Neuendorf

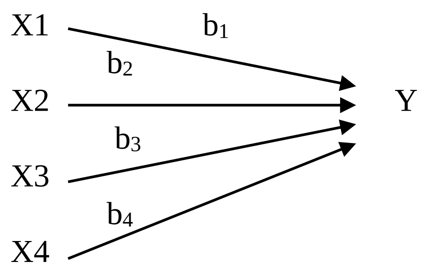
Multiple Regression—Beginnings

Multiple regression investigates the prediction of one DV by 2 or more IVs, all assumed to be measured at the interval/ratio (I/R) level. As we will see, there are two main statistical assessments done by multiple regression. They are:

(1) The proportion of the DV's variance that is explained by one IV or by a set of IVs (an F tests the sig. of an $\mathbf{R^2}$), and

(2) The unique contribution of each IV (an F tests the sig. of each b or β).

Let's suppose that we have collected data on five different variables, and have hypothesized the following model:



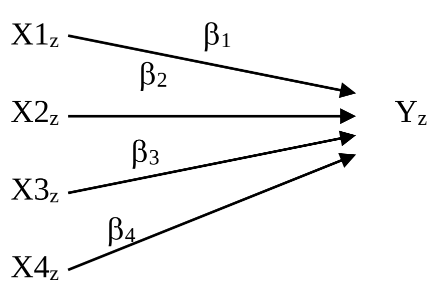
In the original units (as the variables were measured), the equation for the full model with all 4 IVs, with unstandardized variables, is:

$$Y' = a + b_1X1 + b_2X2 + b_3X3 + b_4X4$$

where a = the intercept of the regression line on Y, in other words the value of Y when all X's are zero

b = each unstandardized partial regression coefficient, representing a partial slope, i.e., the impact of that X when controlling for all other X's in the equation

The standardized form of the model is:



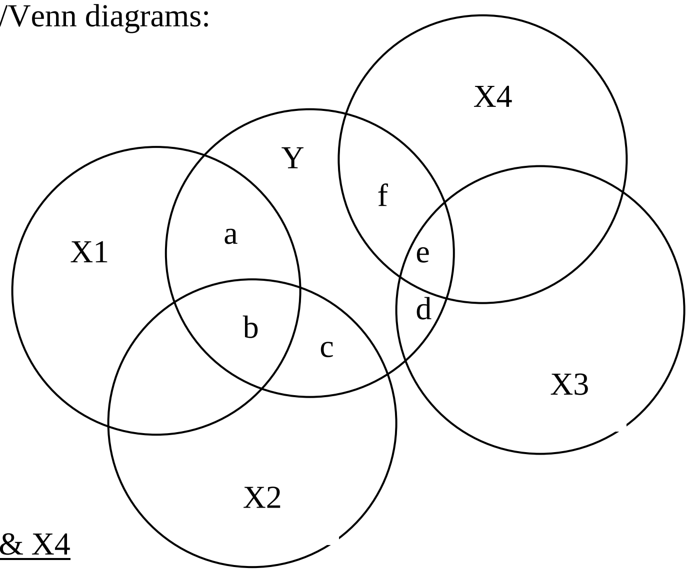
The standardized form of the equation is:

$$Y'_z = \beta_1 X1_z + \beta_2 X2_z + \beta_3 X3_z + \beta_4 X4_z$$

Each partial regression coefficient (unstandardized b or standardized β) is an indicator of the unique contribution of that X to the DV (Y). For example, β_1's statistical significance (tested with an F) indicates whether X1 has a significant linear relationship with Y *when controlling for* X2, X3, and X4. And, β_3's statistical significance (tested with an F) indicates whether X3 has a significant linear relationship with Y *when controlling for* X1, X2, and X4.

We also look at the total variance explained by the IVs. This R^2 (also tested with an F) indicates what proportion of Y's variance is explained by (shared with) the 4 IVs.

This may be shown via Ballantines/Venn diagrams:



Total Contribution of X1, X2, X3, & X4

$$R^2 = a + b + c + d + e + f$$

Unique Contribution for each X

a corresponds to* β_1
c corresponds to* β_2
d corresponds to* β_3
f corresponds to* β_4

* - corresponds to, but does not equal. . . β is not expressed as a proportion; rather, it is a *standardized partial slope* (the amount and direction of change in Y for a unit change in an X, controlling for the other Xs in the equation, assuming all variables are standardized)

Exercise:
While it's unlikely you would find many significant βs with a non-significant R^2, you sometimes do find a significant R^2 without any individually significant βs. What set of relationships among the variables would result in this?

Draw it:

Additional references:

Schroeder, L. D., Sjoquist, D. L., & Stephan, P. E. (1986). *Understanding regression analysis: An introductory guide.* Beverly Hills, CA: Sage.

Lewis-Beck, M. S. (1980). *Applied regression: An introduction*. Beverly Hills, CA: Sage.