1 Considére los datos en el archivo "cars.dat". Estos datos corresponden a calificaciones promedio que cuarenta personas le asignaron a 23 modelos de automóviles y se tiene que las variables (cualidades) evaluadas fueron

	Columna	Cualidad evaluada	Abreviatura (nombre)
X_1	1	Economía	Economy
X_2	2	Servicio	Service
X_3	3	No depreciación de su valor	Value
X_4	4	Precio	Price
X_5	5	El diseño (la apariencia)	Design
X_6	6	aspecto deportivo	Sporty
X_7	7	Grado de Seguridad del vehículo	Safety
X_8	8	Facilidad de Manejo	Easy

Las calificaciones van desde 1 (muy bueno) a 6 (muy malo).

Para este conjunto de datos, la función "factcor.R" lleva a cabo un análisis de factores iterando 5 veces el algoritmo descrito en el curso como "método de factores principales" (PFM), para encontrar $\hat{Q}_Y$ y $\hat{\Psi}_Y$ a partir de la relación

$$\hat{\mathcal{R}} = \hat{Q}_Y \hat{Q}_Y' + \hat{\Psi}_Y .$$

En principio, dada la fórmula en la página 22 para cálculo de los grados de libertad,

$$(p=8) \quad \left.\begin{array}{l} k=2 \Rightarrow d=13 \\ k=3 \Rightarrow d=7 \\ k=4 \Rightarrow d=2 \\ k=5 \Rightarrow d<0 \; ! \end{array}\right\} \cdots\cdots (I)$$

de forma que podríamos tomar hasta 4 factores para llevar a cabo el análisis.

La función "factcor.R" estima preliminarmente a Ψ_Y usando el método (2) en la página 28 de las notas. Hay una función dentro de "factcor.R" (la función "factpf") que en principio propone construir $\hat{Q}_Y$ con aquellas columnas de Γ en la descomposición $\hat{\mathcal{R}}-\hat{\Psi}=\Gamma\Lambda\Gamma'$, para los cuales, los correspondientes valores propios en Δ sean positivos, es decir $\hat{Q}_Y = \Pi_1 \Delta_1^{1/2}$

donde $\gamma_1, \ldots \gamma_k$ son las primeras k_0 columnas de Π

que fueron consideradas con

$$k_0 = \max\{ l \in \{1, 2, \ldots p\} : \lambda_l > 0 \}.$$

cada iteración del algoritmo PFM tendría que llevarse a cabo usando $k = \min(k_0, k_1)$

con k_1 determinado al examinar (I) en la página anterior.

Aunque se puede modificar la función "factor.R" para que considere usar $k > 2$, se procedió a hacer el análisis con $k=2$ para lo cual se obtiene que

$$\frac{\hat{\lambda}_1^R}{p} + \frac{\hat{\lambda}_2^R}{p} = \frac{\hat{\lambda}_1^R + \hat{\lambda}_2^R}{8} = 0.8702$$

= proporción de la varianza muestral total explicada por los dos primeros factores

En la salida de la función estan tres tablas, la primera resume el análisis usando $K=2$ factores sin utilizar rotaciones, la segunda resume un análisis usando $K=2$ factores y utilizando una rotación seleccionada en forma manual, la tercera resume un análisis usando $K=2$ factores y utilizando la rotación varimax.

Las primeras dos columnas en cada tabla son las columnas de la matriz $\hat{Q}_Y$ la tercera columna de cada tabla presenta las comonalidades obtenidas usando el modelo correspon-diente a la tabla, la cuarta columna presenta las varianzas específicas (las varianzas de los factores específicos $\mathbb{U}$). Recordemos (página 18 de las notas) que las entradas de $\hat{Q}_Y$ son las correlaciones muestrales entre las componentes de $\mathbb{F}$ y las componentes del vector de datos estandarizados. La figura "PFM cars.pdf" presenta gráficas de las parejas ordenadas

$$(\hat{q}^{\gamma}_{i1}, \hat{q}^{\gamma}_{i2}) = (r_{X_iF_1}, r_{X_iF_2}) \qquad i=1,2,\ldots,8$$

correlación entre F_1 y X_i

correlación entre F_2 y X_i

La primera gráfica (izquierda) corresponde a la primera tabla en la salida de "factcar.R", la segunda gráfica (a la derecha) corresponde a la segunda tabla en la salida de "factcar.R"