Matrix Regression

When we extend the regression model to include additional predictors, the equations become increasingly complex. With just two predictors, we have a much larger equation if we wish to compute the regression coefficient, for example.

$$B_1 = \frac{\left[\sum(X_1 - \bar{X}_1)(Y - \bar{Y})\right]\left[\sum(X_1 - \bar{X}_1)^2\right] - \left[\sum(X_2 - \bar{X}_2)(Y - \bar{Y})\right]\left[\sum(X_1 - \bar{X}_1)(X_2 - \bar{X}_2)\right]}{\left[\sum(X_1 - \bar{X}_1)^2\right]\left[\sum(X_2 - \bar{X}_2)^2\right] - \left[\sum(X_1 - \bar{X}_1)(X_2 - \bar{X}_2)\right]^2}$$

We can substitute deviations scores, $x = X - \bar{X}$, as Cohen and colleagues (2003) do, but the deviations scores save only so much space and time, and the equations still become unwieldly when there are additional predictors.

One shortcut is to use matrix algebra, which is commonly used for many advanced statistics. Matrix algebra (or sometimes "linear algebra")[1] is a system for performing operations on many sets of numbers at once, making it possible to solve many sets of equations simultaneously. Numbers are organized into rows and columns, and each matrix is designated by a letter, usually in bold. For example, all of the predictor variables X_1 to X_k are organized in a single matrix called $\mathbf{X}$. Every matrix is described by dimensions, the number or rows (often r) and number of columns (often c), because the special rules needed for adding, subtracting, multiplying, and dividing matrices depend on the dimensions. For the $\mathbf{X}$ matrix, each predictor, X_1 through X_k, has a separate column and each case (participant, subject), i, has a separate row. Because we have only one dependent variable, there is only one column for the $\mathbf{y}$ matrix, which uses lowercase here because it is a special type of matrix known as a "vector." Regression coefficients, including the intercept, and the residuals (errors) are organized into vectors as well. Below I show the general form for the multiple regression model (using Pedhazur's notation, where a is used for the intercept and b_k is used for slopes) and, below that, the corresponding notation is given for the matrix form.[2]

General form of the multiple regression model:

$$Y = a + b_1X_1 + b_2X_2 + \ldots + b_kX_k + e \qquad (6.1)$$

Matrix form of the multiple regression model:

$$\underset{N\times 1}{\mathbf{y}} = \underset{N\times(1+k)}{\mathbf{X}} \quad \underset{(1+k)\times 1}{\mathbf{b}} + \underset{N\times 1}{\mathbf{e}} \qquad (6.2)$$

Notice the dimensions given below each matrix, where N is used for the number of cases and k is used for the number of predictors. If we were to show all of the elements in each of these matrices, it would look something like this:

---

[1] This is a common term in mathematics to refer to matrix algebra, but it can be confusing because it is not helpful in distinguishing linear algebra using matrices and linear algebra using more traditional ("scalar") equations. The term "vector algebra" may also be used, in which "vector" refers to a matrix has just one row or one column.

[2] Out of laziness, I borrowed copies of the equations from the optional reading by Pedhazur, E.J. (1997). *Multiple regression in behavioral research: Explanation and prediction* (3rd ed.). New York, NY: Wadsworth (pp.136-137).

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \cdot \\ \cdot \\ \cdot \\ Y_N \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{12} & X_{13} & \cdots & X_{1k} \\ 1 & X_{21} & X_{22} & X_{23} & \cdots & X_{2k} \\ 1 & X_{31} & X_{32} & X_{33} & \cdots & X_{3k} \\ \cdot & \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & \cdot & & \cdot \\ 1_N & X_{N1} & X_{N2} & X_{N3} & \cdots & X_{Nk} \end{bmatrix} \begin{bmatrix} a \\ b_1 \\ b_2 \\ \cdot \\ \cdot \\ \cdot \\ b_k \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ \cdot \\ \cdot \\ \cdot \\ e_N \end{bmatrix}$$

Using the special operations for matrices, we could compute all of the regression coefficients (in the **b** matrix) at once. The prime symbol ′ indicates the matrix is transposed, with all rows and columns switched, a necessary operation for multiplication when squaring variables. The $^{-1}$ superscript is for the inverse operation, which is essentially the matrix operation for dividing.

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \qquad (6.3)$$

Of course, any of the other important regression quantities can be computed via matrix algebra as well. Statistical computer packages allow users to perform matrix operations with some special commands[3] and many software programs use matrix algebra in the computer code for solving statistical equations efficiently.

References

Cohen, J., Cohen, P., West, S.G., & Aiken, L.S. (2003). *Applied multiple regression/correlation analysis in the behavioral sciences (Third Edition)*. Mahwah, NJ: Erlbaum.

Kutner, M.H., Nachtsheim, C., & Neter, J. (2004). *Applied linear regression models (4th Edition)*. Chicago, IL: IRWIN.

Pedhazur, E.J. (1997). *Multiple regression in behavioral research (3rd Edition)*. Fort Worth, TX: Harcourt Brace.

[3] The handout "Matrix Operation Examples in SPSS and R" available on the webpage for my structural equation modeling course gives examples of basic operations and how to use them to obtain several regression quantities.