Multiple Regression—Beginnings

Multiple regression investigates the prediction of one DV by 2 or more IVs, all assumed to be measured at the 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 a set of IVs (an F tests the sig. of $R^2$), and (2) the unique contribution of each IV (an F tests the sig. of each $\beta$).

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

\[
\begin{align*}
X_1 & \rightarrow b_1 \\
X_2 & \rightarrow b_2 \\
X_3 & \rightarrow b_3 \\
X_4 & \rightarrow b_4 \\
\end{align*}
\]

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

\[
Y' = a + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4
\]

With the four variables each standardized, the model would be:

\[
\begin{align*}
X_{z1} & \rightarrow 3_1 \\
X_{z2} & \rightarrow 3_2 \\
X_{z3} & \rightarrow 3_3 \\
X_{z4} & \rightarrow 3_4 \\
\end{align*}
\]

The standardized form of the equation is:

\[
Y_{z'} = \beta_1X_{z1} + \beta_2X_{z2} + \beta_3X_{z3} + \beta_4X_{z4}
\]

Each partial regression coefficient (unstandardized $b$ or standardized $\beta$) is an indicator of the unique contribution of that X to the DV (Y). For example, $\beta_1$'s statistical significance (tested with an F) indicates whether $X_1$ has a significant linear relationship with $Y$ when controlling for $X_2$, $X_3$, and $X_4$.

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:

Unique Contribution
- a corresponds to $\beta_1$
- c corresponds to $\beta_2$
- d corresponds to $\beta_3$
- f corresponds to $\beta_4$

* - corresponds to, but does not equal. . . $\beta$ 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)

Total Contribution

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

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

Draw it: