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

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.

METODO VARIMAX PARA ENCONTRAR ROTACION

(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 $\hat{Q}^{*}_{\theta} = \hat{Q}\,\mathcal{T}(\theta)$. Sea $\hat{q}^{*}_{ij}$ el elemento i en la columna j de $\hat{Q}^{*}_{\theta}$ y sea $\tilde{q}_{ij} = \left(\hat{q}^{*}_{ij}\right)^{2}$, 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

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

Kaiser (1985), propuso seleccionar θ, donde V_{θ} alcanza un máximo.

Para el conjunto de datos correspondiente a los 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 $F_1, \ldots, F_k$, para cada individuo en una muestra $x_{G1}, \ldots, x_{Gn}$.

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

$$(20) \ldots \mathrm{VAR}\begin{pmatrix} \mathbb{X} - \mathbb{M} \\ \mathbb{F} \end{pmatrix} = \begin{pmatrix} \mathbb{Q}\mathbb{Q}' + \Psi & \mathbb{Q} \\ \mathbb{Q}' & \mathbb{I}_k \end{pmatrix} \quad \leftarrow \text{dims } (p+k)\times(p+k)$$

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

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

así como con matriz de varianzas-covarianzas

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

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

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

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

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

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

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

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

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

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