Research Hypotheses and Multiple Regression

- Kinds of multiple regression questions
- Ways of forming reduced models
- Comparing "nested" models
- Comparing "non-nested" models

When carefully considered, almost any research hypothesis or question involving multiple predictors has one of four forms:

1. The impact of ***adding or deleting*** one or more particular predictors from to a specified model

   - Whether or not adding that subset will "help" the model, (i.e., increase the R^2 significantly)

   - This involves comparing "nested models" using the R^2_Δ

2. The impact of ***substituting*** one or more predictors for one or more others

   - Whether the one model "works better" than the other (I.e., has a larger R^2)

   - This involves comparing "non-nested models" (t- or Z-test)

Research Hypotheses and Multiple Regression, cont.

3. The differential performance of a specific model across two or more groups (populations, treatments, etc.)

   - Whether the model produces equivalent R^2 for the groups

   - Allows us to look for much more than "mean differences"

   - Important for population generalizability questions

   - This involves comparing "nested models" (Fisher's Z-test)

4. The differential performance of a specific model for predicting two or more different criterion variables

   - Whether the model produces equivalent R^2 for the criterion variables

   - Important when get "conflicting" results across measures

   - This involves comparing "correlated R^2" (t- or Z-test)

About comparing Full vs. Reduced (nested) models …

Full model -- model involving "all the variables"

Reduced model -- model involving "some subset" of the variables

Ways of forming reduced models:

Theory -- some variables are "more important" from a theoretical
    perspective, and the question is if a subset of
    variables accounts for the criterion variable as well as
    the full model (e.g., will adding MMPI scores improve a
    model of drinking behavior that is based on only
    demographic variables?)

Pragmatic -- will a subset of the "less costly" predictors do as well
    as a model that included them and the more expensive
    ones (e.g., will adding a full scale IQ measure (would cost
    us $250) to a model using GRE scores ($0 for us)
    improve our selection of graduate students?)

Summary of ways of constructing reduced models:

• only include variables with significant simple correlations

    • nah -- ignores suppressor variables & is *atheoretical*

• only include variables with significant contributions to full model

    • nah -- ignores collinearity patterns & is *atheoretical*

• use automated/statistical model construction techniques

    • nah -- doesn't work as intended & is *atheoretical*

• select a subset of variables based on theory or
    availability/economics that might be "sufficient" (perform
    equivalently to the full model)

    • yeah !!!

Keep in mind that the hypothesis/question might involve
    comparing two reduced models - one nested in the other.

Comparing "nested models"

$$R^2_{y.x1,x2,x3,x4} \quad \text{vs.} \quad R^2_{y.x1,x2}$$

$$H0: R^2_{y.x1,x2,x3,x4} = R^2_{y.x1,x2}$$

$$F = \frac{(R_L^2 - R_S^2) / (k_L - k_S)}{(1 - R_L^2) / (N - k_L - 1)}$$

R_L^2 = R^2 of the larger model
R_S^2 = R^2 of the smaller model
k_L = # preds of larger model
k_S = # preds of smaller model
N = total number of subjects

Find F-critical using df = $(k_L - k_S)$ & $(N-k_L-1)$

If retain H0: $R_L^2 = R_S^2$ Larger model "does no better"

If reject H0: $R_L^2 > R_S^2$ Larger model "does better" than smaller

1st Important thing about comparing models using R^2_Δ...

   A model with better R^2 predicts y' better "on average"
                              (averaging across participants)

 The R^2_Δ is computed using the whole sample

 • Different people have different "best predictors"

 • Adding one or more predictors that increase R^2 for most

      participants (or "on average") can actually decrease

      predictive accuracy for some individuals

 • This can happen for whole subgroups (strata)

    • a model can "get better on average" and "get worse for a
       subgroup" at the same time

    • major issue for both correlation and prediction research!!!!!

 Good ideal to look at distribution of residuals from any model –
              looking for subgroup differences!!!

2nd Important thing about comparing models using R^2_Δ…
       The additional predictors contribute to the model "on average"
                              (averaging across predictors)

Notice the numerator of the F-test

• The change in R^2 is divided by the number of predictors changed

• This makes sense, because an R^2 change of .20 from adding or
       dropping a single predictor is "more impressive" than the
       same change of .20 from adding or dropping 4 predictors

• This test actually asks if the "average contribution to the R^2" of
       the particular variables that are added to or dropped from
       the model is significantly different from 0.00.

   • So, a significant R^2_Δ does not mean all the added predictors
      contribute to the model!!!

• The impact of adding or dropping predictors from a model may
       depend upon "packaging", how many are involved, etc.

 Good ideal to look at what predictors do and don't contribute to
                  any model you are comparing!!!

Let's start with a model having 4 predictors -- R^2 = .30
        & a RH about adding 4 of other predictors (x1, x2, x3, x4)

• (for this example) given the sample size, etc, let's say the R^2-
       change will be significant if the average contribution of the
       added predictors is 5% (.05)

• Let's also say that two of the four predictors (x1, x2) have
       contributions of 8% each and two (x3 and x4) have
       contributions of 4% each

• If we add the four predictors simultaneously, the average R^2
       change of 6% will be significant -- adding the four
       predictors will be interpreted as "helping" the model

• IF we add just x1 and x2, the average increase of 8% will be
     significant -- adding just these two "helps" -- the same
     story we got about these two when we added them
     along with the others

• IF we add just x3 and x4, the average increase of 4% will not be
     significant -- adding these two "doesn't help" -- not the
     story we got when we added them along with the other two

Consider what the various results would be if an average R^2-
     change of 5% or 3% were necessary for a significant change

Applying this knowledge of R^2-change will allow us to consider
changes in multiple predictor models, for example…

We know …

• dropping a contributing predictor from a model will lower R^2
     significantly

• dropping a non-contributing predictor from a model will lower R^2
     numerically, but not significantly

• adding a contributing predictor to a model will raise R^2
     significantly

• adding a non-contributing predictor to a model will raise R^2
     numerically, but not significantly

Usually (in most texts) we are told that we can't accurately
anticipate the results of adding or dropping more than one
variable at a time -- but this is not strictly correct !!!

Consider what would happen if we dropped 2 non-contributing
predictors from a multiple regression model

• we are told that when we've dropped one of the predictors, that
     the other might now contribute (we've changed the
     collinearity mix)

• but consider this…

   • neither of the predictors, if dropped by itself will produce a
     significant R^2 change

   • so the average R^2 change from dropping the two shouldn't
     be significant either

   • thus, we can drop both and expect that the R^2 change won't
     be significant)

• this logic is useful, but becomes increasingly precarious as
     sample size drops, collinearity increases or the number of
     predictors in the model or being dropped increases

Similarly…

Dropping two predictors that both contribute should produce an average R^2 change that is significant (same logic in reverse)

However, things get "squirrelly" when considering dropping one contributing and one non-contributing predictor

- we have no good way of considering whether the average R^2 change will or will not be significant

We will consider these issues, their applications and some "variations" when we look at the workings of statistical/automated modeling procedures.

The moral of the story…

• Because this R^2-change tests really tests the average R^2-change of the set of added or dropped predictors, the apparent contribution   of a added variable may depend upon the variables along with which it is added or dropped

• Adding or dropping large sets of variables simultaneously can make the results harder to interpret properly

Because of this...

• Good RHs usually call for arranging the addition or deletions of items in small, carefully considered sets

• Thus, most Rhyps of this type use the addition or removal of multiple sets (each with a separate R^2-change test)

• This is called hierarchical modeling -- the systematic addition or removal of hypothesized sets of variables

---

But wait, there's more…

When we plan to add multiple sets of predictors, we have to carefully consider the ORDER in which we add the sets! WHY???

Remember when we considered the meaning of a predictors regression weight in a multiple regression model … we were careful to point out that the test of that regression weight (or the R^2-change when dropping that predictor from the model) only tests the contribution of that predictor to that particular model.

Same thing when adding (or dropping) a set of predictors from a model -- the test of the R^2-change only tests the contribution of that set of predictors to that particular model.

In other words ... whether or not a set of predictors contributes to "a model" may depend upon the particular model from which they are added (or dropped) -- or the order in which the groups are added (or dropped)

In general, this type of hierarchical modeling starts with the "smallest model", and then proceeds by adding selected variables of sets of variables, until the "full model" is reached.

So, how does one select the order of adding variables or groups of variables from the model ??

There's a general rule --The more important the variable is to your hypothesis, the more conservatively (later in the sequence) you should test it"

In hierarchical modeling, this means that the most interesting variables are entered "later", so that they must have a unique contribution to the most complete model.

Said differently, the most conservative test of a variable is whether or not it contributes to a larger (rather than a smaller) model.

Examples of applying this to testing theoretical RHs

Many hierarchical modeling efforts have three basic steps…

1. Enter the demographic variables

2. Enter the "known predictors" (based on earlier work)

3. Enter the "new variables" (the ones you are proposing make up an important, but as yet unidentified, part of understanding this criterion variable)

This provides a conservative test of the "new variables", because they must "compete" with all the other variables and each other in order to become a contributing predictor in the model.

To show that your "new variables" are correlated with the criterion, but don't contribute beyond what's accounted for by the "demo + old" model, often isn't sufficient (your variables don't add anything)

An important variation of hierarchical modeling, that is applied in psychometric and selection situations, is the demonstration of incremental validity. A predictor or instrument has incremental validity, when it increases the predictive accuracy (R^2) of an existing model.

A common situation is when someone (theoretician or salesperson) wants to "add a new instrument" to the set of predictors already being used. What is convincing evidence??

• A "significant" correlation between the "new" predictor and the criterion isn't sufficient

  • Even if the "new" predictor has a stronger correlation with the criterion than any of the predictors already in use

• To be impressive, the new predictor must "add to" the predictor set already in use -- that's incremental validity

• the "new" is added into the model last, and the R^2-change tested -- providing a conservative test of its utility

Comparing "non-nested models"

Another important type of hypothesis tested using multiple regression is about the substitution of one or more predictors for one or more others.

Common bases for substitution hypotheses :

• Often two collinear variables won't both contribute to a model -- you might check of there is an advantage of one vs. the other being included in the model

• You might have a hypothesis that a variable commonly used in a model (theory) can be substituted for with some other variable

• You might be interested in substituting one (or more) inexpensive (or otherwise more available) predictors for one that is currently used

Non-nested models are compared using either Hotelling's t or Rosenthall's Z (which may be slightly more robust) -- the same tests used to compare "correlated correlations"

   • highlights that R is really r (or that $r_{y,x} = r_{y,y'}$)

These formulas includes not only the R^2 of each model, but also the collinearity of the two models ($r_{x1,x2}$)

Some basic principles of these tests . . .

• the more correlated the models are to each other, the less likely they are to be differentially correlated with the criterion

   • obviously the more predictors two models share, the more collinear they will be -- seldom does the substitution of a single predictor within a larger model have a significant effect

   • the more collinear the variables being substituted, the more collinear they will be -- for this reason there can be strong collinearity between two models that share no predictors

• the weaker the two models (lower R^2), the less likely they are to be differentially correlated with the criterion

•these are not very powerful tests !!!

   • compared to avoiding a Type II error when looking for a given r , you need nearly twice the sample size to avoid a Type II error when looking for an r-r of the same magnitude

   • these tests are also less powerful than tests comparing nested models

So, be sure to consider sample size, power and the magnitude of the r-difference between the non-nested models you compare !

Important thing to consider about comparing non-nested models…

Non-nested models can have different numbers of predictors!!

If so, some argue, the larger model "has an advantage"

• one mathematical solution is to compare the two models using the R^2/k from each model, rather than R^2s

• those who like this approach say it makes the R^2s "more comparable", so that a larger model doesn't have an advantage

• those who like this approach say this helps prevent "hiding" a noncontributing predictor in a well-performing model

• same folks often suggest only comparing models for which all predictors contribute – which has its own problems…

• those who don't like this approach say it is now comparing not "the models" but the "average contribution of the predictors" – not the same thing!