$\Gamma_1 = (\gamma_1, \ldots, \gamma_k)$; $\Delta_1 = \mathrm{diag}(\lambda_1, \ldots, \lambda_k)$.

Finalmente, se corrige $\hat{\Psi}_Y$ a través de

$$\hat{\psi}_{jj} = 1 - \sum_{l=1}^{k} \hat{q}_{jl}^2 \; ; \; j = 1, 2, \ldots p. \cdots (\text{phi})$$

Notemos que el procedimiento descrito puede iterarse:

(I) Usando (phi) construimos $\hat{\Psi}$.

(II) Calculamos la matriz de correlaciones reducida $\hat{R} - \hat{\Psi}$.

(III) Calculamos la descomposición de Jordan de $\hat{R} - \hat{\Psi}$, consideramos los valores propios $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$ y los vectores propios $\gamma_1, \ldots, \gamma_p$.

(IV) Encontramos k t.q. $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k > 0$ y definimos los pesos del factor l

$$\hat{\underset{\sim}{q}}_l = \sqrt{\lambda_l}\, \gamma_l \; ; \; l = 1, 2, \ldots, k$$

estos son las columnas de $\hat{Q}_Y$.

VI Usando las entradas de $\hat{Q}_Y$ regresamos a calcular (phi).

Una regla de paro podría ser detener la repetición de estos pesos cuando los valores $\{\hat{\psi}_{jj}\}_{j=1}^{p}$ se estabilicen ó converjan.

<u>ejemplo</u>: Datos de mediciones de billetes en el banco suizo

La matriz de correlaciones muestral es

$$\hat{R} = \begin{pmatrix} 1.0 & 0.2313 & 0.1518 & -0.1898 & -0.0613 & 0.1943 \\ 0.2313 & 1.0 & 0.7433 & 0.4138 & 0.3623 & -0.5032 \\ 0.1518 & 0.7433 & 1.0 & 0.4868 & 0.4007 & -0.5165 \\ -0.1898 & 0.4138 & 0.4868 & 1.0 & 0.1419 & -0.6230 \\ -0.0613 & 0.3623 & 0.4007 & 0.1419 & 1.0 & -0.5940 \\ 0.1943 & -0.5032 & -0.5165 & -0.6230 & -0.5940 & 1.0 \end{pmatrix}$$

Los estimadores iniciales de las comunalidades se toman como $\hat{h}_j^2 = \max\{|r_{x_j x_l}| : l \neq j \; l \in \{1, 2, \dots, p\}\}$, estas quedan dadas por

$$\underset{\sim}{\hat{h}} = (\hat{h}_1^2, \dots, \hat{h}_6^2) = (0.2313, 0.7433, 0.7433, 0.6230, 0.5940, 0.6230)$$

correspondientemente los estimadores preliminares de ψ_{jj} son $\hat{\psi}_{jj} = 1 - h_j^2$

$$\hat{\underset{\sim}{\psi}} = (\hat{\psi}_{11}, \ldots, \hat{\psi}_{66}) = (0.7687, 0.2567, 0.2567, 0.3770, 0.4060, 0.3770)$$

La matriz de correlación reducida es $\hat{R} - \mathrm{diag}(\underset{\sim}{\hat{\psi}})$

$$= \hat{R} - \hat{\psi}$$

$$= \begin{pmatrix} 0.2313 & 0.2313 & 0.1518 & -0.1898 & -0.0613 & 0.1943 \\ 0.2313 & 0.7433 & 0.7433 & 0.4138 & 0.3623 & -0.5032 \\ 0.1518 & 0.7433 & 0.7433 & 0.4868 & 0.4007 & -0.5165 \\ -0.1898 & 0.4138 & 0.4868 & 0.6230 & 0.1419 & -0.6230 \\ -0.0613 & 0.3623 & 0.4007 & 0.1419 & 0.5940 & -0.5940 \\ 0.1943 & -0.5032 & -0.5165 & -0.6230 & -0.5940 & 0.6230 \end{pmatrix}$$

Calculamos la descomposición de Jordan de $\hat{R} - \hat{\psi}$, los valores propios son

$$\underset{\sim}{\lambda} = (2.6214, 0.7232, 0.4765, 0.0054, -0.0845, -0.184)$$

Los últimos dos valores propios son negativos, de acuerdo al algoritmo propuesto se podría tomar $K = K_1 = 4$ para continuar, pero este valor de K podría <u>no</u> tener sentido de acuerdo a la discusión sobre los grados de libertad d que hicimos en las páginas 22 y 23.

Asumiendo que hicimos este análisis para d antes de hechar a andar el "metodo de factores principales", sea K_0 el más grande valor de K tal que $d \geq 0$, entonces podemos continuar tomando $K = \min(K_0, K_1)$[1].

El hecho de que existan valores propios negativos quiere decir que $\hat{R} - \hat{\Psi}$ no es positivo definido, esta situación puede aparecer a lo largo de las iteraciones del algoritmo, por lo cual es necesario tomarla en cuenta en los programas de cómputo. Para nuestro caso si $p=6$ $K_0 = 3$ $\Rightarrow d=0$ (que no resulta con una solución interpretable), entonces conviene tomar $K_0 = 2$, luego $K = \min(2, 4) = 2$ para continuar con el algoritmo.

(1) En la siguiente iteración del algoritmo K_1 puede cambiar, pero al tomar de nuevo $K = \min(K_0, K_1)$ podemos continuar

$$\Gamma_1 \Lambda_1^{1/2} = \begin{pmatrix} 0.0011 & -0.6225 \\ -0.4832 & -0.4509 \\ -0.5019 & -0.3314 \\ -0.3974 & 0.3488 \\ -0.3542 & 0.1660 \\ 0.4807 & -0.3871 \end{pmatrix} \begin{pmatrix} \sqrt{2.6214} & 0 \\ 0 & \sqrt{0.7232} \end{pmatrix} = \hat{Q},$$

$$\hat{Q} = \begin{pmatrix} 0.0018 & -0.5294 \\ -0.7824 & -0.3835 \\ -0.8127 & -0.2818 \\ -0.6435 & 0.2967 \\ -0.5736 & 0.1412 \\ 0.7783 & -0.3292 \end{pmatrix}$$

La corrección a $\hat{\Psi}$ es $\hat{\Psi}_{11} = 1-\left(0.0018^2 + (-0.5294)^2\right)$

$\hat{\Psi}_{22} = 1-\left((-0.7824)^2 + (-0.3835)^2\right)$, ...

$\hat{\Psi}_{66} = 1-\left(0.7783^2 + (-0.3292)^2\right)$

$$\hat{\Psi} = \begin{pmatrix} \hat{\Psi}_{11} & 0 & 0 & 0 & 0 & 0 \\ 0 & \hat{\Psi}_{22} & 0 & \cdot & \cdot & \cdot \\ \vdots & & & & & \\ \cdot & \cdot & \cdot & \cdot & & \hat{\Psi}_{pp} \end{pmatrix}$$

y luego se repiten II, III, ... etc.

Después de 10 iteraciones se obtiene

	Estimaciones de los pesos de los Factores		Comunalidad	Varianzas Específicas
	$\hat{q}_1$	$\hat{q}_2$	$\hat{h}_j^2$	$\hat{\psi}_{jj} = 1 - \hat{h}_j^2$
X_1 longitud	+0.0047	−0.5369	0.2883	0.711
X_2 ancho (izq.)	−0.79	−0.4157	0.7975	0.202
X_3 ancho (der.)	−0.7976	−0.2983	0.7253	0.274
X_4 long. borde inferior	−0.5920	0.1929	0.3877	0.612
X_5 long. borde superior	−0.5106	0.1069	0.2722	0.727
X_6 long. diag.	0.8816	−0.4482	0.9782	0.021

Por último y de acuerdo con la discusión en la página 19, podemos usar una rotación de los ejes, para tratar de obtener interpretabilidad

Estimaciones de los pesos de los Factores

	$\hat{q}_1$	$\hat{q}_2$	Comunalidad $\hat{h}_j^2$	Varianzas Específicas $\hat{\psi}_{jj} = 1 - \hat{h}_j^2$
X_1 longitud	+0.0047	−0.5369	0.2883	0.711
X_2 ancho (izq.)	−0.79	−0.4157	0.7975	0.202
X_3 ancho (der.)	−0.7976	−0.2983	0.7253	0.274
X_4 long. borde inferior	−0.5920	0.1929	0.3877	0.612
X_5 long. borde superior	−0.5106	0.1069	0.2722	0.727
X_6 long. diag.	0.8816	−0.4482	0.9782	0.021

Por último y de acuerdo con la discusión en la página 19, podemos usar una rotación de los ejes, para tratar de obtener interpretabilidad

Método de componentes Principales (PCM)

Para dar una forma de estimación alternativa al método de factores principales, comencemos por aproximar Q (en lugar de aproximar Ψ, como lo hace el PFM). Para lo anterior sea $\hat{\Sigma} = \hat{\Gamma}\hat{\Lambda}\hat{\Gamma}'$ la descomposición de Jordan de $\hat{\Sigma}$.

Entonces el método de componentes principales propone aproximar $\mathbb{Q}$ con

$$(\text{PCM1}) \cdots \quad \hat{\mathbb{Q}} = \hat{\Gamma}_1 \hat{\Lambda}_1^{1/2},$$

donde $\hat{\Gamma}_1 = (\hat{\gamma}_1, \ldots, \hat{\gamma}_k)$ y $\Lambda_1 = \begin{pmatrix} \hat{\lambda}_1 & 0 & 0 & \cdots & 0 \\ 0 & \hat{\lambda}_2 & 0 & \cdots & 0 \\ \vdots & \vdots & & \ddots & \\ 0 & 0 & & & \hat{\lambda}_k \end{pmatrix}$

$\hat{\Lambda}_1$ es una matriz diagonal que contiene en su diagonal a los primeros k valores propios, $\hat{\Gamma}_1$ es de dimensiones $p \times k$ y sus columnas son los primeros k vectores propios de $\hat{\Gamma}$

Los estimadores de las varianzas especificas ψ_{jj} son entonces los elementos en la diagonal de la matriz $\hat{\Sigma} - \hat{\mathbb{Q}}\hat{\mathbb{Q}}'$

$$(\text{PCM2}) \cdots \quad \hat{\Psi} = \begin{pmatrix} \hat{\psi}_{11} & 0 & 0 & \cdots & 0 \\ 0 & \hat{\psi}_{22} & 0 & \cdots & 0 \\ \vdots & \vdots & & & \\ 0 & 0 & & & \hat{\psi}_{pp} \end{pmatrix} ; \quad \hat{\psi}_{jj} = \hat{\Sigma}_{jj} - \sum_{l=1}^{k} \hat{q}_{jl}^2 .$$

Para medir la bondad de estas aproximaciones consideremos la matriz $\hat{\Sigma} - (\hat{Q}\hat{Q}' + \hat{\psi})$.

Se puede probar que

$$\sum_i \sum_j (\Sigma - \hat{Q}\hat{Q}' - \hat{\psi})_{ij}^2 \leq \lambda_{k+1}^2 + \cdots + \lambda_p^2 .$$

Entonces, entre más pequeño sea el tamaño o la magnitud de los valores propios que no se consideraron, el error de la aproximación será menor. Una regla intuitiva para seleccionar k, es considerar la proporción de la varianza muestral total atribuida al j-ésimo factor

$$\widehat{VAR(X_j)} = \sum_{l=1}^{k} \hat{q}_{jl}^2 + \hat{\psi}_{jj} = \sum_{l=1}^{k} \hat{\lambda}_l \hat{\gamma}_{jl}^2 + \hat{\psi}_{jj}$$

$$= \hat{\lambda}_1 \hat{\gamma}_{j1}^2 + \hat{\lambda}_2 \hat{\gamma}_{j2}^2 + \cdots + \hat{\lambda}_k \hat{\gamma}_{jk}^2 + \hat{\psi}_{jj}$$

$$T = \sum_{j=1}^{p} \widehat{VAR(X_j)} = \hat{\lambda}_1 \sum_{j=1}^{p} \hat{\gamma}_{j1}^2 + \hat{\lambda}_2 \sum_{j=1}^{p} \hat{\gamma}_{j2}^2 + \cdots + \hat{\lambda}_k \sum_{j=1}^{p} \hat{\gamma}_{jk}^2 + \hat{\psi}_{jj}$$

$$= \hat{\lambda}_1 + \hat{\lambda}_2 + \cdots + \hat{\lambda}_k + \hat{\psi}_{jj}$$

Ya que $\hat{\Gamma}$ es ortonormal

$$1 = \frac{\hat{\lambda}_1}{T} + \frac{\hat{\lambda}_2}{T} + \cdots + \frac{\hat{\lambda}_k}{T} + \frac{\hat{\psi}_{jj}}{T}$$

de donde $\dfrac{\hat{\lambda}_j}{\sum_{i=1}^{p} \widehat{VAR}(X_i)}$ es la proporción de la varianza de los datos, explicada por el factor j ; $j = 1, 2, \ldots, K$

<u>NOTA</u>: Si todo el proceso se hace a partir de $\hat{\mathcal{R}} = \hat{\Gamma}_{\hat{\mathcal{R}}} \hat{\Lambda}_{\hat{\mathcal{R}}} \hat{\Gamma}'_{\hat{\mathcal{R}}}$, entonces

$\dfrac{\hat{\lambda}_j^{\mathcal{R}}}{p}$ es la proporción a considerar.

<u>Ejemplo</u>: En un estudio de preferencias de los consumidores, estos fueron encuestados para valorar varios atributos de un producto nuevo. Con sus respuestas, se construyó la siguiente matriz de correlaciones

X_1 = Gusto. X_2 = Relación calidad-Precio.

X_3 = Sabor. X_4 = Posibilidad de usarse como botana.

X_5 = Grado de energías que provee

$$
\begin{array}{c}
X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5
\end{array}
\begin{pmatrix}
1.00 & 0.02 & \boxed{0.96} & 0.42 & 0.01 \\
0.02 & 1.0 & 0.13 & 0.71 & \boxed{0.85} \\
0.96 & 0.13 & 1.0 & 0.5 & 0.11 \\
0.42 & 0.71 & 0.5 & 1.0 & \boxed{0.79} \\
0.01 & 0.85 & 0.11 & 0.79 & 1.0
\end{pmatrix} = \hat{R}
$$

De esta matriz se puede apreciar que las variables 1 y 3, así como las variables 2 y 5 tienen alta correlación. La variable 4 tiene más correlacion con las variables 2 y 5 que con las variables 1 y 3. Podríamos pensar en grupos de variables $\{2,5\}, \{1,3\}, \{4\}$ ó también $\{2,5,4\}, \{1,3\}$, lo anterior sugiere considerar $K=3$ ó 2 factores en un modelo de factores.

Los valores propios de $\hat{R}$ son

$(2.85, 1.80, 0.204, 0.102, 0.033)$,

con esta información obtenemos que un modelo con $K=2$ factores comúnes debe describir un total de

$$\frac{\lambda_1 + \lambda_2}{p} = \frac{2.85 + 1.8}{5} = 0.93$$

de la varianza muestral. Procedemos a usar PCM para calcular (estimar) los pesos de los factores, las comunalidades y las varianzas específicas, a partir de (PCM1) y (PCM2)

Variable	Pesos de Factores $\hat{q}_1$	$\hat{q}_2$	Comunalidades $\hat{h}_j^2$	Varianzas específicas $\psi_{jj} = 1 - \hat{h}_j^2$
1 Gusto	0.56	-0.82	0.98	0.02
2 Relación Calidad-Precio	0.78	0.52	0.88	0.12
3 Sabor	0.65	-0.75	0.98	0.02
4 Posibilidad uso botana	0.94	0.10	0.89	0.11
5 Grado de energía que da	0.80	0.54	0.93	0.07
valores propios	2.85	1.81		
Proporción de varianza muestral explicada (acumulativa)	0.571	0.932		

Con las estimaciones $\hat{Q} = \begin{pmatrix} 0.56 & -0.82 \\ 0.78 & 0.52 \\ 0.65 & -0.75 \\ 0.94 & 0.10 \\ 0.80 & 0.54 \end{pmatrix}$ y

$$\hat{\Psi} = \begin{pmatrix} 0.02 & 0 & 0 & 0 & 0 \\ 0 & 0.12 & 0 & 0 & 0 \\ 0 & 0 & 0.02 & 0 & 0 \\ 0 & 0 & 0 & 0.11 & 0 \\ 0 & 0 & 0 & 0 & 0.07 \end{pmatrix}$$

calculamos

$$\hat{Q}\hat{Q}' + \hat{\Psi} = \begin{pmatrix} 1.0 & 0.01 & 0.97 & 0.44 & 0.0 \\ 0.01 & 1.0 & 0.11 & 0.79 & 0.91 \\ 0.97 & 0.11 & 1.0 & 0.53 & 0.11 \\ 0.44 & 0.79 & 0.53 & 1.0 & 0.81 \\ 0.0 & 0.91 & 0.11 & 0.81 & 1.0 \end{pmatrix}$$

y esta última matriz está muy cercana a la matriz de correlacion muestral $\hat{R}$. Las comunalidades (0.98, 0.88, 0.98, 0.89, 0.93) nos indican que los dos factores explican en forma adecuada la varianza muestral de cada una de las variables $X_1, X_2, \ldots, X_5$.

Se puede aprovechar la no unicidad de los pesos de los factores, para encontrar una rotación que permita una mejor interpretación.

<u>METODO VARIMAX PARA ENCONTRAR ROTACION</u>

(Kaiser, 1985)

Supóngase que la matriz $\mathcal{G} = \mathcal{G}(\theta)$ representa una rotación, en el sentido de las manecillas del reloj, de los ejes de coordenadas, donde θ es el ángulo que define la rotación.

Los pesos transformados (rotados) con $\mathcal{G}(\theta)$

están dados por $\hat{Q}^*_\theta = \hat{Q}\,\mathcal{G}(\theta)$. Sea $\hat{q}^*_{ij}$ el elemento i en la columna j de $\hat{Q}^*_\theta$ y sea $\tilde{q}_{ij} = (\hat{q}^*_{ij})^2$, la varianza de los cuadrados de los pesos en la columna l de $\hat{Q}^*_\theta$, $l = 1, 2, \ldots, k$ es

$$\widehat{VAR}(\tilde{q}_l) = \frac{1}{p}\sum_{j=1}^{p} \tilde{q}_{jl}^{2} - \left(\bar{\tilde{q}}_{\cdot l}\right)^2 , \quad (1)$$

donde $\bar{\tilde{q}}_{\cdot l} = \frac{1}{p}\sum_{j=1}^{p} \tilde{q}_{jl}$.

Al sumar estas varianzas para $l = 1, 2, \ldots, k$, tenemos

$$V_\theta \equiv \sum_{l=1}^{k} \widehat{VAR}(\tilde{q}_l) = \frac{1}{p}\sum_{l=1}^{k}\left\{\sum_{j=1}^{p}(\hat{q}^*_{jl})^4 - \left(\frac{1}{p}\sum_{j=1}^{p}(\hat{q}^*_{jl})^2\right)^2\right\}.$$

Kaiser (1985), propuso seleccionar θ, donde V_θ alcanza un máximo.

Para el conjunto de datos correspondiente a las preferencias de los consumidores, los pesos de los factores 1 y 2 tienen magnitudes que

(1) $y_1, \ldots, y_n$ obs. $\widehat{VAR}(Y) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2 = \frac{1}{n}\sum_{i=1}^{n} y_i^2 - \bar{y}^2$

no facilitan una interpretación de estos factores.

Al aplicar una rotación determinada por el método varimax obtenemos

	$\tilde{q}_1$	$\tilde{q}_2$	
1 Gusto	0.02	0.99	
2 Relación Calidad-Precio	0.94	-0.01	
3 Sabor	0.13	0.98	
4 Posibilidad uso botana	0.84	0.43	
5 Grado de energía	0.97	-0.02	

Podemos interpretar ahora, que el primer factor está determinado por las variables 2, 4 y 5. (este factor parece ser más bien nutricional).

Por otra parte, el segundo factor está determinado por las variables 1 y 3 (este factor parece ser más bien relacionado al gusto de los consumidores)

Factor Scores (estimación de los Factores)

Los valores estimados de los factores suelen llamarse "factor scores" y pueden ser de utilidad para análisis y diagnósticos en los modelos. Los scores son estimaciones de los factores $F_1, \dots, F_k$, para cada individuo en una muestra $x_{G1}, \dots, x_{Gn}$.

Supuesto: La distribución conjunta de $(\mathbb{X} - \mathbb{M})$ y $\mathbb{F}$ es normal multivariada.

Del modelo IV, página 7, la matriz de varianzas covarianzas conjunta de $(\mathbb{X} - \mathbb{M})$ y $\mathbb{F}$ es

$$(10) \dots \mathrm{VAR}\begin{pmatrix} \mathbb{X} - \mathbb{M} \\ \mathbb{F} \end{pmatrix} = \begin{pmatrix} \mathbb{Q}\mathbb{Q}' + \Psi & \mathbb{Q} \\ \mathbb{Q}' & \mathbb{I}_k \end{pmatrix} \quad \leftarrow \text{dims } (p+k) \times (p+k)$$

la submatriz $\mathbb{Q}\mathbb{Q}' + \Psi$ de dimensiones $p \times p$ equivale a la varianza de $\mathbb{X} - \mathbb{M}$, es decir Σ. Del supuesto de normalidad conjunta de $(\mathbb{X} - \mathbb{M})$ y $\mathbb{F}$, se tienen resultados sobre la distribución de $\mathbb{F}$ condicional a $\mathbb{X}$ (véase, por ej. Wikipedia, página sobre el tema "Multivariate Normal Distribution", o bien cualquier texto sobre

Análisis Multivariado), esta distribución condicional ($f_{\mathbb{F}|\mathbb{X}}$) es normal multivariada con vector de medios dado por

$$\mathbb{E}(\mathbb{F} \mid \mathbb{X} = x_G) = \mathbb{Q}' \Sigma^{-1} (\mathbb{X} - \mathbb{M}),$$

así como con matriz de varianzas-covarianzas

$$\mathrm{VAR}(\mathbb{F} \mid \mathbb{X} = x_G) = \mathbb{I}_K - \mathbb{Q}' \Sigma^{-1} \mathbb{Q}.$$

Para construir el estimador $\hat{f}_i$ de $\mathbb{F}_i$, reemplazamos $\mathbb{Q}$, Σ y $\mathbb{M}$ por sus correspondientes estimadores

$$\hat{f}_i = \hat{\mathbb{Q}}' \hat{\Sigma}^{-1} (x_{Gi} - \bar{x}_{G_n}) \;;\quad i = 1, 2, \ldots, n.$$

La misma idea se puede seguir si se quiere usar $\hat{\mathbb{R}}$ en lugar de $\hat{\Sigma}$, en tal caso $\mathbb{Z} \equiv \mathbb{D}^{-1/2}(\mathbb{X} - \mathbb{M})$ con $\mathbb{D} = \mathrm{diag}(\mathrm{VAR}(X_1), \ldots, \mathrm{VAR}(X_p))$ y en lugar de (2A) tenemos

$$(2B) \cdots \quad \mathrm{VAR}\begin{pmatrix} \mathbb{Z}_1 \\ \mathbb{F} \end{pmatrix} = \begin{pmatrix} \mathbb{Q}_z \mathbb{Q}_z' + \Psi_z & \mathbb{Q}_z \\ \mathbb{Q}_z & \mathbb{I}_K \end{pmatrix}$$

Los estimadores $\hat{f}_i$ en este caso son

$$\hat{f}_i = \hat{\mathbb{Q}}' \hat{\mathcal{R}}^{-1} z_i ,$$

donde $z_i \equiv \hat{D}^{-1/2}(x_i - \bar{x}_n)$, $\hat{\mathbb{Q}}$ es la matriz de pesos obtenida a partir de $\hat{\mathcal{R}}$ y $\hat{D} = \text{diag}(\widehat{VAR}(X_1), \ldots, \widehat{VAR}(X_p))$.

<u>nota</u>: Si se usa una rotación $\mathcal{G}$ para mejorar la interpretación de los factores $\mathbb{F}$, entonces se debe rotar a los factor scores

$$\hat{f}_i^* = \mathcal{G} \hat{f}_i$$