Important Regression Stuff

What things are comparable?

The following are all really the same question …

1. Does X1 contribute to a model also involving X2 and X3?

    This is most directly tested using the t-test of the b-weight for X1 in this multivariate model…

    $$Y' = bX1 + bX2 + bX3 + a$$

2. Does dropping X1 from the X1, X2 & X3 model decrease the R^2 of the model?

    This is most directly tested using the R^2? F-test comparing these models …

    $$R^2y.X1,X2,X3 > R^2y.X2,X3$$

3. Will adding X1 to a model including X1 & X3 increase the R^2 of the model?

    This is most directly tested using the R^2? F-test comparing these models …

    $$R^2y.X2,X3 < R^2y. X1,X2,X3$$

    The difference between 2 & 3 is simple whether we conceptualize these nested models as (2) deleting a predictor from the full model or (3) adding a predictor to the reduced model to make the full model.

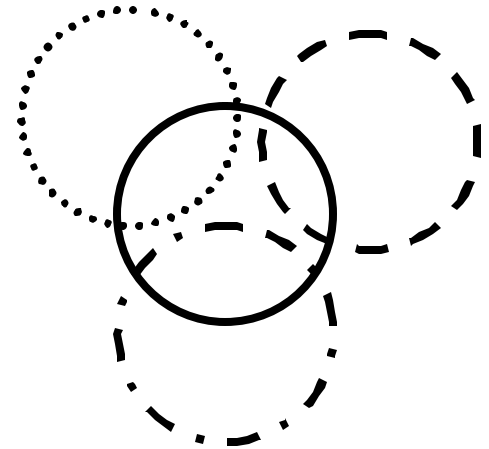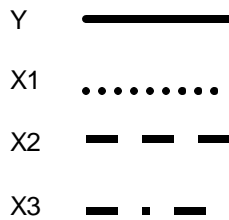4. Is there a multiple semi-partial correlation between Y and X1 after correcting X1 for X2 & X3?

    $$r^2_{y(1.2,3)}$$ the significance test for this is the same F-test as for 2 & 3 above

Adding things up to get R²

Multiple regression can be seen as a way of adjusting a set of collinear predictors of a criterion so that they are "uncorrelated". **OR** Multiple regression is a of way of asking, "What is the part of each predictor that is independent of the other predictors, and how does that unique part of each predictor contribute to the multivariate model?"

If there is no collinearity among the predictors, then R² is easy to calculate…

$$R^2_{y.1,2,3} = r^2_{y,1} + r^2_{y,2} + r^2_{y,3}$$

Y   ——————

X1 ··········

X2 — — — —

X3 — · —

If there is collinearity, we must adjust each predictor for each "previous" predictor in the model (these 3 are equivalent in that they produce the same R²)…

1) $R^2_{y.1,2,3} = r^2_{y,1} + r^2_{y(2.1)} + r^2_{y(3.12)}$

OR

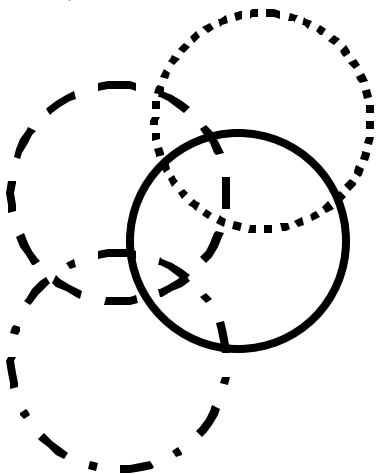2) $R^2_{y.1,2,3} = r^2_{y,2} + r^2_{y(1.2)} + r^2_{y(3.21)}$

OR

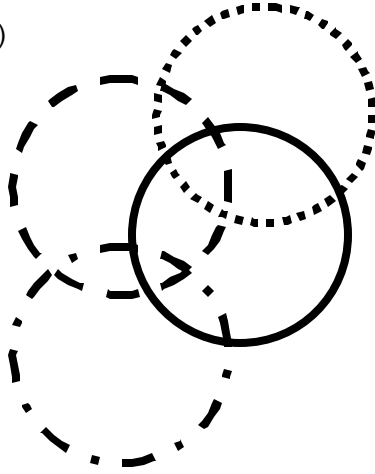3) $R^2_{y.1,2,3} = r^2_{y,3} + r^2_{y(1.3)} + r^2_{y(2.31)}$

Notice: The "order of entry" of predictors into the model doesn't change the model's R²! Nor will it change the β of any predictor *in the full model* (more below)

Notice: Since we are trying to "adjust" for collinearity among the predictors, the contribution of each predictor is controlled for previous predictors included in the model (see below for the difference between this and β)
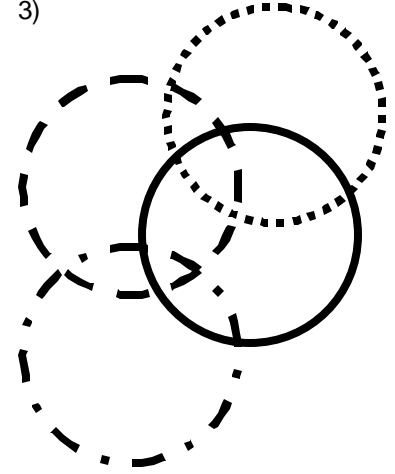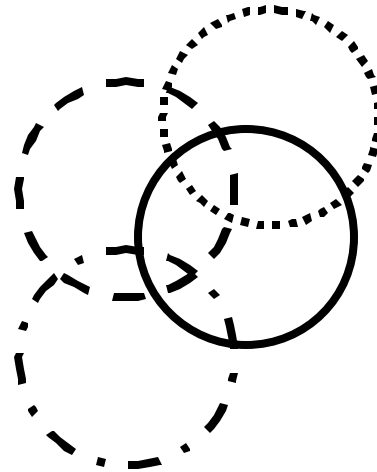
1)

2)

3)

Notice: The "order of entry" of predictors into the model doesn't change the model's R^2!
Nor will it change the β of any predictor *in the full model* (more below)

Notice: Since we are trying to "adjust" for collinearity among the predictors, the contribution of each predictor is controlled for previous predictors included in the model (see below for the difference between this and β)

Pay attention, this is where it gets tricky…. (the difference between **R^2** and accumulated **b** weights)
You cannot simply accumulate variance accounted for by each predictor to get the R^2.
- This is because this process ignores the collinearity among the predictors
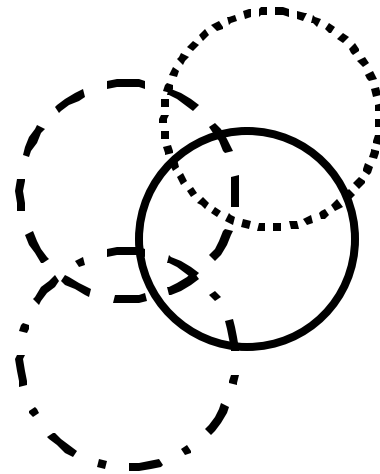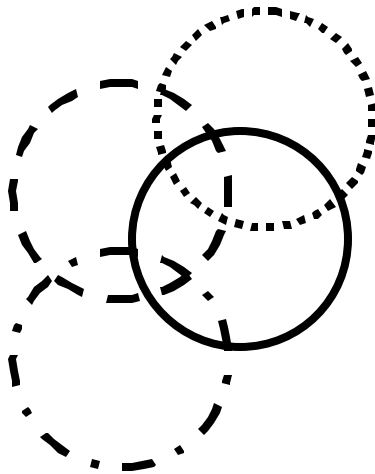- So… $R^2_{y.1,2,3} < r^2_{y,1} + r^2_{y,2} + r^2_{y,3}$

Neither can you simply accumulate the the squared semi-partials **or** the squared β weights from all the predictors to find R^2.
- This is because the semi-partials and βs don't include the variance that predictors share with the criterion and each other (the collinear part)! So, we get…

$$R^2_{y.1,2,3} > r^2_{y(1.2,3)} + r^2_{y(2.1,3)} + r^2_{y(3.1,2)} \qquad \textbf{\&} \qquad R^2_{y.1,2,3} > \beta_1^2 + \beta_2^2 + \beta_3^2$$

Carefully compare the model on the left with the top model of the three above – that model, but not this, includes in R^2 the part of the y,1 relationship that is shared with 2 & 3. Likewise, that model, but not this, shares the part of the y,2 relationship that is independent of 2, but related with 3. Those parte **are** the differences between "contribution to the model" and "independent contribution to the model" – only the latter is expressed by the multiple semi-partial correlations and regression weights.