Multiple Regression Models

- Advantages of multiple regression
- Important preliminary analyses
- Parts of a multiple regression model & interpretation
- Differences between r, bivariate b, multivariate b &
- Steps in examining & interpreting a full regression model

Advantages of Multiple Regression

Practical issues …

- better prediction from multiple predictors
- can "avoid" picking/depending on a single predictor
- can "avoid" non-optimal combinations of predictors (e.g., total scores)

Theoretical issues …

- even when we know in our hearts that the design will not support causal interpretation of the results, we have thoughts and theories of the causal relationships between the predictors and the criterion -- and these thoughts are about multi-causal relationships
- multiple regression models allow the examination of more sophisticated research hypotheses than is possible using simple correlations
- gives a "link" among the various correlation and ANOVA models

Before launching into the various hypotheses tests and other types of analyses, be sure to "get familiar" with your data and determine if it has any "problems" …

1. Get means and standard deviations for each variable
   - do they "make sense" for these measures & this population?
2. Get the minimum and maximum values for each variable
   - any "extreme" scores -- coding errors, etc. ?
3. Consider the shape of the distribution for each variable
   - any extreme nonnormality or discontinuity?
4. Consider the scatterplot of each predictor with the criterion
   - any important nonlinearities or bivariate outliers?
5. Consider the correlations of each variable with the criterion
   - do they "make sense" for these measures & this population?
6. Consider the correlations among the predictors (collinearities)
   - do they make sense for these measures & this population?
   - will there be a "collinearity problem" ?

raw score regression $y' = b_1x_1 + b_2x_2 + b_3x_3 + a$

each b

- represents the unique and independent contribution of that predictor to the model
- for a quantitative predictor tells the expected direction and amount of change in the criterion for a 1-unit increase in that predictor, while holding the value of all the other predictors constant
- for a binary predictor (with unit coding -- 0,1 or 1,2, etc.), tells direction and amount of group mean difference on the criterion variable, while holding the value of all the other predictors constant

a

- the expected value of the criterion if all predictors have a value of 0

Let's practice -- Tx (0 = control, 1 = treatment)

depression' = (2.0 * stress) - (1.5 * support) - (3.0 * Tx) + 35

- apply the formula patient has stress score of 10, support score of 4 and was in the treatment group dep' =
- interpret "b" for stress -- for each 1-unit increase in stress, depression is expected to by , when holding all other variables constant
- interpret "b" for support -- for each 1-unit increase in support, depression is expected to by , when holding all other variables constant
- interpret "b" for tx – those in the Tx group are expected to have a mean depression score that is than the control group, when holding all other variables constant
- interpret "a" -- if a person has a score of "0" on all predictors, their depression is expected to be

standard score regression $Z_y' = \beta Z_{x1} + \beta Z_{x2} + \beta Z_{x3}$

each β

- for a quantitative predictor the expected Z-score change in the criterion for a 1-Z-unit increase in that predictor, holding the values of all the other predictors constant
- for a binary predictor, tells size/direction of group mean difference on criterion variable in Z-units, holding all other variable values constant

As for the standardized bivariate regression model there is no "a" or "constant" because the mean of Zy' always = Zy = 0

The most common reason to refer to standardized weights is when you (or the reader) is unfamiliar with the scale of the criterion. A second reason is to promote comparability of the relative contribution of the various predictors (but see the important caveat to this discussed below!!!).

Other versions of the regression equation…

using b_0 as constant $$y' = b_0 + b_1x_1 + b_2x_2 + b_3x_3$$

y = y' + error $$y = b_1x_1 + b_2x_2 + b_3x_3 + a + e$$

combining these $$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + e$$

standard score version $$Z_y = \beta_1 Z_{x1} + \beta_2 Z_{x2} + \beta_3 Z_{x3} + e$$

It is important to discriminate among the information obtained from ...

bivariate r & bivariate regression model weights

r -- simple correlation

tells the direction and strength of the linear relationship between two variables ($r = \beta$ for bivariate models)

r^2 -- squared correlation

tells how much of the Y variability is "accounted for,"
. "predicted from" or "caused by" X ($r = \beta$ for bivariate models)

b -- raw regression weight from a bivariate model

tells the expected change (direction and amount) in the criterion for a 1-unit increase in the predictor

β -- standardized regression wt. from a bivariate model

tells the expected change (direction and amount) in the criterion in Z-score units for a 1-Z-score unit increase in that predictor

It is important to discriminate among the information obtained from ...

multivariate R & multivariate regression model weights

R^2 -- squared multiple correlation

tells how much of the Y variability is "accounted for,"
. "predicted from" or "caused by" the multiple regression model

R -- multiple correlation (not used that often)

tells the strength of the relationship between Y and the
. multiple regression model

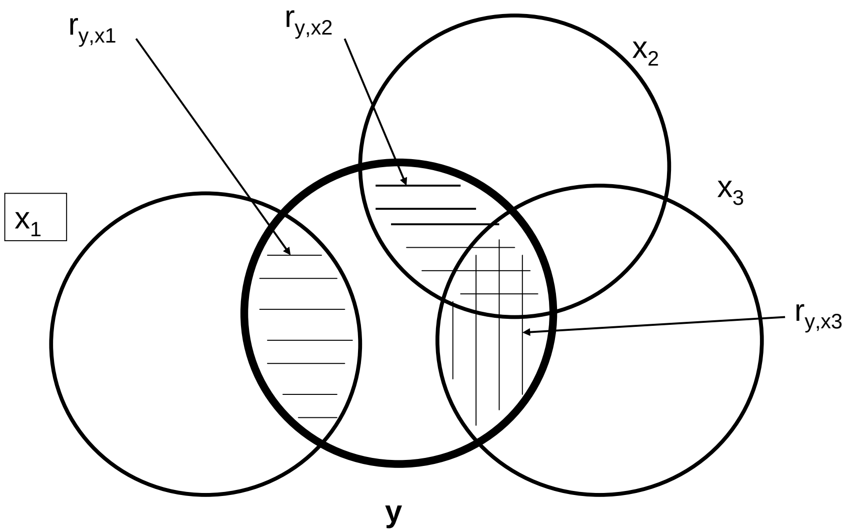
b_i -- raw regression weight from a multivariate model

tells the expected change (direction and amount) in the criterion for a 1-unit increase in that predictor, holding the value of all the other predictors constant

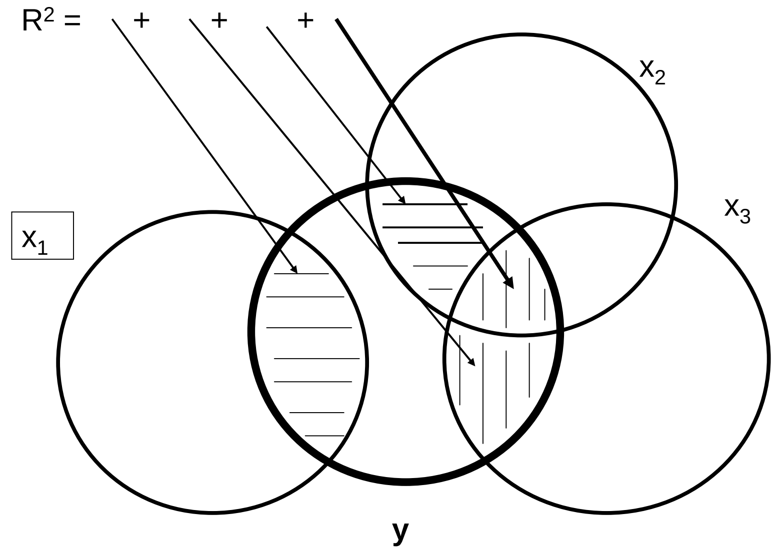
β -- standardized regression wt. from a multivariate model

tells the expected change (direction and amount) in the criterion in Z-score units for a 1-Z-score unit increase in that predictor, holding the value of all the other predictors constant

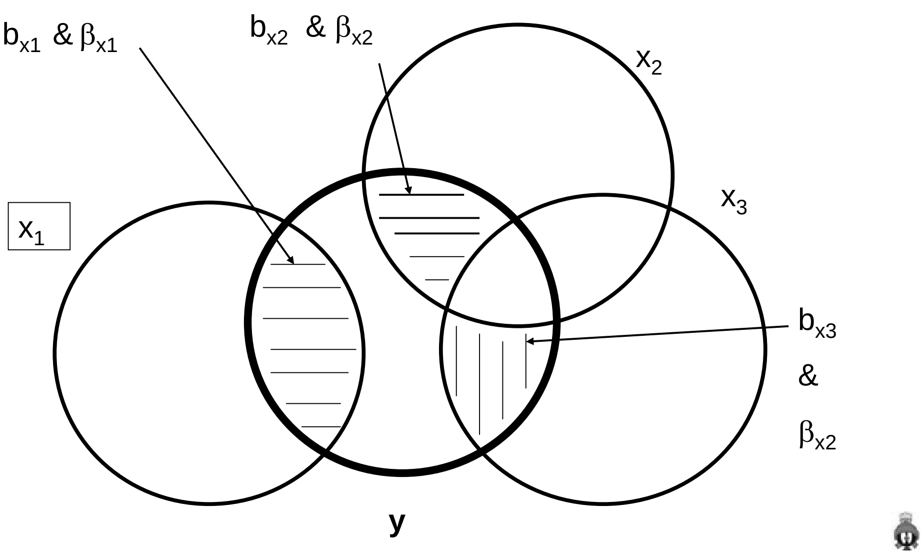
Venn diagrams representing r, b and R^2



Remember R^2 is the total variance shared between the model (all of the predictors) and the criterion (not just the accumulation of the parts uniquely attributable to each predictor).



Remember that the b of each predictor represents the part of that predictor shared with the criterion that is not shared with any other predictor -- the unique contribution of that predictor to the model



Inspecting and describing the results of a multiple regression formula …

0. Carefully check the bivariate results

1. Does the model work?

Multiple -- F-test (ANOVA) of H0: $R^2 = 0$ (R=0)

$$F = \frac{(R^2) / k}{(1 - R^2) / (N - k - 1)}$$

k = # preds of in the model

N = total number of subjects

Find F-critical using df = k & N-k-1

2. How well does the model work?
   - R^2 is an "effect size estimate" telling the proportion of variance of the criterion variable that is accounted for by the model
   - adjusted R^2 is an attempt to correct R^2 for the inflation possible when the number if predictors is large relative to the sample size (gets "mixed reviews" -- replication is better!!)

3. Which variables contribute to the model ??
   - t-test of H0: b = 0 for each variable

4. Which variables contribute "most" to the model
   - ***careful*** comparison of the predictor's βs
   - don't compare predictor's bs – more about why later!

5. Identify the difference between the "bivariate story" told by the correlations and the "multivariate story" told by the multiple regression weights – more later