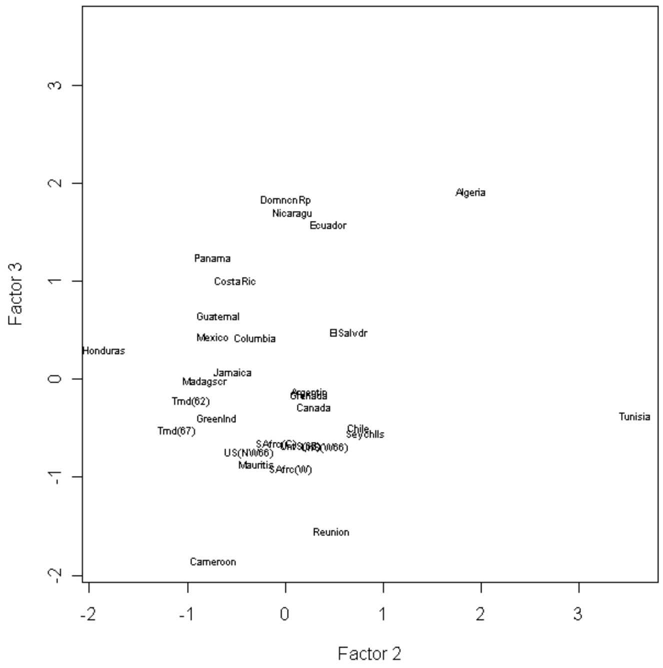

```
pairs(scores)
```




```
par(pty="s")
plot(scores[,1],scores[,2],
ylim=range(scores[,1]),
xlab="Factor 1",ylab="Factor 2",type="n",lwd=2)
text(scores[,1],scores[,2],
labels=abbreviate(row.names(life),minlength=8),cex=0.6,lwd=2)

par(pty="s")
plot(scores[,1],scores[,3],
ylim=range(scores[,1]),
xlab="Factor 1",ylab="Factor 3",type="n",lwd=2)
text(scores[,1],scores[,3],
labels=abbreviate(row.names(life),minlength=8),cex=0.6,lwd=2)

par(pty="s")
plot(scores[,2],scores[,3],
ylim=range(scores[,2]),
xlab="Factor 2",ylab="Factor 3",type="n",lwd=2)
text(scores[,2],scores[,3],
labels=abbreviate(row.names(life),minlength=8),cex=0.6,lwd=2)
```


Seychlls
Chile
ElSalvdr
Reunion
Ecuador
UnS(W66)
Canada
Grenada
Argentin
UntS(66)
Nicaragu
SAfrc(W)
DomncnRp
SAfrc(C)
Columbia
Mauritis
US(NW66)
CostaRic
Jamaica
Cameroon
Guatemal
Mexico
Greenlnd
Panama
Madagscr
Trnd(62)
Trnd(67)
Honduras
Factor 2
Factor 1


Panama
CostaRic
Guatemal
ElSalvdr
Columbia
Mexico
Honduras
Jamaica
Madagscr
Argentin
Trnd(62)
Grenada
Canada
Tunisia
Greenlnd
Chile
Seychlls
Trnd(67)
SAfrc(C)
US(NW66)
Mauritis
SAfrc(W)
Reunion
Cameroon
Factor 3
Factor 1


Factor 3
Factor 2
-2
-1
0
1
2
3
Algeria
DomncnRp
Nicaragu
Ecuador
Panama
CostaRic
Guatemal
ElSalvdr
Mexico
Columbia
Honduras
Jamaica
Madagscr
Grenada
Canada
Tmd(62)
GreenInd
Tunisia
Tmd(67)
Chile
Seychlls
SAfrc(C)
US(NW66)
Mauritis
SAfrc(W)
Reunion
Cameroon

 # Leo la matriz de correlación de los datos

```
drogasusas.cor <- read.table("c:\\...\\drogasFA.txt")
dimnames(drogasusas.cor)[[1]] <- c("cigarettes", "beer", "wine", "liquor",
"cocaine", "tranquillizers", "drug store medication", "heroin", "marijuana",
"hashish", inhalants", "haluucinogenics", "amphetamine")
dimnames(drogasusas.cor)[[2]] <- dimnames(drogasusas.cor)[[1]]

drogasusas.cor
                 cigar beer  wine  liquor   coca    tranq   medin  heroin marija
hashish
cigarettes       1.000 0.447 0.422  0.435   0.114   0.203   0.091  0.082   0.513   0.304
beer             0.447 1.000 0.619  0.604   0.068   0.146   0.103  0.063   0.445   0.318
wine             0.422 0.619 1.000  0.583   0.053   0.139   0.110  0.066   0.365   0.240
liquor           0.435 0.604 0.583  1.000   0.115   0.258   0.122  0.097   0.482   0.368
cocaine          0.114 0.068 0.053  0.115   1.000   0.349   0.209  0.321   0.186   0.303
tranquillizers   0.203 0.146 0.139  0.258   0.349   1.000   0.221  0.355   0.315   0.377
medication       0.091 0.103 0.110  0.122   0.209   0.221   1.000  0.201   0.150   0.163
heroin           0.082 0.063 0.066  0.097   0.321   0.355   0.201  1.000   0.154   0.219
marijuana        0.513 0.445 0.365  0.482   0.186   0.315   0.150  0.154   1.000   0.534
hashish          0.304 0.318 0.240  0.368   0.303   0.377   0.163  0.219   0.534   1.000
inhalants        0.245 0.203 0.183  0.255   0.272   0.323   0.310  0.288   0.301   0.302
haluucinogenics  0.101 0.088 0.074  0.139   0.279   0.367   0.232  0.320   0.204   0.368
amphetamine      0.245 0.199 0.184  0.293   0.278   0.545   0.232  0.314   0.394   0.467

                          inhal         haluuci     ampheta
cigarettes                0.245          0.101       0.245
beer                      0.203          0.088       0.199
wine                      0.183          0.074       0.184
liquor                    0.255          0.139       0.293
cocaine                   0.272          0.279       0.278
tranquillizers            0.323          0.367       0.545
medication                0.310          0.232       0.232
heroin                    0.288          0.320       0.314
marijuana                 0.301          0.204       0.394
hashish                   0.302          0.368       0.467
inhalants                 1.000          0.340       0.392
haluucinogenics           0.304          1.000       0.511
amphetamine               0.392          0.511       1.000
```

 # Se analiza a partir de la matriz de correlaciones de 1 a 6 factores como solución

```
drogasusas.fa <- vector("list",6)
for(i in 1:6) {
drogasusas.fa[[i]] <- factanal(covmat=drogasusas.cor, factors=i, method="mle")
 }
```

 # Se considera la solución con 6 factores como la adecuada

```
drogasusas.fa[[6]]
Call:
factanal(factors = i, covmat = drogasusas.cor, method = "mle")

Uniquenesses:
           cigarettes                  beer                  wine
                0.562                 0.368                 0.374
               liquor               cocaine        tranquillizers
                0.411                 0.681                 0.525
drug store medication                heroin             marijuana
                0.748                 0.664                 0.322
              hashish             inhalants       haluucinogenics
                0.010                 0.595                 0.633
          amphetamine
                0.005

Loadings:
                      Factor1 Factor2 Factor3 Factor4 Factor5 Factor6
cigarettes             0.494                           0.411
beer                   0.775                           0.113
wine                   0.785
liquor                 0.721   0.123   0.105   0.115   0.162
cocaine                        0.519           0.132           0.159
tranquillizers         0.131   0.563   0.321   0.103   0.144
drug store medication          0.248                           0.424
heroin                         0.540   0.101                   0.179
marijuana              0.428   0.160   0.154   0.260   0.606   0.103
hashish                0.244   0.281   0.190   0.877   0.195
inhalants              0.171   0.324   0.165           0.149   0.465
```

```
haluucinogenics               0.401   0.344   0.186           0.231
amphetamine           0.151   0.341   0.888   0.141   0.139   0.166

               Factor1 Factor2 Factor3 Factor4 Factor5 Factor6
SS loadings      2.301   1.448   1.133   0.956   0.680   0.584
Proportion Var   0.177   0.111   0.087   0.074   0.052   0.045
Cumulative Var   0.177   0.288   0.376   0.449   0.501   0.546

The degrees of freedom for the model is 15 and the fit was 0.0148
```

 # Se calcula la diferencia entre las correlaciones observadas y predichas
 # con 6 factores
pred <- drogasusas.fa[[6]]$loadings%*%t(drogasusas.fa[[6]]$loadings) +
diag(drogasusas.fa[[6]]$uniquenesses)
round(drogasusas.cor-pred, digits=3)

```
                      cigarettes   beer   wine liquor cocaine tranquillizers
cigarettes                 0.000 -0.001  0.015 -0.018   0.010          0.000
beer                      -0.001  0.000 -0.002  0.004   0.004         -0.011
wine                       0.015 -0.002  0.000 -0.001  -0.001         -0.005
liquor                    -0.018  0.004 -0.001  0.000  -0.007          0.020
cocaine                    0.010  0.004 -0.001 -0.007   0.000          0.003
tranquillizers             0.000 -0.011 -0.005  0.020   0.003          0.000
drug store medication     -0.020 -0.001  0.007 -0.002   0.003          0.011
heroin                    -0.005  0.007  0.008 -0.018   0.002         -0.004
marijuana                  0.002  0.002 -0.004  0.003  -0.004         -0.003
hashish                    0.000  0.000  0.000  0.000   0.000          0.000
inhalants                  0.013 -0.004 -0.008  0.012  -0.002         -0.002
haluucinogenics           -0.003  0.006 -0.001 -0.005  -0.007         -0.010
amphetamine                0.000  0.000  0.000  0.000   0.000          0.000

                      drug store medication heroin marijuana hashish inhalants
cigarettes                           -0.020 -0.005     0.002       0     0.013
beer                                 -0.001  0.007     0.002       0    -0.004
wine                                  0.007  0.008    -0.004       0    -0.008
liquor                               -0.002 -0.018     0.003       0     0.012
cocaine                               0.003  0.002    -0.004       0    -0.002
tranquillizers                        0.011 -0.004    -0.003       0    -0.002
drug store medication                 0.000 -0.019     0.007       0     0.004
heroin                               -0.019  0.000     0.006       0     0.004
marijuana                             0.007  0.006     0.000       0    -0.008
hashish                               0.000  0.000     0.000       0     0.000
inhalants                             0.004  0.004    -0.008       0     0.000
haluucinogenics                       0.005  0.022     0.003       0    -0.009
amphetamine                           0.000  0.000     0.000       0     0.000

                      haluucinogenics amphetamine
cigarettes                     -0.003           0
beer                            0.006           0
wine                           -0.001           0
liquor                         -0.005           0
cocaine                        -0.007           0
tranquillizers                 -0.010           0
drug store medication           0.005           0
heroin                          0.022           0
marijuana                       0.003           0
hashish                         0.000           0
inhalants                       0.027           0
haluucinogenics                 0.000           0
amphetamine                     0.000           0
```

```
# Se calcula la diferencia entre las correlaciones observadas y predichas
# con 3 factores
pred <- drogasusas.fa[[3]]$loadings%*%t(drogasusas.fa[[3]]$loadings)+
diag(drogasusas.fa[[3]]$uniquenesses)
round(drogasusas.cor-pred, digits=3)

                      cigarettes   beer   wine liquor cocaine tranquillizers
cigarettes                 0.000 -0.001  0.009 -0.013   0.011          0.010
beer                      -0.001  0.000 -0.001  0.002   0.002         -0.014
wine                       0.009 -0.001  0.000  0.000  -0.002         -0.004
liquor                    -0.013  0.002  0.000  0.000  -0.008          0.023
cocaine                    0.011  0.002 -0.002 -0.008   0.000          0.029
tranquillizers             0.010 -0.014 -0.004  0.023   0.029          0.000
drug store medication     -0.011  0.000  0.012 -0.018   0.038         -0.022
heroin                    -0.004  0.005  0.013 -0.020   0.081          0.024
marijuana                  0.002 -0.001  0.001 -0.001  -0.002         -0.001
hashish                   -0.026  0.019 -0.017  0.013   0.040         -0.017
inhalants                  0.039 -0.003 -0.007 -0.002   0.028         -0.032
haluucinogenics           -0.016  0.010  0.004 -0.015  -0.026         -0.056
amphetamine                0.002 -0.007  0.002  0.005  -0.077          0.040

                      drug store medication heroin marijuana hashish inhalants
cigarettes                           -0.011 -0.004     0.002  -0.026     0.039
beer                                  0.000  0.005    -0.001   0.019    -0.003
wine                                  0.012  0.013     0.001  -0.017    -0.007
liquor                               -0.018 -0.020    -0.001   0.013    -0.002
cocaine                               0.038  0.081    -0.002   0.040     0.028
tranquillizers                       -0.022  0.024    -0.001  -0.017    -0.032
drug store medication                 0.000  0.021     0.006  -0.040     0.117
heroin                                0.021  0.000     0.006  -0.035     0.037
marijuana                             0.006  0.006     0.000   0.001     0.002
hashish                              -0.040 -0.035     0.001   0.000    -0.031
inhalants                             0.117  0.037     0.002  -0.031     0.000
haluucinogenics                       0.003 -0.002    -0.003   0.037    -0.017
amphetamine                          -0.038 -0.050    -0.002   0.009    -0.009

                      haluucinogenics amphetamine
cigarettes                     -0.016       0.002
beer                            0.010      -0.007
wine                            0.004       0.002
liquor                         -0.015       0.005
cocaine                        -0.026      -0.077
tranquillizers                 -0.056       0.040
drug store medication           0.003      -0.038
heroin                         -0.002      -0.050
marijuana                      -0.003      -0.002
hashish                         0.037       0.009
inhalants                       0.019      -0.009
haluucinogenics                 0.000       0.045
amphetamine                     0.045       0.000
```

```
# Se calcula la diferencia entre las correlaciones observadas y predichas
# con 4 factores
pred <- drogasusas.fa[[4]]$loadings%*%t(drogasusas.fa[[4]]$loadings)+
diag(drogasusas.fa[[4]]$uniquenesses)
round(drogasusas.cor-pred, digits=3)

                      cigarettes   beer   wine liquor cocaine tranquillizers
cigarettes                 0.000 -0.001  0.009 -0.012   0.010          0.008
beer                      -0.001  0.000 -0.001  0.001   0.001         -0.016
wine                       0.009 -0.001  0.000  0.000  -0.001         -0.005
liquor                    -0.012  0.001  0.000  0.000  -0.005          0.029
cocaine                    0.010  0.001 -0.001 -0.005   0.000          0.019
tranquillizers             0.008 -0.016 -0.005  0.029   0.019          0.000
drug store medication     -0.014 -0.002  0.012 -0.015  -0.015         -0.021
heroin                    -0.007  0.003  0.014 -0.016   0.003          0.023
marijuana                  0.001 -0.001  0.001 -0.001  -0.003         -0.001
hashish                   -0.023  0.018 -0.020  0.018   0.032          0.000
inhalants                  0.037 -0.005 -0.008  0.001  -0.016         -0.027
haluucinogenics           -0.019  0.007  0.002 -0.009  -0.023         -0.023
amphetamine                0.000  0.000  0.000  0.000   0.000          0.000

                      drug store medication heroin marijuana hashish inhalants
cigarettes                           -0.014 -0.007     0.001  -0.023     0.037
beer                                 -0.002  0.003    -0.001   0.018    -0.005
wine                                  0.012  0.014     0.001  -0.020    -0.008
liquor                               -0.015 -0.016    -0.001   0.018     0.001
cocaine                              -0.015  0.003    -0.003   0.032    -0.016
tranquillizers                       -0.021  0.023    -0.001   0.000    -0.027
drug store medication                 0.000 -0.018     0.004  -0.042     0.096
heroin                               -0.018  0.000     0.003  -0.038     0.005
marijuana                             0.004  0.003     0.000   0.000     0.000
hashish                              -0.042 -0.038     0.000   0.000    -0.028
inhalants                             0.096  0.005     0.000  -0.028     0.000
haluucinogenics                       0.013  0.010    -0.002   0.059    -0.004
amphetamine                           0.000  0.000     0.000   0.000     0.000

                      haluucinogenics amphetamine
cigarettes                     -0.019           0
beer                            0.007           0
wine                            0.002           0
liquor                         -0.009           0
cocaine                        -0.023           0
tranquillizers                 -0.023           0
drug store medication           0.013           0
heroin                          0.010           0
marijuana                      -0.002           0
hashish                         0.059           0
inhalants                       0.032           0
haluucinogenics                 0.000           0
amphetamine                     0.000           0
```

ANALISIS FACTORIAL (con SAS)

```
/* Analisis Factorial */
options ls=80 nodate nonumber;
title 'Analisis Factorial de datos de mortalidad';
data muerte;
infile 'c:\...\datFAsas.txt';
input m0 m25 m50 m75 w0 w25 w50 w75;
run;

proc factor data=muerte nfactors=3 rotate=varimax;
run;
```

```
                 Analisis Factorial de datos de mortalidad

                           The FACTOR Procedure
                 Initial Factor Method: Principal Components

                     Prior Communality Estimates: ONE


       Eigenvalues of the Correlation Matrix: Total = 8  Average = 1

               Eigenvalue    Difference    Proportion    Cumulative

          1    5.60241029    4.24422874        0.7003        0.7003
          2    1.35818155    0.85885454        0.1698        0.8701
          3    0.49932700    0.19120110        0.0624        0.9325
          4    0.30812590    0.15343627        0.0385        0.9710
          5    0.15468962    0.09605584        0.0193        0.9903
          6    0.05863378    0.04581216        0.0073        0.9977
          7    0.01282163    0.00701139        0.0016        0.9993
          8    0.00581023                      0.0007        1.0000

            3 factors will be retained by the NFACTOR criterion.


                              Factor Pattern

                       Factor1         Factor2         Factor3

            m0         0.84572        -0.50388        -0.00222
            m25        0.79338        -0.26526         0.35122
            m50        0.86398         0.20688         0.33902
            m75        0.67493         0.63213         0.22620
            w0         0.83139        -0.50529        -0.16243
            w25        0.97181        -0.11019        -0.16718
            w50        0.94762         0.17659        -0.18122
            w75        0.72258         0.54119        -0.35028
```

Variance Explained by Each Factor

Factor1	Factor2	Factor3
5.6024103	1.3581815	0.4993270

Final Communality Estimates: Total = 7.459919

m0	m25	m50	m75
0.96913658	0.82316394	0.90419710	0.90628012

w0	w25	w50	w75
0.97292213	0.98450040	0.96201103	0.93770753

Orthogonal Transformation Matrix

	1	2	3
1	0.71166	0.53803	0.45174
2	-0.69599	0.62747	0.34912
3	-0.09562	-0.56286	0.82100

Rotated Factor Pattern

	Factor1	Factor2	Factor3
m0	0.95277	0.14010	0.20431
m25	0.71565	0.06273	0.55414
m50	0.43846	0.40383	0.74086
m75	0.01873	0.63245	0.71129
w0	0.95888	0.22168	0.06581
w25	0.78427	0.54781	0.26329
w50	0.56880	0.72265	0.34095
w75	0.17106	0.92551	0.22778

Variance Explained by Each Factor

Factor1	Factor2	Factor3
3.4998442	2.3146695	1.6454052

Final Communality Estimates: Total = 7.459919

m0	m25	m50	m75
0.96913658	0.82316394	0.90419710	0.90628012

w0	w25	w50	w75
0.97292213	0.98450040	0.96201103	0.93770753