# MAT-468: Sesión 3, Cálculos en regresión

**Felipe Osorio**

http://fosorios.mat.utfsm.cl

Departamento de Matemática, UTFSM

Antes de describir una variante para obtener la descomposición ortogonal-triangular (QR) de una matriz es conveniente revisar algunas propiedades fundamentales de las matrices ortogonales:

- $\boldsymbol{Q}\boldsymbol{Q}^\top = \boldsymbol{Q}^\top\boldsymbol{Q} = \boldsymbol{I}$.
- $\langle \boldsymbol{Q}\boldsymbol{x}, \boldsymbol{Q}\boldsymbol{y}\rangle = \boldsymbol{x}^\top\boldsymbol{Q}^\top\boldsymbol{Q}\boldsymbol{y} = \boldsymbol{x}^\top\boldsymbol{y} = \langle \boldsymbol{x}, \boldsymbol{y}\rangle$.
- $\|\boldsymbol{Q}\boldsymbol{x}\| = \|\boldsymbol{x}\|$.
- Si $\boldsymbol{B} = \boldsymbol{Q}^\top\boldsymbol{A}\boldsymbol{Q}$, entonces $\boldsymbol{A}$ y $\boldsymbol{B}$ tienen los mismos valores propios para $\boldsymbol{Q}$ matriz ortogonal.

Existen diversar variantes del algoritmo para implementar la descomposición QR. A continuación veremos una basada en transformaciones Householder.

### Problema 1

Para $\boldsymbol{x} \in \mathbb{R}^p$, $\boldsymbol{x} \neq \boldsymbol{0}$, hallar una matriz ortogonal $\boldsymbol{M} \in \mathbb{R}^{p \times p}$ tal que

$$\boldsymbol{M}^\top \boldsymbol{x} = \|\boldsymbol{x}\| \, \boldsymbol{e}_1,$$

donde $\boldsymbol{e}_1 = (1, 0, \dots, 0)^\top$ denota el primer vector unidad.

### Definición 1 (Reflexión)

Sea $\boldsymbol{u}$ y $\boldsymbol{v}$ vectores ortonormales y $\boldsymbol{x}$ vector generado por $\boldsymbol{u}$ y $\boldsymbol{v}$. Entonces

$$\boldsymbol{x} = c_1 \boldsymbol{u} + c_2 \boldsymbol{v},$$

para escalares $c_1$, $c_2$. El vector

$$\widetilde{\boldsymbol{x}} = -c_1 \boldsymbol{u} + c_2 \boldsymbol{v},$$

el llamado una reflexión de $\boldsymbol{x}$ a través de la línea definida por el vector $\boldsymbol{v}$ (o $\boldsymbol{u}^\perp$).

**Definición 2 (Transformación Householder)**

Sea $\boldsymbol{x} = c_1\boldsymbol{u} + c_2\boldsymbol{v}$, con $\boldsymbol{u}$ y $\boldsymbol{v}$ vectores generadores de $\boldsymbol{x}$ y considere la matriz

$$\boldsymbol{H} = \boldsymbol{I} - \lambda\boldsymbol{u}\boldsymbol{u}^\top, \qquad \lambda = 2/\boldsymbol{u}^\top\boldsymbol{u}.$$

Note que $\boldsymbol{H}\boldsymbol{x} = \widetilde{\boldsymbol{x}}$, es decir $\boldsymbol{H}$ es un reflector.

La transformación Householder satisface las siguientes propiedades:

- $\boldsymbol{H}\boldsymbol{u} = -\boldsymbol{u}$.
- $\boldsymbol{H}\boldsymbol{v} = \boldsymbol{v}$ para cualquier $\boldsymbol{v}$ ortogonal a $\boldsymbol{u}$.
- $\boldsymbol{H}^\top = \boldsymbol{H}$.
- $\boldsymbol{H}^{-1} = \boldsymbol{H}^\top$.

Considere $\boldsymbol{u} = \boldsymbol{x} + \delta \boldsymbol{e}_1$, donde $\delta^2 = \boldsymbol{x}^\top \boldsymbol{x}$. Es fácil notar que,

$$\begin{aligned}\boldsymbol{u}^\top \boldsymbol{u} &= (\boldsymbol{x} + \delta \boldsymbol{e}_1)^\top (\boldsymbol{x} + \delta \boldsymbol{e}_1) = \boldsymbol{x}^\top \boldsymbol{x} + 2\delta \boldsymbol{x}^\top \boldsymbol{e}_1 + \delta^2 \boldsymbol{e}_1^\top \boldsymbol{e}_1 \\ &= \delta^2 + 2\delta x_1 + \delta^2 = 2(\delta^2 + \delta x_1),\end{aligned}$$

así

$$\lambda = \frac{2}{\|\boldsymbol{u}\|^2} = \frac{1}{\delta^2 + \delta x_1}.$$

Además,

$$\boldsymbol{u}^\top \boldsymbol{x} = (\boldsymbol{x} + \delta \boldsymbol{e}_1)^\top \boldsymbol{x} = \boldsymbol{x}^\top \boldsymbol{x} + \delta \boldsymbol{e}_1^\top \boldsymbol{x} = \delta^2 + \delta x_1.$$

De este modo, $\boldsymbol{H}$ satisface:[1]

$$\begin{aligned}\boldsymbol{H}\boldsymbol{x} &= (\boldsymbol{I} - \lambda \boldsymbol{u}\boldsymbol{u}^\top)\boldsymbol{x} = \boldsymbol{x} - \lambda(\boldsymbol{u}^\top \boldsymbol{x})\boldsymbol{u} \\ &= \boldsymbol{x} - \lambda(\delta^2 + \delta x_1)(\boldsymbol{x} + \delta \boldsymbol{e}_1) = \boldsymbol{x} - \boldsymbol{x} - \delta \boldsymbol{e}_1 \\ &= -\delta \boldsymbol{e}_1.\end{aligned}$$

[1] Note que $\boldsymbol{H}\boldsymbol{x}$ puede ser obtenido usando un axpy.

La descomposición QR de una matriz $\boldsymbol{A} \in \mathbb{R}^{n\times p}$ $(n > p)$, puede ser construída a través de una secuencia de matrices $\boldsymbol{Q}_1, \dots, \boldsymbol{Q}_p$ tales que

$$\boldsymbol{Q}_p \cdots \boldsymbol{Q}_1 \boldsymbol{A} = \begin{pmatrix} \boldsymbol{R} \\ \boldsymbol{0} \end{pmatrix},$$

donde todas $\boldsymbol{Q}_1, \dots, \boldsymbol{Q}_p$ son ortogonales. De este modo,

$$\boldsymbol{A} = \boldsymbol{Q}_1^\top \cdots \boldsymbol{Q}_p^\top \begin{pmatrix} \boldsymbol{R} \\ \boldsymbol{0} \end{pmatrix} = \boldsymbol{Q} \begin{pmatrix} \boldsymbol{R} \\ \boldsymbol{0} \end{pmatrix}.$$

A continuación se describe el algoritmo para obtener la descomposición QR usando transformaciones Householder.[2] Sea $\boldsymbol{M}(\boldsymbol{x})$ la matriz ortogonal desde el Problema 1 basada en un vector $\boldsymbol{x}$.

[2] Otro método popular para obtener la descomposición QR es usando rotaciones Givens.

# Descomposición QR

**Algoritmo 1:** Descomposición QR

**Entrada:** Matriz $\boldsymbol{A} \in \mathbb{R}^{n\times p}$.

**Salida :** Factores $\boldsymbol{Q}$ y $\boldsymbol{R}$, matrices ortogonal y triangular superior, respectivamente.

**begin**

Hacer $\boldsymbol{Q} = \boldsymbol{I}_n$ y $\boldsymbol{R} = \boldsymbol{A}$

**for** $i = 1$ **to** $p$ **do**

$\boldsymbol{x} = (R_{1i}, \dots, R_{pi})^\top$

$$\boldsymbol{Q}_i = \begin{pmatrix} \boldsymbol{I}_{i-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{M}(\boldsymbol{x}) \end{pmatrix}$$

/* $\boldsymbol{M}(\boldsymbol{x})$ `obtenido usando reflexiones Householder` */

$\boldsymbol{Q} = \boldsymbol{Q}_i \boldsymbol{Q}$

$\boldsymbol{R} = \boldsymbol{Q}_i \boldsymbol{R}$

**end**

$\boldsymbol{Q} = \boldsymbol{Q}^\top$

$\boldsymbol{R} = (R_{ij})$ para $i, j = 1, \dots, p$.

**end**

Sea el modelo de regresión lineal:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

donde $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ con $\operatorname{rg}(\boldsymbol{X}) = p$ y $\mathsf{E}(\boldsymbol{\epsilon}) = \mathbf{0}$ y $\mathsf{Cov}(\boldsymbol{\epsilon}) = \sigma^2 \boldsymbol{I}$. El estimador mínimos cuadrados (LS) de $\boldsymbol{\beta}$ es:

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{Y}, \qquad \text{con} \qquad \mathsf{Cov}(\widehat{\boldsymbol{\beta}}) = \sigma^2(\boldsymbol{X}^\top\boldsymbol{X})^{-1}.$$

Además,

$$s^2 = \frac{1}{n-p}\|\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\|^2.$$

Adicionalmente, si $\boldsymbol{\epsilon} \sim \mathsf{N}_n(\mathbf{0}, \sigma^2\boldsymbol{I})$, entonces[3]

$$\widehat{\boldsymbol{\beta}} \sim \mathsf{N}_p(\boldsymbol{\beta}, \sigma^2(\boldsymbol{X}^\top\boldsymbol{X})^{-1}),$$
$$\frac{(n-p)s^2}{\sigma^2} \sim \chi^2(n-p).$$

[3] En cuyo caso, $\widehat{\boldsymbol{\beta}}$ corresponde al estimador ML de $\boldsymbol{\beta}$.

## Aplicación en regresión lineal

Las ecuaciones de estimación para obtener $\widehat{\boldsymbol{\beta}}$ son $\boldsymbol{X}^\top(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}) = \boldsymbol{0}$, o bien

$$\boldsymbol{X}^\top\boldsymbol{X}\widehat{\boldsymbol{\beta}} = \boldsymbol{X}^\top\boldsymbol{Y}. \tag{1}$$

Así, podemos resolver (1) usando la descomposición Cholesky, de

$$\boldsymbol{X}^\top\boldsymbol{X} = \boldsymbol{U}^\top\boldsymbol{U},$$

con $\boldsymbol{U}$ matrix triangular superior. De este modo:

$$\boldsymbol{U}^\top\boldsymbol{z} = \boldsymbol{X}^\top\boldsymbol{Y}, \qquad \text{y} \qquad \boldsymbol{U}\widehat{\boldsymbol{\beta}} = \boldsymbol{z},$$

para obtener $s^2$ considere

$$RSS = \|\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\|^2 = \boldsymbol{Y}^\top\boldsymbol{Y} - \boldsymbol{z}^\top\boldsymbol{z}.$$

Invirtiendo $\boldsymbol{U}$ (in-place)[4], podemos hacer

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

[4] Haciendo $\boldsymbol{U} \leftarrow \boldsymbol{U}^{-1}$, tenemos $(\boldsymbol{X}^\top\boldsymbol{X})^{-1} = \boldsymbol{U}\boldsymbol{U}^\top$

# Aplicación en regresión lineal

Considere

$$\boldsymbol{Z} = (\boldsymbol{X}, \boldsymbol{Y}) \in \mathbb{R}^{n\times(p+1)},$$

luego

$$\boldsymbol{Z}^\top\boldsymbol{Z} = \begin{pmatrix} \boldsymbol{X}^\top\boldsymbol{X} & \boldsymbol{X}^\top\boldsymbol{Y} \\ \boldsymbol{Y}^\top\boldsymbol{X} & \boldsymbol{Y}^\top\boldsymbol{Y} \end{pmatrix} \in \mathbb{R}^{(p+1)\times(p+1)}.$$

Aplicando el operador Sweep sobre las primeros $p$ elementos diagonales de $\boldsymbol{Z}^\top\boldsymbol{Z}$, obtenemos:

$$\begin{aligned} \boldsymbol{B} &= \prod_{i=1}^{p} \operatorname{Sweep}(i)\boldsymbol{Z}^\top\boldsymbol{Z} \\ &= \begin{pmatrix} (\boldsymbol{X}^\top\boldsymbol{X})^{-1} & (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{Y} \\ -\boldsymbol{Y}^\top\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1} & \boldsymbol{Y}^\top\boldsymbol{Y} - \boldsymbol{Y}^\top\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{Y} \end{pmatrix} \\ &= \begin{pmatrix} (\boldsymbol{X}^\top\boldsymbol{X})^{-1} & \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{\beta}}^\top & RSS \end{pmatrix}. \end{aligned}$$

Considere la descomposición QR (Ortogonal-Triangular) de $\boldsymbol{X}$, como:

$$\boldsymbol{X} = \boldsymbol{Q}\boldsymbol{R}, \qquad \boldsymbol{R} = \begin{pmatrix} \boldsymbol{R}_1 \\ \boldsymbol{0} \end{pmatrix}$$

con $\boldsymbol{R}_1 \in \mathbb{R}^{p \times p}$ matriz triangular superior $(n > p)$. Si $\operatorname{rg}(\boldsymbol{X}) = p$, entonces $\boldsymbol{R}_1$ es no singular. Además, considere la transformación:

$$\boldsymbol{Q}^\top \boldsymbol{Y} = \boldsymbol{c}, \qquad \boldsymbol{c} = (\boldsymbol{c}_1^\top, \boldsymbol{c}_2^\top)^\top.$$

El estimador LS minimiza la función objetivo:

$$\begin{aligned} \|\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}\|^2 = \|\boldsymbol{Q}^\top(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})\|^2 &= \|\boldsymbol{Q}^\top\boldsymbol{Y} - \boldsymbol{Q}^\top\boldsymbol{Q}\boldsymbol{R}\boldsymbol{\beta}\|^2 \\ &= \|\boldsymbol{c} - \boldsymbol{R}\boldsymbol{\beta}\|^2, \end{aligned}$$

Es fácil notar que:

$$\|\boldsymbol{c} - \boldsymbol{R}\boldsymbol{\beta}\|^2 = \|\boldsymbol{c}_1 - \boldsymbol{R}_1\boldsymbol{\beta}\|^2 + \|\boldsymbol{c}_2\|^2.$$

Finalmente, el estimador de mínimos cuadrados $\widehat{\boldsymbol{\beta}}$ está dado por la solución del sistema triangular:

$$\boldsymbol{R}_1\widehat{\boldsymbol{\beta}} = \boldsymbol{c}_1.$$

El mínimo de la función objetivo está dado por $\|\boldsymbol{c}_2\|^2$. Note además que

$$\widehat{\sigma}^2 = \frac{1}{n}\,\|\boldsymbol{c}_2\|^2,$$

corresponde al estimador máximo verosímil para $\sigma^2$.[5] Por otro lado,

$$\boldsymbol{R}_1^{-1}\boldsymbol{R}_1^{-\top} = (\boldsymbol{X}^\top\boldsymbol{X})^{-1},$$

permite obtener $\text{Cov}(\widehat{\boldsymbol{\beta}})$.

[5] bajo el supuesto de normalidad.

## Aplicación en regresión lineal

Considere la descomposición valor singular (SVD) de $\boldsymbol{X}$,

$$\boldsymbol{X} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^\top,$$

donde $\boldsymbol{U} \in \mathbb{R}^{n\times p}$ tal que $\boldsymbol{U}^\top\boldsymbol{U} = \boldsymbol{I}_p$, $\boldsymbol{D} = \operatorname{diag}(d_1, \dots, d_p)$ y $\boldsymbol{V}$ es matriz ortogonal $p \times p$. De este modo, podemos escribir el modelo:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{\alpha} + \boldsymbol{\epsilon},$$

con $\boldsymbol{\alpha} = \boldsymbol{V}^\top\boldsymbol{\beta}$. Haciendo $\boldsymbol{Z} = \boldsymbol{U}^\top\boldsymbol{Y}$, tenemos el modelo en forma canónica:

$$\boldsymbol{Z} = \boldsymbol{D}\boldsymbol{\alpha} + \boldsymbol{\delta}, \qquad \boldsymbol{\delta} = \boldsymbol{U}^\top\boldsymbol{\epsilon},$$

donde $\mathsf{E}(\boldsymbol{\delta}) = \boldsymbol{0}$ y $\mathsf{Cov}(\boldsymbol{\delta}) = \sigma^2\boldsymbol{U}^\top\boldsymbol{U} = \sigma^2\boldsymbol{I}_p$.

# Aplicación en regresión lineal

El estimador LS de $\boldsymbol{\alpha}$ en el modelo canónico es:

$$\widehat{\boldsymbol{\alpha}} = \boldsymbol{D}^{-1}\boldsymbol{Z}, \qquad \Rightarrow \qquad \widehat{\boldsymbol{\beta}} = \boldsymbol{V}\widehat{\boldsymbol{\alpha}}.$$

Además,

$$\|\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\|^2 = \|\boldsymbol{Y} - \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^\top\widehat{\boldsymbol{\beta}}\|^2 = \|\boldsymbol{Z} - \boldsymbol{D}\widehat{\boldsymbol{\alpha}}\|^2.$$

Finalmente,

$$(\boldsymbol{X}^\top\boldsymbol{X})^{-1} = (\boldsymbol{V}\boldsymbol{D}^2\boldsymbol{V}^\top)^{-1} = \boldsymbol{V}\boldsymbol{D}^{-2}\boldsymbol{V}^\top.$$