MAT-269: Sesión 13
Regresión Multivariada I

Felipe Osorio

fosorios.mat.utfsm.cl

Departamento de Matemática, UTFSM

En esta sección extendemos el modelo de regresión lineal considerando que ahora se dispone de k variables de respuesta. El modelo lineal multivariado asume la forma

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{B} + \boldsymbol{U},$$

donde $\boldsymbol{Y}$ y $\boldsymbol{U}$ son matrices aleatorias $n \times k$, $\boldsymbol{X}$ es matriz de diseño $n \times p$ y $\boldsymbol{B}$ es matriz de coeficientes de regresión $p \times k$. Asumiremos que $\operatorname{rg}(\boldsymbol{X}) = p$ y $n \geq p + k$.

Además supondremos que las filas de la matriz de disturbios aleatorios son independientes $\mathsf{N}_p(\mathbf{0}, \boldsymbol{\Sigma})$. Es decir,

$$\boldsymbol{U} \sim \mathsf{N}_{n,k}(\mathbf{0}, \boldsymbol{I}_n \otimes \boldsymbol{\Sigma}),$$

o análogamente

$$\boldsymbol{Y} \sim \mathsf{N}_{n,k}(\boldsymbol{X}\boldsymbol{B}, \boldsymbol{I}_n \otimes \boldsymbol{\Sigma}).$$

Resultado 1

Si $\boldsymbol{Y} \sim \mathsf{N}_{n,k}(\boldsymbol{X}\boldsymbol{B}, \boldsymbol{I}_n \otimes \boldsymbol{\Sigma})$ y $n \geq k + p$ los estimadores máximo verosímiles de $\boldsymbol{B}$ y $\boldsymbol{\Sigma}$ son

$$\begin{aligned} \widehat{\boldsymbol{B}} &= (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{Y}, \\ \widehat{\boldsymbol{\Sigma}} &= \frac{1}{n}(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{B}})^\top (\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{B}}) \end{aligned}$$

Además $(\widehat{\boldsymbol{B}}, \widehat{\boldsymbol{\Sigma}})$ es suficiente para $(\boldsymbol{B}, \boldsymbol{\Sigma})$.

Demostración:

En efecto, como $\boldsymbol{Y} \sim \mathsf{N}_{n,k}(\boldsymbol{X}\boldsymbol{B}, \boldsymbol{I}_n \otimes \boldsymbol{\Sigma})$, la función de densidad conjunta de $\boldsymbol{Y}$ es

$$f(\boldsymbol{Y}) = (2\pi)^{-np/2}|\boldsymbol{\Sigma}|^{-n/2}\exp\big\{-\tfrac{1}{2}\operatorname{tr}(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{B})\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{B})^\top\big\},$$

ignorando términos que no involucran $\boldsymbol{\theta} = (\boldsymbol{B}, \boldsymbol{\Sigma})$, tenemos

$$\ell_n(\boldsymbol{B}, \boldsymbol{\Sigma}) = -\frac{n}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\operatorname{tr}\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{B})^\top(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{B}).$$

Diferenciando con relación a $\boldsymbol{B}$ y $\boldsymbol{\Sigma}$ obtenemos

$$\begin{aligned}
\mathsf{d}_B\,\ell_n(\boldsymbol{B}, \boldsymbol{\Sigma}) &= -\tfrac{1}{2}\operatorname{tr}\boldsymbol{\Sigma}^{-1}\,\mathsf{d}_B\,Q(\boldsymbol{B}) \\
\mathsf{d}_\Sigma\,\ell_n(\boldsymbol{B}, \boldsymbol{\Sigma}) &= -\tfrac{n}{2}\operatorname{tr}\boldsymbol{\Sigma}^{-1}\,\mathsf{d}\,\boldsymbol{\Sigma} + \tfrac{1}{2}\operatorname{tr}\boldsymbol{\Sigma}^{-1}(\mathsf{d}\,\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}Q(\boldsymbol{B})
\end{aligned}$$

donde $Q(\boldsymbol{B}) = (\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{B})^\top(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{B})$.

En efecto,

$$\mathsf{d}_B\, Q(\boldsymbol{B}) = -(\mathsf{d}\,\boldsymbol{B})^\top \boldsymbol{X}^\top(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B}) - (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B})^\top \boldsymbol{X}\,\mathsf{d}\,\boldsymbol{B},$$

de este modo

$$\mathsf{d}_B\, \ell_n(\boldsymbol{B}, \boldsymbol{\Sigma}) = \tfrac{1}{2}\operatorname{tr}\boldsymbol{\Sigma}^{-1}\big\{(\mathsf{d}\,\boldsymbol{B})^\top \boldsymbol{X}^\top(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B}) + (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B})^\top \boldsymbol{X}\,\mathsf{d}\,\boldsymbol{B}\big\},$$

recordando que $\operatorname{tr}\boldsymbol{A} = \operatorname{tr}\boldsymbol{A}^\top$ y como $\boldsymbol{\Sigma}$ es simétrica

$$\mathsf{d}_B\, \ell_n(\boldsymbol{B}, \boldsymbol{\Sigma}) = \operatorname{tr}\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B})^\top \boldsymbol{X}\,\mathsf{d}\,\boldsymbol{B},$$

el diferencial es cero si y sólo si

$$\boldsymbol{X}^\top(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B}) = \boldsymbol{0},$$

es decir $\widehat{\boldsymbol{B}}$ es solución del sistema de ecuaciones

$$\boldsymbol{X}^\top\boldsymbol{X}\widehat{\boldsymbol{B}} = \boldsymbol{X}^\top\boldsymbol{Y},$$

como $\operatorname{rg}(\boldsymbol{X}) = p$, tenemos $\widehat{\boldsymbol{B}} = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{Y}$.

Por otro lado,

$$\mathsf{d}_{\Sigma}\, Q(\boldsymbol{B}) = -\tfrac{1}{2}\operatorname{tr}\boldsymbol{\Sigma}^{-1}(n\boldsymbol{\Sigma} - Q(\boldsymbol{B}))\boldsymbol{\Sigma}^{-1}\,\mathsf{d}\,\boldsymbol{\Sigma},$$

y por tanto,

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{n} Q(\widehat{\boldsymbol{B}}).$$

Note además que

$$\begin{aligned}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B})^\top(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B}) &= (\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{B}} + \boldsymbol{X}\widehat{\boldsymbol{B}} - \boldsymbol{X}\boldsymbol{B})^\top(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{B}} + \boldsymbol{X}\widehat{\boldsymbol{B}} - \boldsymbol{X}\boldsymbol{B}) \\ &= Q(\widehat{\boldsymbol{B}}) + (\widehat{\boldsymbol{B}} - \boldsymbol{B})^\top\boldsymbol{X}^\top\boldsymbol{X}(\widehat{\boldsymbol{B}} - \boldsymbol{B})\end{aligned}$$

(pues $\boldsymbol{X}^\top(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{B}}) = \boldsymbol{0}$), de ahí que

$$\begin{aligned}L(\boldsymbol{B}, \boldsymbol{\Sigma}) &= |\boldsymbol{\Sigma}|^{-n/2}\exp\big\{-\tfrac{1}{2}\operatorname{tr}\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B})^\top(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B})\big\} \\ &= |\boldsymbol{\Sigma}|^{-n/2}\exp\big\{-\tfrac{1}{2}\operatorname{tr}\boldsymbol{\Sigma}^{-1}\big(Q(\widehat{\boldsymbol{B}}) + (\widehat{\boldsymbol{B}} - \boldsymbol{B})^\top\boldsymbol{X}^\top\boldsymbol{X}(\widehat{\boldsymbol{B}} - \boldsymbol{B})\big)\big\} \\ &= |\boldsymbol{\Sigma}|^{-n/2}\exp\big\{-\tfrac{n}{2}\operatorname{tr}\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1} + \tfrac{1}{2}\operatorname{tr}\boldsymbol{\Sigma}^{-1}(\widehat{\boldsymbol{B}} - \boldsymbol{B})^\top\boldsymbol{X}^\top\boldsymbol{X}(\widehat{\boldsymbol{B}} - \boldsymbol{B})\big\},\end{aligned}$$

es decir $(\widehat{\boldsymbol{B}}, \widehat{\boldsymbol{\Sigma}})$ es suficiente para $(\boldsymbol{B}, \boldsymbol{\Sigma})$.

Resultado 2

Si $\boldsymbol{Y} \sim \mathsf{N}_{n,k}(\boldsymbol{X}\boldsymbol{B}, \boldsymbol{I}_n \otimes \boldsymbol{\Sigma})$ los estimadores máximo verosímiles $\widehat{\boldsymbol{B}}$ y $\widehat{\boldsymbol{\Sigma}}$ son independientemente distribuídos como:

$$\begin{aligned} \widehat{\boldsymbol{B}} &\sim \mathsf{N}_{q,k}(\boldsymbol{B}, (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \otimes \boldsymbol{\Sigma}), \\ n\widehat{\boldsymbol{\Sigma}} &\sim \mathsf{W}_k(n-p, \boldsymbol{\Sigma}). \end{aligned}$$

Demostración:

Sea $\boldsymbol{E} = \boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{B}} = \boldsymbol{Y} - \boldsymbol{H}\boldsymbol{Y} = (\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y}$, con

$$\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top.$$

Sea $\boldsymbol{M} = \boldsymbol{I} - \boldsymbol{H}$, entonces

$$\boldsymbol{M}\boldsymbol{X} = (\boldsymbol{I} - \boldsymbol{H})\boldsymbol{X} = \boldsymbol{0}, \qquad \boldsymbol{M}^2 = \boldsymbol{M}.$$

De este modo,

$$\boldsymbol{E}^\top \boldsymbol{E} = \boldsymbol{Y}^\top (\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y} = \boldsymbol{Y}^\top \boldsymbol{M}\boldsymbol{Y} = n\widehat{\boldsymbol{\Sigma}}.$$

Considere la matriz

$$\begin{pmatrix} \widehat{\boldsymbol{B}} \\ \boldsymbol{E} \end{pmatrix} = \begin{pmatrix} (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \\ \boldsymbol{M} \end{pmatrix} \boldsymbol{Y}$$

luego $(\widehat{\boldsymbol{B}}^\top, \boldsymbol{E}^\top)^\top$ es normal con media

$$\mathsf{E}\begin{pmatrix} \widehat{\boldsymbol{B}} \\ \boldsymbol{E} \end{pmatrix} = \begin{pmatrix} (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \\ \boldsymbol{M} \end{pmatrix} \mathsf{E}(\boldsymbol{Y}) = \begin{pmatrix} \widehat{\boldsymbol{B}} \\ \boldsymbol{E} \end{pmatrix} = \begin{pmatrix} (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \\ \boldsymbol{M} \end{pmatrix} \boldsymbol{X}\boldsymbol{B} = \begin{pmatrix} \boldsymbol{B} \\ \boldsymbol{0} \end{pmatrix}$$

mientras que la covarianza es:

$$\begin{aligned}
&\left(\begin{pmatrix} (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \\ \boldsymbol{M} \end{pmatrix} \otimes \boldsymbol{I}_p\right) (\boldsymbol{I}_n \otimes \boldsymbol{\Sigma})\Big((\boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}, \boldsymbol{M}^\top) \otimes \boldsymbol{I}_p\Big) \\
&= \begin{pmatrix} (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \\ \boldsymbol{M} \end{pmatrix} (\boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}, \boldsymbol{M}^\top) \otimes \boldsymbol{\Sigma} \\
&= \begin{pmatrix} (\boldsymbol{X}^\top \boldsymbol{X})^{-1} & (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{M}^\top \\ \boldsymbol{M}\boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1} & \boldsymbol{M}\boldsymbol{M}^\top \end{pmatrix} \otimes \boldsymbol{\Sigma}
\end{aligned}$$

En efecto,

$$\mathsf{Cov}\left(\operatorname{vec}\begin{pmatrix}\widehat{\boldsymbol{B}}\\ \boldsymbol{E}\end{pmatrix}\right) = \begin{pmatrix}(\boldsymbol{X}^\top\boldsymbol{X})^{-1} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{M}\end{pmatrix} \otimes \boldsymbol{\Sigma}.$$

De ahí que

$$\widehat{\boldsymbol{B}} \sim \mathsf{N}(\boldsymbol{B}, (\boldsymbol{X}^\top\boldsymbol{X})^{-1} \otimes \boldsymbol{\Sigma}), \qquad \boldsymbol{E} \sim \mathsf{N}(\boldsymbol{0}, \boldsymbol{M} \otimes \boldsymbol{\Sigma}),$$

luego $\widehat{\boldsymbol{B}}$ y $\boldsymbol{E}$ son independientes. La parte final del resultado sigue de notar que $n\widehat{\boldsymbol{\Sigma}}$ puede ser escrito como $\boldsymbol{E}^\top\boldsymbol{E}$.

EX UMBRA IN SOLEM

Ejemplo: Datos de dializadores (Vonesh y Carter, 1987)[1]

Datos de un estudio para evaluar las características de ultrafiltración in vivo de un grupo de dializadores.

Los dializadores se evaluaron en tres centros y cada uno de ellos utilizó un tipo diferente de sistema de administración de dializado.

Los datos corresponden a la tasa de ultrafiltración de los 4 dializadores Y_1, Y_2, Y_3, Y_4.

Este conjunto de datos ha sido usado en varios artículos científicos y bajo el supuesto de normalidad. Además, este supuesto no es rechazado por el test de Mardia (1974), basado en las medidas de sesgo y kurtosis multivariadas.

[1] Biometrics **43**, 617-628.

Ejemplo: Datos de dializadores

```
# Carga biblioteca 'heavy' y datos de ejemplo
> library(heavy)
> data(dialyzer)

> dialyzer
    y1   y2   y3   y4 centre
1  600 1026 1470 1890      1
2  516  930 1380 1770      1
3  480  900 1380 1860      1
4  528  930 1410 1872      1
5  540  978 1410 1920      1
6  564  996 1422 1920      1
7  564 1062 1500 1980      1
8  492  900 1392 1860      1
9  516  960 1380 1800      1
10 528  930 1356 1860      1
11 564 1020 1380 1884      1

...

38 480  780 1140 1710      3
39 540  840 1200 1650      3
40 780  780 1290 1680      3
```

Vamos a considerar un modelo de regresión multivariada con matriz de respuestas

$$\boldsymbol{Y} = \big(\boldsymbol{Y}_1, \boldsymbol{Y}_2, \boldsymbol{Y}_3, \boldsymbol{Y}_4\big),$$

y matriz de diseño

$$\boldsymbol{X} = \begin{pmatrix} \mathbf{1}_{17} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_{12} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1}_{11} \end{pmatrix}.$$

Por tanto tenemos que $\boldsymbol{B} \in \mathbb{R}^{3\times 4}$ y $\boldsymbol{\Sigma}$ es matriz simétrica y definida positiva 4×4.

Ejemplo: Datos de dializadores

```
# Ajuste de un modelo de regresión lineal multivariado
# (bajo errores normales)
> fm <- heavyLm(cbind(y1,y2,y3,y4) ~ -1 + centre, data = dialyzer,
+      family = normal())

# Salida con estimación de los coeficientes de regresión
> fm
Call:
heavyLm(formula = cbind(y1, y2, y3, y4) ~ -1 + centre, data = dialyzer,
      family = normal())
Converged in 1 iterations

Coefficients:
         y1         y2         y3          y4
centre1  541.0588   973.4118  1404.3529  1873.4118
centre2  472.5000   830.5000  1230.5000  1653.0000
centre3  591.8182   850.9091  1276.3636  1655.4545

Degrees of freedom: 40 total; 37 residual
```

Ejemplo: Datos de dializadores

```
# Salida de función 'summary'
> summary(fm)
Multivariate regression under heavy-tailed distributions
 Data: dialyzer; Family: normal()

Coefficients:
         y1        y2        y3         y4
centre1  541.0588  973.4118  1404.3529  1873.4118
centre2  472.5000  830.5000  1230.5000  1653.0000
centre3  591.8182  850.9091  1276.3636  1655.4545

Scatter matrix estimate:
   y1        y2        y3        y4
y1 4402.5848
y2  380.6878 2337.0659
y3  867.8210 2712.4908 6160.8440
y4  506.2871 1987.2839 4808.4436 7134.5452

Degrees of freedom: 40 total; 37 residual
Log-likelihood: -761.7318 on 22 degrees of freedom
```

Es decir, tenemos que

$$\widehat{\boldsymbol{B}} = \begin{pmatrix} 541.059 & 973.412 & 1404.353 & 1873.412 \\ 472.500 & 830.500 & 1230.500 & 1653.000 \\ 591.818 & 850.909 & 1276.364 & 1655.455 \end{pmatrix}.$$

y

$$\widehat{\boldsymbol{\Sigma}} = \begin{pmatrix} 4402.585 & 380.688 & 867.821 & 506.287 \\ 380.688 & 2337.066 & 2712.491 & 1987.284 \\ 867.821 & 2712.491 & 6160.844 & 4808.444 \\ 506.287 & 1987.284 & 4808.444 & 7134.545 \end{pmatrix}.$$

Además, $\ell_n(\widehat{\boldsymbol{B}}, \widehat{\boldsymbol{\Sigma}}) = -761.732$.

Ejemplo: Datos de dializadores

```
# Alternativamente, podemos usar la función 'lm' de R
> f0 <- lm(cbind(y1,y2,y3,y4) ~ -1 + centre, data = dialyzer)

# Salida:
> f0

Call:
lm(formula = cbind(y1, y2, y3, y4) ~ -1 + centre, data = dialyzer)

Coefficients:
          y1      y2      y3      y4
centre1   541.1   973.4   1404.4  1873.4
centre2   472.5   830.5   1230.5  1653.0
centre3   591.8   850.9   1276.4  1655.5
```