Regression with matrices

Oliver Snellman
oliver.snellman@helsinki.fi

October 6, 2021

In this document I present a simple problem to be studied with a linear regression. The problem can be expressed by using a system of equations 1 or using matrices 2, which I show to be equivalent. To summarize: Matrix notation provides a compact way of representing systems of equations.

Recall the relevant **rules for matrix multiplication and summation**, with $a, b, c, d, x, y \in \mathbb{R}$ and dimensions displayed below (first number for rows, second for columns)

$$\underbrace{\begin{bmatrix} a & b \\ c & d \end{bmatrix}}_{2\times2} \underbrace{\begin{bmatrix} x \\ y \end{bmatrix}}_{2\times1} = \underbrace{\begin{bmatrix} ax+by \\ cx+dy \end{bmatrix}}_{2\times1} \qquad \underbrace{\begin{bmatrix} a \\ c \end{bmatrix}}_{2\times1} + \underbrace{\begin{bmatrix} x \\ y \end{bmatrix}}_{2\times1} = \underbrace{\begin{bmatrix} a+x \\ c+y \end{bmatrix}}_{2\times1}$$

The research question is: How much does education and IQ influence the salary. To answer this we gather information from John and Sue (two data points), by asking three questions each, summarized in the table below.

Name	Salary	Education	IQ
John	5000	4	110
Sue	7000	8	120

Instead of numbers, let's express the data in variable form

Name	Salary	Education	IQ
John	y_1	x_1	z_1
Sue	y_2	x_2	z_2

where y is the variable you wish to explain, whereas x and z are the variables which do the explaining. The lower indexes and colors of y, x and z refer to the person; 1 for John and 2 for Sue (I'm sorry Simone).

We form a system of two regression equations

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 x_1 + \beta_2 z_1 + \varepsilon_1 \\ y_2 &= \beta_0 + \beta_1 x_2 + \beta_2 z_2 + \varepsilon_2 \end{aligned} \tag{1}$$

The aim is to estimate the unknown parameters β_0, β_1 and β_2 with OLS to find the answer to the research question.

Thse equations can be expressed in matrix form as

$$\underbrace{\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}}_{y} = \underbrace{\begin{bmatrix} 1 & x_1 & z_1 \\ 1 & x_2 & z_2 \end{bmatrix}}_{\mathbf{X}} \underbrace{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}}_{\beta} + \underbrace{\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix}}_{\varepsilon} \tag{2}$$

which can be expressed even more compactly as

$$y = X\beta + \varepsilon \qquad \text{General matrix notation for linear regression}$$

This expressions can represent any linear regression. If we, for example, interview more people, we just add rows to y, X and ε. If we add explanatory variables to x and z, then we add columns to X and rows to β. In the latter case we might want to replace the different alphabets $x_i, z_i, q_i \dots$ with x_{ik} notation, where i stands for the person and k for the explanatory variable.

Why are Equations 1 and 2 equivalent?

Let's carry out the matrix multiplication $X\beta$ on the right hand side of Equation 2

$$\begin{bmatrix} 1 & x_1 & z_1 \\ 1 & x_2 & z_2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 1 * \beta_0 + x_1 * \beta_1 + z_1 * \beta_2 \\ \\ 1 * \beta_0 + x_2 * \beta_1 + z_2 * \beta_2 \end{bmatrix} = \begin{bmatrix} \beta_0 + \beta_1 x_1 + \beta_2 z_1 \\ \\ \beta_0 + \beta_1 x_2 + \beta_2 z_2 \end{bmatrix}$$

Notice, that the end result of $X\beta$ matrix multiplication is just a vector with two elements. Plugging this back into Equation 2 gives

$$\begin{bmatrix} y_1 \\ \\ y_2 \end{bmatrix} = \underbrace{\begin{bmatrix} \beta_0 + \beta_1 x_1 + \beta_2 z_1 \\ \\ \beta_0 + \beta_1 x_2 + \beta_2 z_2 \end{bmatrix}}_{X\beta} + \begin{bmatrix} \varepsilon_1 \\ \\ \varepsilon_2 \end{bmatrix}$$

Using the summation rule to combine the two vectors on the right hand side

$$\begin{bmatrix} y_1 \\ \\ y_2 \end{bmatrix} = \begin{bmatrix} \beta_0 + \beta_1 x_1 + \beta_2 z_1 + \varepsilon_1 \\ \\ \beta_0 + \beta_1 x_2 + \beta_2 z_2 + \varepsilon_2 \end{bmatrix}$$

We are left with 2×1 vectors on both sides of the equation. This means, that the elements of these vectors are equal to each other row by row:

$$y_1 = \beta_0 + \beta_1 x_1 + \beta_2 z_1 + \varepsilon_1$$

$$y_2 = \beta_0 + \beta_1 x_2 + \beta_2 z_2 + \varepsilon_2$$

leading us back to the starting point, Equation 1. $\square$