ANALISIS DE FACTORES

<u>Idea</u>: Modelar la variabilidad (y dependencias entre sus componentes) de $\mathbb{X}$ en términos de un número (usualmente más pequeño que p)[a] de variables <u>no</u> observadas (variables latentes) llamadas "factores".

ejemplo: Supóngase que $\mathbb{X}$ reporta los consumos de p diferentes bienes en un estudio sobre gastos en los hogares durante cierto mes. Estos gastos podrían ser: agua, luz, internet, teléfono, gas. La forma en que las componentes de $\mathbb{X}$ varían (así como la forma en que X_i covaría con X_j ; $i,j=1,2,\ldots,p$, $i\neq j$), podría quedar

(a) $\mathbb{X}$ vector aleatorio de dimensión p $\mathbb{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix}$.
Se pretende que el número de factores sea más pequeño que p, es decir "reducir la dimensión".

explicada por dos o tres factores relacionados con la conducta social de los habitantes. Ejemplos de estos factores serían: búsqueda de comodidad, deseo de alcanzar cierta clase social, intención de cubrir necesidades básicas.

En la práctica, estos factores no observados resultan de mayor interés para quienes estudian estos fenómenos que las mismas observaciones $x_{G1}, \dots, x_{Gn}$. En general estos factores se interpretan como características comúnes de $x_{G1}, \dots, x_{Gn}$, las cuales no son observadas.

Una forma de plantear esta situación en un modelo sería si para cada x_{Gj} tenemos

$$(\mathrm{I}) \dots \quad X_j = \sum_{l=1}^{K} a_{jl} f_l + \mu_j \; ; \; j = 1, 2, \dots, P;$$

f_l ; $l = 1, 2, \dots, K$, son los factores,

se espera que K sea más pequeño (significativamente) que P

Por ejemplo, en estudios psicológicos, $\mathbb{X}$ podría representar a p resultados correspondientes a una prueba para medir "inteligencia". Un factor común a las observaciones $x_{G1}, \dots, x_{Gn}$ un nivel de "intelligencia" general, que se asume para los miembros de la población.

Para un vector aleatorio $\mathbb{X}$ de dimension p, con media $\mathbb{\mu}$ y matriz de covarianzas Σ, un modelo similar a (I) se puede escribir en notación matricial

$$\text{(II)} \quad \dots \quad \mathbb{X} = Q\mathbb{F} + \mathbb{\mu},$$

donde $\mathbb{F}$ es una matriz (vector de dimensión $k \times 1$), $\mathbb{Q}$ es una matriz de dimensiones $p \times k$ cuyos renglones son los pesos (en la ecuación I) asociados a los componentes de $\mathbb{X}$. Para (II), es usual asumir que: $\mathbb{E}(\mathbb{F}) = \mathbb{0}$ y $VAR(\mathbb{F}) = \mathbb{I}_k$, recordemos que los factores son variables aleatorias no observadas (latentes).

A continuación veremos que si se asume que los últimos $p-k$ valores propios de Σ valen 0, entonces la ecuación (II) nos sirve para expresar un modelo de factores.

Sea $\Sigma = \Gamma \Lambda \Gamma'$ la descomposición de Jordan para Σ y supongamos que $\lambda_{k+1}, \ldots, \lambda_p$ valen 0. En tal caso Σ es singular, pero podemos escribir

$$\Sigma = (\Gamma_1\ \Gamma_2) \begin{pmatrix} \Lambda_1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \Gamma_1' \\ \Gamma_2' \end{pmatrix}$$

Γ_1 $p \times k$

Γ_2 $p \times (p-k)$

$\Lambda_1 = \mathrm{DIAG}(\lambda_1, \ldots \lambda_k)$

$$= \sum_{l=1}^{k} \lambda_l \gamma_l \gamma_l',$$

γ_l columna l de Γ_1; $l = 1, 2, \ldots, k$.

Recordemos ahora que los componentes principales para Σ se definen como $Y = \Gamma'(X - \mu)$.

Podemos escribir esta relación como

$\Gamma Y = X - \mu$ es decir $X - \mu = \Gamma Y$

$$= \Gamma_1 Y_1 + \Gamma_2 Y_2$$

donde

$$\mathbb{Y} = \begin{pmatrix} \mathbb{Y}_1 \\ \mathbb{Y}_2 \end{pmatrix} = \begin{pmatrix} \mathbb{\Gamma}_1' \\ \mathbb{\Gamma}_2' \end{pmatrix} (\mathbb{X} - \mathbb{M})$$

es un vector aleatorio con momentos

$$\mathbb{E}(\mathbb{Y}) = \mathbb{\Gamma}' \,\mathbb{E}(\mathbb{X} - \mathbb{M}) = \mathbb{O}$$

$$\begin{aligned} VAR(\mathbb{Y}) &= VAR(\mathbb{\Gamma}'(\mathbb{X} - \mathbb{M})) \\ &= \mathbb{\Gamma}' \, VAR(\mathbb{X} - \mathbb{M})\, \mathbb{\Gamma} = \mathbb{\Gamma}' \Sigma \mathbb{\Gamma} \\ &= \mathbb{\Lambda} = \begin{pmatrix} \mathbb{\Lambda}_1 & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{pmatrix} . \qquad \ldots\ldots \underline{III} \end{aligned}$$

Notemos que el vector $\mathbb{Y}_2$ tiene una distribucion singular con media $\mathbb{O}$ y matriz de covarianzas $\mathbb{O}$, luego

$$\mathbb{X} - \mathbb{M} = \mathbb{\Gamma}_1 \mathbb{Y}_1$$

$$\begin{aligned} \mathbb{X} &= \mathbb{\Gamma}_1 \mathbb{\Lambda}_1^{1/2} \mathbb{\Lambda}_1^{-1/2} \mathbb{Y}_1 + \mathbb{M} \\ &= \mathbb{Q}\mathbb{F} + \mathbb{M} , \end{aligned}$$

con $\underset{(p\times k)}{\mathbb{Q}} = \mathbb{\Gamma}_1 \underset{(k\times k)}{\mathbb{\Lambda}_1^{1/2}}$ y $\mathbb{F} = \underbrace{\mathbb{\Lambda}_1^{-1/2}}_{(k\times k)} \underset{(k\times 1)}{\mathbb{Y}_1}$.

notemos que

$$\mathbb{\Sigma} = \mathbb{E}\left((\mathbb{X}-\mathbb{M})(\mathbb{X}-\mathbb{M})'\right) = \mathbb{E}(\mathbb{Q}\mathbb{F})(\mathbb{Q}\mathbb{F})')$$

$$= \mathbb{Q}\,\mathbb{E}(\mathbb{F}\mathbb{F}')\,\mathbb{Q}' = \mathbb{Q}\,\mathbb{E}(\mathbb{\Lambda}_1^{-1/2}\mathbb{Y}_1(\mathbb{Y}_1'\mathbb{\Lambda}_1^{-1/2}))$$

$$= \mathbb{Q}\,\mathbb{E}(\mathbb{\Lambda}_1^{-1/2}\mathbb{Y}_1\mathbb{Y}_1'\mathbb{\Lambda}_1^{-1/2})\,\mathbb{Q}'$$

$$= \mathbb{Q}\,\mathbb{\Lambda}_1^{-1/2}\,\mathbb{E}(\mathbb{Y}_1\mathbb{Y}_1')\,\mathbb{\Lambda}_1^{-1/2}\mathbb{Q}'$$

ecuación III pág 5 $\longrightarrow$

$$= \mathbb{Q}\,\mathbb{\Lambda}_1^{-1/2}\,\mathbb{\Lambda}_1\,\mathbb{\Lambda}_1^{-1/2}\,\mathbb{Q}' = \mathbb{Q}\mathbb{Q}'$$

$$= \sum_{j=1}^{k} \lambda_j\, \gamma_j\, \gamma_j'$$

Entonces bajo el supuesto $\lambda_{k+1} = \lambda_{k+2} = \cdots = \lambda_p$, $\mathbb{X}$ se puede expresar como una combinación lineal de k factores ($k<p$) que son no correlacionados (las componentes de $\mathbb{\Lambda}_1^{-1/2}\mathbb{Y}_1$).

<u>Problema</u>: En la práctica $\hat{\mathbb{\Sigma}}$ no resulta singular, casi siempre.

En la práctica, una parte del modelo[1] que resulta útil, es considerar que hay dos tipos de factores: (1) Factores comunes a todas las variables contempladas,

(2) Factores específicos de cada variable.

Trabajaremos con una generalización del modelo en (II) que está dada por

$$\mathbb{X} = \mathbb{Q}\,\mathbb{F} + \mathbb{U} + \mu, \quad \ldots \quad \boxed{\text{IV}}$$

donde $\mathbb{Q}$ es una matriz de dimensión $p \times k$, cuyas entradas representan pesos asociados a los factores comunes $\mathbb{F}$, este último es un vector de dimensión $k \times 1$. El vector $\mathbb{U}$ de dimensión $p \times 1$, tiene entradas que representan a los factores específicos de cada variable[2].

(1) una parte de la especificación del modelo, un supuesto.

(2) Tanto $\mathbb{F}$ como $\mathbb{U}$ son latentes, i.e. no se observaron.

Se asume además que la correlación entre los componentes de $\mathbb{F}$ vale $\mathbb{O}$, la correlación entre los componentes de $\mathbb{U}$ también vale 0 y adicionalmente $COV(\mathbb{F}, \mathbb{U}) = \mathbb{O}_{K \times p}$. Los supuestos son entonces:

$$
(2) \begin{cases}
\mathbb{E}(\mathbb{F}) = \mathbb{O}_{K \times 1} \; , \\
VAR(\mathbb{F}) = \mathbb{I}_K \; , \\
\mathbb{E}(\mathbb{U}) = \mathbb{O}_{p \times 1} \; , \\
COV(U_i, U_j) = 0 \; , \; \forall \, i \neq j \; , \\
COV(\mathbb{F}, \mathbb{U}) = \mathbb{O}_{K \times p} \; .
\end{cases}
$$

Además denotaremos $VAR(\mathbb{U}) = \Psi$ $= DIAG(\psi_{11}, \dots, \psi_{pp})$.

Para el jésimo componente de $\mathbb{X}$ tenemos

$$
(\underline{V}) \dots \quad X_j = \sum_{l=1}^{K} q_{jl} F_l + U_j + \mu_j \, , \quad j = 1, 2, \dots, p.
$$

$\mathbb{E}(X_j) = \mu_j$; $F_1, \dots, F_K$ factores comunes

U_j factor específico de X_j

$q_{j1}, \dots, q_{jk}$ pesos de $F_1, \dots, F_k$ asociados a X_j.

(IV) ó (V) se llaman "Modelo Ortogonal de Factores" ó "Modelo de Factores Ortogonales"

Usando (V) y (Z)

$$VAR(X_j) = \sum_{l=1}^{k} q_{jl}^2 + \Psi_{jj}$$

llamaremos a $h_j^2 = \sum_{l=1}^{k} q_{jl}^2$ "la comunalidad" y a Ψ_{jj} "la varianza específica".

$$\begin{aligned}
\Sigma &= \mathbb{E}\left[(\mathbb{X}-\mathbb{M})(\mathbb{X}-\mathbb{M})'\right] \\
&= \mathbb{E}\left[(\mathbb{Q}\mathbb{F}+\mathbb{U})(\mathbb{Q}\mathbb{F}+\mathbb{U})'\right] \\
&= \mathbb{Q}\,\mathbb{E}(\mathbb{F}\mathbb{F}')\mathbb{Q}' + \mathbb{E}(\mathbb{U}\mathbb{U}') \\
&= \mathbb{Q}\,VAR(\mathbb{F})\,\mathbb{Q}' + \Psi \\
&= \mathbb{Q}\mathbb{Q}' + \Psi \qquad \text{-----} \quad (Y)
\end{aligned}$$

En el modelo (IV), $\mathbb{U}$ se puede interpretar como un término de error tal que sus componentes U_i tienen la posibilidad de tener varianzas distintas, por lo que permite captar variaciones en las componentes de $\mathbb{X}$ (variaciones específicas de cada variable).

El termino μ corresponde a $\mathbb{E}(\mathbb{X})$ y el termino $Q\mathbb{F}$ tiene el objetivo de explicar la estructura de correlación entre las componentes de $\mathbb{X}$.

Nuestro objetivo será encontrar estimadores de Q y de Ψ.

Invarianza a cambios de escala

Supóngase que se re-escalan los datos para obtener $\mathbb{Y} = \mathbb{C}\mathbb{X}$ con $\mathbb{C} = \mathrm{DIAG}(c_1, \ldots, c_p)$, entonces asumiendo que en verdad tenemos que

$$\mathbb{X} = \mathbb{Q}\mathbb{F} + \mathbb{U} + \mu,$$

con los supuestos (z)

$$\mathrm{VAR}(\mathbb{Y}) = \mathbb{C}\Sigma\mathbb{C}' \overset{(y)}{=} \mathbb{C}\mathbb{Q}\mathbb{Q}'\mathbb{C}' + \mathbb{C}\Psi\mathbb{C}',$$

de manera que el mismo modelo de K factores se satisface para $\mathbb{Y}$, tomando $\mathbb{Q}_Y \equiv \mathbb{C}\mathbb{Q}$ y $\Psi_Y = \mathbb{C}\Psi\mathbb{C}'$:

$$\mathbb{Y} = \mathbb{Q}_Y\mathbb{F} + \mathbb{U}' + \mu',$$

con $\mathbb{U}' = \mathbb{C}\mathbb{U}$ y $\mu' = \mathbb{C}\mu$.

Por esta razón, en la literatura, se considera conveniente considerar los datos estandarizados $\mathbb{Y} = D^{-1/2}(\mathbb{X} - \mu)$, en cuyo caso (y) pag 9 es

$$R = \mathbb{Q}_Y\mathbb{Q}_Y' + \Psi_Y$$

<u>Resultado 1</u>: Si $\mathbf{x}_1, \ldots, \mathbf{x}_n \overset{\text{i.i.d.}}{\sim} N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

entonces los estimadores de máxima verosimilitud para $\boldsymbol{\mu}$ y $\boldsymbol{\Sigma}$ son

$$\hat{\boldsymbol{\mu}} = \overline{\mathbf{x}}_{(n)} \quad \text{y} \quad \hat{\boldsymbol{\Sigma}} = \frac{1}{n} \sum_{j=1}^{n} (\mathbf{x}_j - \overline{\mathbf{x}}_{(n)})(\mathbf{x}_j - \overline{\mathbf{x}}_{(n)})'$$

<u>Dem</u> ver las notas MLE-MULTVAR-NORMAL

<u>Resultado 2</u>: Si $\mathbf{x}_1, \ldots, \mathbf{x}_n \overset{\text{i.i.d.}}{\sim} N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

entonces $n\hat{\boldsymbol{\Sigma}} \sim \text{Wishart}_p(\boldsymbol{\Sigma}, n-1)$

Este Teorema se revisó en la página 55 de las notas para componentes principales

<u>Resultado 3</u>: Si $\mathbf{x}_1, \ldots, \mathbf{x}_n \overset{\text{i.i.d.}}{\sim} N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

entonces $\hat{\boldsymbol{\mu}} = \overline{\mathbf{x}}_{(n)}$ y $n\hat{\boldsymbol{\Sigma}}$ <u>son</u> <u>independientes</u>

<u>Dem</u> Véase Press, S. J., (2005), "Applied Multivariate Analysis, Using Bayesian and Frequentist Methods of Inference", 2nd Edition, Dover, Teorema (7.1.2), pag 181, sección 7.1.

Resultado 4 : Si $\mathbb{X}_1, \ldots, \mathbb{X}_n \overset{i.i.d.}{\sim} N_p(\mu, \Sigma)$

entonces $(\hat{\mu}, n\hat{\Sigma})$ es suficiente para

$\Theta = (\mu, \Sigma)$.

Dem Véase Press, S. J. (2005) ...,

Teorema (7.1.1), página 181, sección 7.1

Estimación de máxima verosimilitud para

Q y Ψ (Lawley y Maxwell, 1963. Anderson

y Rubin, 1956).

Argumento (caso variables discretas):

Sean $f(x_1, \ldots, x_n, \theta)$ la densidad de $X_1, \ldots, X_n$

y $f(x_1, \ldots, x_n, t, \theta)$ la densidad conjunta de

$x_1, \ldots, x_n, T(x_1, \ldots, x_n)$, donde T es una

estadístico suficiente para θ.

$f(x_1, \ldots, x_n, t, \theta) = \mathbb{P}(X_1 = x_1, \ldots, X_n = x_n, T = t; \theta)$

$\underset{\uparrow}{=} \mathbb{P}(X_1 = x_1; \ldots; X_n = x_n; \theta) = f(x_1, \ldots, x_n, \theta)$

si t es tal que $T(x_1, \ldots, x_n) = t$

$$f(x_1,\dots,x_n,\theta) = f(x_1,\dots,x_n,t,\theta) =$$

$$f(x_1,\dots,x_n \mid t,\theta)\, f(t,\theta)$$

$$\underset{\substack{\uparrow \\ t \text{ es} \\ \text{suficiente}}}{=} f(x_1,\dots,x_n \mid t)\, f(t,\theta)$$

Como el primer factor no depende de θ, para maximizar $f(x_1,\dots,x_n,\theta)$ c.r. a θ debemos maximizar $f(t,\theta)$ c.r. a θ.

Para poder aplicar este argumento al caso(1), recordemos que $\overline{\mathbf{x}}_{(n)} \sim N_p(\mu, \frac{1}{n}\Sigma)$, entonces por el resultado 3

$$f_{\hat{\mu},\, n\hat{\Sigma}}(t_1, t_2) = f_{\hat{\mu}}(t_1)\, f_{n\hat{\Sigma}}(t_2)$$

$$\overset{(2)}{=} N_p(t_1; \mu, \tfrac{1}{n}\Sigma) \cdot \text{Wishart}_p(t_2; \Sigma, n-1)$$

que es la densidad conjunta de $T(\mathbf{x}_1,\dots,\mathbf{x}_n) = (\hat{\mu}, n\hat{\Sigma})^{(3)}$

(1) $\mathbf{x}_1,\dots,\mathbf{x}_n \overset{iid}{\sim} N_p(\mu,\Sigma)$

(2) Usando: Resultado 2.

(3) $T(\mathbf{x}_1,\dots,\mathbf{x}_n)$ suficiente — Resultado 4:

$\tilde{L}_T(t, \mu, \Sigma) = f_{\hat{\mu}, n\hat{\Sigma}}(t_1, t_2)$

De acuerdo al argumento dado arriba, hay que maximizar a $\tilde{L}_T(t, \theta) = \tilde{L}_T(t, \mu, \Sigma)$ como función de μ, Σ, pero como lo que nos interesa son los parámetros Q y Ψ, entonces podemos substituir $\Sigma = QQ' + \Psi$ en $\tilde{L}_T(t, \mu, \Sigma)$ y maximizar esta última c.r.a Q y Ψ. Tomando logaritmos

$$\ell(\mu, Q, \Psi, x) = \log \tilde{L}_T(t, \mu, QQ' + \Psi)$$

$$= \log\left\{N_p\left(t_1, \mu, \tfrac{1}{n}(QQ'+\Psi)\right) \times \text{Wishart}_p\left(t_2; QQ'+\Psi, n-1\right)\right\}$$

$$= c + \left(\frac{n-p-2}{2}\right)\log|A| - \frac{n}{2}\log|QQ'+\Psi|$$

$$-\frac{1}{2}\operatorname{tr}\left[(QQ'+\Psi)^{-1}A\right] - \frac{n}{2}(\bar{x}_{(n)} - \mu)'(QQ'+\Psi)^{-1}(\bar{x}_{(n)} - \mu),$$

$A = n\hat{\Sigma}$

donde c es una constante que no depende de Q, Ψ y x. Notemos que al hacer $\mu = \hat{\mu}$ el último término se cancela, después de esto, si se deriva con respecto a Q y $\Psi_{11}, \Psi_{22}, \dots, \Psi_{pp}$ y se igualan estas derivadas a 0, se obtiene un sistema de ecuaciones

$$\mathrm{Diag}(\hat{\Psi} + \hat{Q}\hat{Q}') = \mathrm{Diag}(\hat{\Sigma}),$$

$$\hat{\Sigma}\,\hat{\Psi}^{-1}\hat{Q} = \hat{Q}(\mathbb{I} + \hat{Q}'\hat{\Psi}^{-1}Q),$$

que deben satisfacer los estimadores de máxima verosimilitud $\hat{Q}$ y $\hat{\Psi}$.

Modelo para datos estandarizados

$$\mathbb{X} = \mathbb{Q}\mathbb{F} + \mathbb{U} + \mathbb{M}$$

donde denotando $\mathcal{D} = \text{DIAG}(\text{VAR}(X_1), \ldots, \text{VAR}(X_p))$

tenemos que

$$\mathcal{D}^{-1/2}(\mathbb{X} - \mathbb{M}) = (\mathcal{D}^{-1/2}\mathbb{Q})\mathbb{F} + \mathcal{D}^{-1/2}\mathbb{U} \ldots (\text{I}).$$

$\mathcal{D}^{-1/2}\mathbb{Q} \equiv \mathbb{Q}_Y$; $\mathbb{Y} \equiv \mathcal{D}^{-1/2}(\mathbb{X} - \mathbb{M})$; $\mathcal{D}^{-1/2}\mathbb{U} = \mathbb{U}_Y$

$$\mathbb{Y} = \mathbb{Q}_Y \mathbb{F} + \mathbb{U}_Y, \quad \ldots\ldots \quad (\text{II})$$

donde

$$\begin{aligned} \text{VAR}(\mathbb{U}_Y) &= \text{COV}(\mathcal{D}^{-1/2}\mathbb{U}, \mathcal{D}^{-1/2}\mathbb{U}) \\ &= \mathcal{D}^{-1/2}\,\text{COV}(\mathbb{U},\mathbb{U})\,\mathcal{D}^{-1/2} \\ &= \mathcal{D}^{-1/2}\,\Psi\,\mathcal{D}^{-1/2}. \end{aligned}$$

Al calcular VAR(·) en ambos lados de (II) (usando los supuestos (z) en la página 8 tenemos

matriz de correlaciones de $\mathbb{X}$ $\longrightarrow$

$$\begin{aligned} \mathbb{R} &= \text{COV}(\mathcal{D}^{-1/2}\mathbb{Q}\mathbb{F}, \mathcal{D}^{-1/2}\mathbb{Q}\mathbb{F}) + \text{COV}(\mathcal{D}^{-1/2}\mathbb{U}, \mathcal{D}^{-1/2}\mathbb{U}) \\ &= \mathcal{D}^{-1/2}\mathbb{Q}\,\text{COV}(\mathbb{F},\mathbb{F})\,\mathbb{Q}'\mathcal{D}^{-1/2} + \mathcal{D}^{-1/2}\,\Psi\,\mathcal{D}^{-1/2} \\ &= \mathbb{Q}_Y \mathbb{Q}_Y' + \Psi_Y \end{aligned}$$

Para estimar Q_Y y Ψ_Y en (II)

usamos los MLE's $\hat{Q}$ y $\hat{\Psi}$ obtenidos en las páginas 15-16, de donde

$$\hat{Q}_Y = \hat{D}^{-1/2}\,\hat{Q}$$

$$\hat{\Psi}_Y = \hat{D}^{-1/2}\,\hat{\Psi}\,\hat{D}^{-1/2}$$

$$\hat{D} = \mathrm{DIAG}\left(\widehat{VAR}(X_1), \dots, \widehat{VAR}(X_p)\right)$$

Interpretación de los Factores

En el caso en que el modelo haya resultado adecuado (se haya logrado explicar un porcentaje significativo de la varianza y de las covarianzas en las componentes de $\mathbb{X}$, a través de los k factores), hay que interpretar los resultados del modelo, por lo cual es importante entender cómo correlacionan los factores con las variables originales $X_1, \ldots, X_p$. Usando los supuestos en (2), página 8, obtenemos que

$$\mathrm{COV}(\mathbb{X}, \mathbb{F}) = \Sigma_{\mathbb{X}\mathbb{F}} = \mathbb{E}\left[(\mathbb{X} - \mathbb{M})(\mathbb{F} - \mathbb{M}_{\mathbb{F}})'\right] = \mathbb{E}\left[(\mathbb{Q}\mathbb{F} + \mathbb{U})\mathbb{F}'\right]$$

$$= \mathbb{Q},$$

de forma que

$$\mathrm{CORR}(\mathbb{X}, \mathbb{F}) = D^{-1/2}\mathbb{Q}\, D_{\mathbb{F}}^{-1/2}, \quad \text{donde } D = \mathrm{diag}\{\mathrm{VAR}(X_1), \ldots, \mathrm{VAR}(X_p)\}$$

y como $\mathrm{VAR}(\mathbb{F}) = \mathbb{I}_k$ $\quad D_{\mathbb{F}} = \mathrm{diag}\{1, \ldots, 1\}$,

es decir, $\quad \mathrm{CORR}(\mathbb{X}, \mathbb{F}) = D^{-1/2}\mathbb{Q}$. $\qquad \left(P_{\mathbb{X}\mathbb{F}} \equiv \mathrm{CORR}(\mathbb{X}, \mathbb{F})\right)$

$$\widehat{P}_{\mathbb{X}\mathbb{F}} = \hat{D}^{-1/2}\hat{\mathbb{Q}} \quad , \quad \hat{\mathbb{Q}} = \text{mle para } \mathbb{Q} \text{.} \quad \cdots\cdots (\alpha)$$

Con esta estimación es posible construir gráficas similares a las que se usaron para componentes principales con las parejas ordenadas $(P_{X_i,F_j}, P_{X_i,F_k})$, $i=1,2,\ldots,p$, para estudiar dependencias.

Para el caso de datos estandarizados $\mathbb{Y} = D^{-1/2}(\mathbb{X}-\mathbb{M})$

$$\operatorname{COV}(\mathbb{Y},\mathbb{F}) = D^{-1/2}\operatorname{COV}(\mathbb{X}-\mathbb{M},\mathbb{F}') = D^{-1/2}\operatorname{COV}(\mathbb{X},\mathbb{F}')$$

$$= D^{-1/2}\mathbb{Q}$$

$$\operatorname{CORR}(\mathbb{Y},\mathbb{F}) = \mathbb{I}_p \operatorname{COV}(\mathbb{Y},\mathbb{F}')\,\mathbb{I}_K = \operatorname{COV}(\mathbb{Y},\mathbb{F}')$$

$$= D^{-1/2}\mathbb{Q} = \mathbb{Q}_Y$$

$$\widehat{P}_{\mathbb{Y}\mathbb{F}} = \widehat{\operatorname{CORR}(\mathbb{Y},\mathbb{F})} = \hat{D}^{-1/2}\hat{\mathbb{Q}} = \hat{\mathbb{Q}}_Y \quad \cdots\cdots (\beta)$$

Los pesos asociados a los factores $\mathbb{F}$ <u>no</u> son únicos:

Si $\mathbb{G}$ es una matriz t.q. $\mathbb{G}\mathbb{G}' = \mathbb{I}_K$ (matriz ortonormal) entonces podemos re-escribir ($\underline{\overline{\text{IV}}}$) (página 7) como

$$\mathbb{X} = (\mathbb{Q}\,\mathbb{G})(\mathbb{G}'\,\mathbb{F}) + \mathbb{U} + \boldsymbol{\mu} \qquad (\mathrm{IV}')$$

Si el modelo IV es "verdadero" entonces el modelo (IV') también se satisface, con pesos $\mathbb{Q}\mathbb{G}$ y con factores $\mathbb{G}'\mathbb{F}$. Se puede aprovechar este hecho, ya que multiplicar por la izquierda a $\mathbb{F}$ con una matriz ortonormal se puede interpretar como una rotación al sistema de ejes en el cual está representado el vector en $\mathbb{F}$. Si se selecciona una rotación "conveniente" se obtendrá una matriz de pesos $\mathbb{Q}\mathbb{G}$ que será más fácil de interpretar. Por otra parte, como lo establecen (α) y (β) los pesos $\mathbb{Q}$ se usan para calcular las correlaciones entre los factores $\mathbb{F}$ y las variables $X_1, \ldots, X_p$, se puede entonces buscar una rotación $\mathbb{G}$ de forma que los factores $\mathbb{G}'\mathbb{F}$ en donde se maximice la correlación con algunos grupos de variables.

nota: Debido a que para obtener los estimadores de MLE para Q y Ψ requiere la programación de métodos numéricos para resolver ecuaciones complicadas, en la práctica se estudian formas de usar la relación

$$\Sigma = QQ' + \Psi$$

para obtener estimaciones de Q y Ψ

No obstante, desde un punto de vista computacional y numérico la no unicidad de los pesos es una desventaja debido a la multiplicidad de soluciones encontrar Q y Ψ a partir de la ecuación $\Sigma = QQ' + \Psi$ no es un problema que pueda resolver algún algoritmo numérico en forma directa, por ello se requiere imponer restricciones. Si resulta que al imponer restricciones se puede encontrar una solución, esta se puede "corregir" usando una rotación "conveniente", como se mencionó antes.

Algunas restricciones usadas en la práctica son

(i) $Q'\Psi^{-1}Q$ es matriz diagonal ó

(ii) $Q'DQ$ es matriz diagonal.

La relación $\Sigma = QQ' + \Psi$ "parametriza" a Σ con un total de $p \times K + p$ parámetros, pero ya sea (i) ó (ii) introducen $\frac{1}{2}K(K+1)$ restricciones[1]

(1) p. ej. las entradas de $Q'\Psi^{-1}Q$ que están debajo de la diagonal No se requieren.

Ya que se requiere que las matrices sean diagonales

Entonces el número d de "grados de libertad" para el modelo de k factores es

$$d = \begin{pmatrix} \text{número de parámetros de } \Sigma \\ \text{sin considerar restricciones} \end{pmatrix} - \begin{pmatrix} \text{número de parametros de} \\ \Sigma \text{ consirerando} \\ \text{restricciones} \end{pmatrix}$$

$$= \frac{1}{2} p(p+1) - \left(pk + p - \frac{1}{2} k(k-1) \right)$$

$$= \frac{1}{2} (p-k)^2 - \frac{1}{2} (p+k) \quad .$$

Se requiere que $d \geq 0$ [1], para poder determinar el modelo. El caso $d=0$ da una solución única y a veces no resulta útil. En la práctica se tiene $d>0$, (lo cual nos deja más ecuaciones por resolver que el número de incognitas (parámetros), por lo cual, para determinar soluciones se usan aproximaciones.

Evaluar el número de grados de libertad d, es un aspecto importante, porque nos da una idea de una cota superior para el número de factores

(1) $d<0$ significaría que el número de parámetros en el modelo de factores es mayor que el número de parametros del modelo original.

que se espera poder usar en el modelo.

Por ejemplo para $p=4$ y $K=2 \Rightarrow d=-1$

$p=4 \quad K=1 \Rightarrow d=2$ tiene solución

$p=6 \quad K=1 \Rightarrow d=9 \qquad p=6 \quad K=2 \Rightarrow d=4$ tienen solución

$p=6 \quad K=3 \Rightarrow d=0$ hay una sola solución

ejemplo: $p=3 \quad K=1 \Rightarrow d=0$ hay una única solución

$\Sigma = QQ' + \Psi$

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix} = \begin{pmatrix} q_1^2+\psi_{11} & q_1q_2 & q_1q_3 \\ q_1q_2 & q_2^2+\psi_{22} & q_2q_3 \\ q_1q_3 & q_2q_3 & q_3^2+\psi_{33} \end{pmatrix}$$

donde $Q = \begin{pmatrix} q_1 \\ q_2 \\ q_3 \end{pmatrix} \qquad \Psi = \begin{pmatrix} \psi_{11} & 0 & 0 \\ 0 & \psi_{22} & 0 \\ 0 & 0 & \psi_{33} \end{pmatrix}$

para este ejemplo se satisface la restricción

$Q'\Psi Q$ es diagonal, ya que $K=1$.

La solución está dada por

$$q_1^2 = \frac{\sigma_{12}\sigma_{13}}{\sigma_{23}} \;; \quad q_2^2 = \frac{\sigma_{12}\sigma_{23}}{\sigma_{13}} \;; \quad q_3^2 = \frac{\sigma_{13}\sigma_{23}}{\sigma_{12}}$$

$\psi_{11} = \sigma_{11} - q_1^2$; $\psi_{22} = \sigma_{22} - q_2^2$; $\psi_{33} = \sigma_{33} - q_3^2$.

Para este caso particular, la única rotación que se puede considerar es $\mathcal{E} = -1$, de forma que se puede estudiar la solución alternativa

$$-Q = \begin{pmatrix} -q_1 \\ -q_2 \\ -q_3 \end{pmatrix}$$

Habiendo estudiado estos ejemplos con la cantidad poblacional Σ, pasamos a revisar lo que sucede en la práctica, cuando se reemplaza Σ por $\hat{\Sigma}$ (ó cuando se usan datos estandarizados y en lugar de $\hat{\Sigma}$ aparece $\hat{R}$). En tal caso tenemos

$$\hat{\Sigma} = \hat{Q}\hat{Q}' + \hat{\psi}, \quad \cdots\cdots (\delta)$$

dado un estimador $\hat{Q}$ de Q, es natural proponer

$$\hat{\psi}_{jj} = \hat{\Sigma}_{jj} - \sum_{l=1}^{k} \hat{q}_{jl}^2 .$$

$\hat{h}_j^2 \equiv \sum_{l=1}^{k} \hat{q}_{jl}^2$ es un estimador para la comunalidad h_j^2. ¿Cómo encontrar $\hat{Q}$ y $\hat{\Psi}$ t.q. (δ) se satisface?

Para el caso de datos, estandarizados, a menudo es más fácil, resolver el problema de estimar Q y Ψ. Sea $\mathcal{X}_S = \mathcal{H}\mathcal{X}\hat{D}^{-1/2}$ la matriz de datos estandarizados[1] $\left(\hat{D} = \operatorname{diag}(\widehat{VAR}(X_1), \ldots, \widehat{VAR}(X_p))\right)$.

Asumiendo que $\hat{Q} \equiv \hat{Q}_X$ y $\hat{\Psi} \equiv \hat{\Psi}_X$ satisfacen

$$\hat{\Sigma} = \hat{Q}_X \hat{Q}_X' + \hat{\Psi}_X ,$$

entonces si $\hat{\mathcal{R}} = \frac{1}{n}\mathcal{X}_S'\mathcal{X}_S$ es la matriz de correlaciones de $\mathcal{X}$

$$\hat{\mathcal{R}} = \hat{Q}_Y \hat{Q}_Y' + \hat{\Psi}_Y \quad \cdots\cdots (\xi)$$

para $\hat{Q}_Y = \hat{D}^{-1/2}\hat{Q}_X$ y $\Psi_Y = \hat{D}^{-1/2}\hat{\Psi}_X\hat{D}^{-1/2}$

(1) $\mathcal{H} = \mathbb{I}_n - \frac{1}{n}\mathbb{1}_n\mathbb{1}_n'$

La interpretación de los factores (del modelo de Factores) se formula a partir de las correlaciones $\hat{Q}_Y$ (véase la ecuación (β), página 18).

<u>ejemplo:</u> Consideremos datos $\mathcal{X}$ para los cuales

$$\mathcal{X} = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}$$ con matriz de correlaciones

estimada

$$\hat{\mathcal{R}} = \begin{pmatrix} 1 & 0.975 & 0.613 \\ 0.975 & 1 & 0.620 \\ 0.613 & 0.620 & 1 \end{pmatrix}.$$

Tomando $k=1$ tenemos que $d=0$, entonces sólo hay una solución. Planteando la ecuación (ξ)

$$\begin{pmatrix} 1 & 0.975 & 0.613 \\ 0.975 & 1 & 0.620 \\ 0.613 & 0.620 & 1 \end{pmatrix} = \begin{pmatrix} \hat{q}_1^2 + \hat{\psi}_{11} & \hat{q}_1\hat{q}_2 & \hat{q}_1\hat{q}_3 \\ \hat{q}_1\hat{q}_2 & \hat{q}_2^2 + \hat{\psi}_{22} & \hat{q}_2\hat{q}_3 \\ \hat{q}_1\hat{q}_3 & \hat{q}_2\hat{q}_3 & \hat{q}_3^2 + \hat{\psi}_{33} \end{pmatrix},$$

se obtiene $\hat{q}_1^2 = \dfrac{r_{X_1X_2}\, r_{X_1X_3}}{r_{X_2X_3}} \Rightarrow \hat{q}_1 = 0.982$

$\hat{q}_2^2 = \dfrac{r_{X_1X_2}\, r_{X_2X_3}}{r_{X_1X_3}} \Rightarrow \hat{q}_2 = 0.993$

$$\hat{a}_3^2 = \frac{r_{x_1x_3} r_{x_2x_3}}{r_{x_1x_2}} \quad \Rightarrow \quad \hat{a}_3 = 0.624$$

De forma que las estimaciones de l@s comunalidades son $\hat{h}_1^2 = (0.982)^2$; $\hat{h}_2^2 = (0.993)^2$ $\hat{h}_3^2 = (0.624)^2$. Las estimaciones de las varianzas específicas ψ_{jj} $j = 1, 2, 3$ son

$$\hat{\psi}_{11} = 1 - \hat{a}_1^2 = 0.035 \qquad \hat{\psi}_{22} = 1 - \hat{a}_2^2 = 0.014$$

$$\hat{\psi}_{33} = 1 - \hat{a}_3^2 = 0.610 .$$

Como $\hat{h}_1^2 \approx 0.965$ y $\hat{h}_2^2 \approx 0.986$, las las dos comunalidades están cercanas a 1, se concluye que X_1 y X_2 sí están siendo explicados por el factor considerado.

METODO DE FACTORES PRINCIPALES

Su objetivo es descomponer la matriz de correlaciones $\hat{\mathcal{R}}$ (o también $\hat{\Sigma}$ si así se requiere). De la relación

(3) podemos escribir

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

La idea a continuación, es tratar de descomponer[1] $\hat{\mathcal{R}} - \hat{\Psi}_Y$ (previamente substituyendo aquí un estimador preliminar para Ψ_Y).

Para plantear un estimador preliminar para Ψ_Y, proponemos estimadores preliminares de las comunalidades h_j^2, $j=1,2,\dots,p$, hay 2 propuestas usadas en la práctica:

1. $\hat{h}_j^2$ es el cuadrado del coeficiente de correlacion multiple de X_j con $(X_1, \dots, X_{j-1}, X_{j+1}, \dots, X_p)$

2. $\hat{h}_j^2$ esta dado por $\max\left\{ |r_{x_j x_\ell}| : \begin{array}{l} \ell \neq j \\ \ell \in \{1,2,\dots,p\} \end{array} \right\}$

(1) Por ej. descomposición de Jordan ($\hat{\mathcal{R}} - \hat{\Psi}_Y$ es simétrica)

$$\hat{\mathcal{R}} - \hat{\Psi}_Y = \Gamma \Lambda \Gamma' = \sum_{j=1}^{p} \lambda_j \gamma_j \gamma_j' \rightarrow \hat{Q}_Y = \Gamma \Lambda^{1/2}$$

$r_{x_i x_j}$ = entrada i,j de la matriz $\hat{\mathcal{R}}$.

Construimos $\hat{\Psi}_Y$ dada por $\tilde{\Psi}_{jj} = 1-\hat{h}_j^2$ $j=1,2,\dots,p$, y con esta matriz diagonal a su vez se construye la "matriz de correlaciones reducida" $\mathcal{R}-\Psi_Y$.

Aplicamos la descomposición de Jordan para la matriz simétrica $\hat{\mathcal{R}}-\hat{\Psi}_Y$

$$\hat{\mathcal{R}}-\hat{\Psi}_Y = \sum_{j=1}^{p} \lambda_j \hat{\gamma}_j \hat{\gamma}_j',$$

$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ son los valores propios de $\hat{\mathcal{R}}-\hat{\Psi}_Y$ y $\hat{\gamma}_1, \dots, \hat{\gamma}_p$ son sus vectores propios. Asumiendo que los primeros k valores propios $\lambda_1, \dots, \lambda_k$ son positivos y "grandes" comparados con $\lambda_{k+1}, \dots, \lambda_p$, podemos estimar

$$\hat{q}_l = \sqrt{\lambda_l}\, \hat{\gamma}_l \quad ; \quad l=1,2,\dots k,$$

es decir $\hat{Q}_Y = \Gamma_1 \Lambda_1^{1/2}$ (dims $p\times k$)

$\Gamma_1 = (\hat{\gamma}_1, \ldots, \hat{\gamma}_K)$; $\Lambda_1 = \mathrm{diag}(\lambda_1, \ldots, \lambda_K)$.

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

$$\hat{\psi}_{jj} = 1 - \sum_{l=1}^{K} \hat{q}_{jl}^2 \quad ; \quad j = 1, 2, \ldots p.$$