Scores de Bartlett

Imanol Núñez Morales

24 de septiembre de 2015

Teorema (Gauss-Markov)

Sea $\mathrm{Y} = \mathrm{X}\beta + \varepsilon$ *con* $\mathrm{Y}, \varepsilon \in \mathbb{R}^n$, $\beta \in \mathbb{R}^p$ *y* $\mathrm{X} \in \mathbb{R}^{n \times p}$, *en donde* $E[\varepsilon] = 0$ *y* $Var(\varepsilon) = \sigma^2 \mathcal{I}_n$ *con* $\sigma^2 < \infty$. *Entonces el estimador* $\hat{\beta} = \left(\mathrm{X}^T \mathrm{X}\right)^{-1} \mathrm{X}^T \mathrm{Y}$, *de* β, *de mínimos cuadrados es el mejor estimador lineal insesgado en el sentido del error cuadrático medio.*

En análisis factorial tenemos la ecuación

$$\underbrace{\mathrm{X}^{\text{🐱}}}_{k\times 1} = \underbrace{\mathrm{A}}_{k\times m}\underbrace{\mathrm{F}}_{m\times 1} + \underbrace{e}_{k\times 1}, \tag{1}$$

en donde

i) $\mathrm{X}^{\text{🐱}} = \mathcal{D}^{-\frac{1}{2}}\left(\mathrm{X}-\mu\right)$ con $\mathcal{D} = diag\left(\Sigma\right)$,

ii) $\mathrm{F} \sim (0, \mathcal{I}_m)$,

iii) $e \sim (0, \Psi)$ con Ψ una matriz diagonal,

iv) $Cov(\mathrm{F}, e) = 0_{m\times k}$,

v) A se define como la matriz de pesos de los factores.

Para obtener los scores de Bartlett se asume que se conocen $\mathrm{X}^{😺}$ y A en la ecuación (1). Lo que queremos estimar es el valor de F y esto es un caso claro de regresión lineal.

Inmediatamente pensamos en utilizar mínimos cuadrados ordinarios para estimar F, pero como no sabemos si $\Psi = \sigma^2 \mathcal{I}_k$ para alguna $\sigma^2 > 0$, no podemos asegurar que esta estimación sea la mejor.

¿Cómo resolver este problema?

Si premultiplicamos ambos lados de la ecuación (1) por $\Psi^{-\frac{1}{2}}$ obtenemos lo siguiente:

$$\Psi^{-\frac{1}{2}}\mathrm{X}^{\text{😸}} = \Psi^{-\frac{1}{2}}\mathrm{AF} + \Psi^{-\frac{1}{2}}e. \tag{2}$$

Si nombramos $\epsilon = \Psi^{-\frac{1}{2}}e$ es inmediato que $E\left[\epsilon\right] = 0$ y $Var\left(\epsilon\right) = \mathcal{I}_k$, además de que sigue siendo un modelo lineal, por lo que el mejor estimador lineal de F es el que se obtiene de minimizar lo siguiente:

$$\sum_{i=1}^{k} \epsilon_i^2 = \epsilon^T \epsilon \tag{3}$$

Resolvamos el problema.

El problema de minimizar (3) se escribe como sigue:

$$\min \sum_{i=1}^{k} \epsilon_i^2. \tag{4}$$

Veamos que

$$\begin{aligned}
\sum_{i=1}^{k} \epsilon_i^2 &= \epsilon^T \epsilon \\
&= \left(\Psi^{-\frac{1}{2}}\mathrm{X}^{\text{🐱}} - \Psi^{-\frac{1}{2}}\mathrm{AF}\right)^T \left(\Psi^{-\frac{1}{2}}\mathrm{X}^{\text{🐱}} - \Psi^{-\frac{1}{2}}\mathrm{AF}\right) \\
&= \left(\mathrm{X}^{\text{🐱}} - \mathrm{AF}\right)^T \Psi^{-1} \left(\mathrm{X}^{\text{🐱}} - \mathrm{AF}\right) \\
&= \left(\mathrm{X}^{\text{🐱}}\right)^T \Psi^{-1}\mathrm{X}^{\text{🐱}} - 2\left(\left(\mathrm{X}^{\text{🐱}}\right)^T \Psi^{-1}\mathrm{A}\right)\mathrm{F} \\
&\quad + \mathrm{F}^T \left(\mathrm{A}^T \Psi^{-1}\mathrm{A}\right)\mathrm{F}.
\end{aligned}$$

De las operaciones anteriores se deduce que el problema (4) es equivalente a minimizar la función $g(F)$ dada por

$$\left(\mathrm{X}^{\Cat}\right)^{T}\Psi^{-1}\mathrm{X}^{\Cat} - 2\left(\left(\mathrm{X}^{\Cat}\right)^{T}\Psi^{-1}\mathrm{A}\right)\mathrm{F} + \mathrm{F}^{T}\left(\mathrm{A}^{T}\Psi^{-1}\mathrm{A}\right)\mathrm{F}. \tag{5}$$

Para esto necesitamos que $\nabla g(\mathrm{F}) = 0$. Es fácil demostrar que

$$\nabla g(\mathrm{F}) = 2\left(\mathrm{A}^{T}\Psi^{-1}\mathrm{A}\right)\mathrm{F} - 2\mathrm{A}^{T}\Psi^{-1}\mathrm{X}^{\Cat}, \tag{6}$$

por lo que $\nabla g(\hat{\mathrm{F}}) = 0$ si y sólo si

$$\hat{\mathrm{F}} = \left(\mathrm{A}^{T}\Psi^{-1}\mathrm{A}\right)^{-1}\mathrm{A}^{T}\Psi^{-1}\mathrm{X}^{\Cat}. \tag{7}$$

El resultado (7) es lo que se conoce como los scores de Bartlett.

El código para obtener los scores de Bartlett en R se encuentra en: `https://github.com/inuzm/Aplicada3/blob/master/Bartlett/recuperar_factores.R`.
La idea es aprovechar la teoría desarrollada hasta el momento.

[1] Koch, Inge. *Analysis of Multivariate and High-Dimensional Data*, Cambridge University Press, 2013.

[2] https://en.wikipedia.org/wiki/Gauss%E2%80%93Markov_theorem

[3] Documentación de *R*.