Welcome to Talk Stats! Join the discussion today by registering your FREE account. Membership benefits: • Get your questions answered by community gurus and expert researchers. • Exchange your learning and research experience among peers and get advice and insight. **Join Today!**

Thread: Relative weights (people's thoughts)

Thread Tools Display

09-17-2012 08:43 PM #1

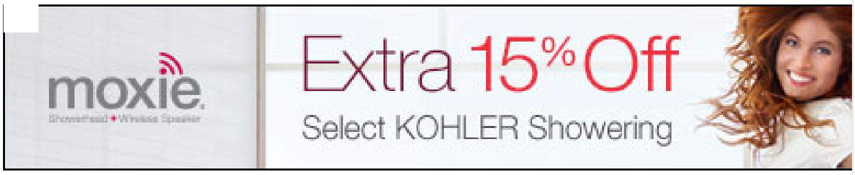
trinker
ggplot2orBust
m.v.c.

Awards:

Location:	Buffalo, NY
Posts:	4,006
Thanks:	1,418

Thanked 775 Times in 690 Posts

Relative weights (people's thoughts)



I recently came across a paper and an R function to calculate relative weights. When I do Hierarchical block-wise multiple regression I'm always bothered by which predictor I put into the model first as having an impact on the delta R squared. THis comes to light with predictors like race and SES (though theory may tell us to do both variables as a block). Anyway I thought it was interesting and wondered what others though:

> Originally Posted by **R in Action**
>
> *There have been many other attempts at quantifying relative importance. Relative*
> *importance can be thought of as the contribution each predictor makes to R-square,*
> *both alone and in combination with other predictors. Several possible approaches to*
> *relative importance are captured in the relaimpo package written by Ulrike Grömping*
> *(http://prof.beuth-hochschule.de/groemping/relaimpo/)*
> *A new method called relative weights shows significant promise. The method closely*
> *approximates the average increase in R-square obtained by adding a predictor variable*
> *across all possible submodels (Johnson, 2004; Johnson and Lebreton, 2004; LeBreton*
> *and Tonidandel, 2008). A function for generating relative weights is provided in the*
> *next listing.*
> *Listing 8.16 relweights() function for calculating relative*

> *importance of predictors*

Code: select all

```
relweights <- function(fit,...){

  R <- cor(fit$model)
  nvar <- ncol(R)
  rxx <- R[2:nvar, 2:nvar]
  rxy <- R[2:nvar, 1]
  svd <- eigen(rxx)
  evec <- svd$vectors
  ev <- svd$values
  delta <- diag(sqrt(ev))
  lambda <- evec %*% delta %*% t(evec)
  lambdasq <- lambda ^ 2
  beta <- solve(lambda) %*% rxy
  rsquare <- colSums(beta ^ 2)
  rawwgt <- lambdasq %*% beta ^ 2
  import <- (rawwgt / rsquare) * 100
  lbls <- names(fit$model[2:nvar])
  rownames(import) <- lbls
  colnames(import) <- "Weights"
  barplot(t(import),names.arg=lbls,
     ylab="% of R-Square",
     xlab="Predictor Variables",
     main="Relative Importance of Predictor Variables
",
     sub=paste("R-Square=", round(rsquare, digits=3))
,
     ...)
```

> Originally Posted by **R in Action**
>
> *NOTE The code in listing 8.16 is adapted from an SPSS program generously*
> *provided by Dr. Johnson. See Johnson (2000, Multivariate Behavioral Research, 35,*
> *1–19) for an explanation of how the relative weights are derived.*
> *In listing 8.17 the relweights() function is applied to the states data with murder*
> *rate predicted by the population, illiteracy, income, and temperature.*
> *You can see from figure 8.19 that the total amount of variance accounted for by*
> *the model (R-square=0.567) has been divided among the predictor variables. Illiteracy*
> *accounts for 59 percent of the R-square, Frost accounts for 20.79 percent, and so forth.*
> *Based on the method of relative weights, Illiteracy has the greatest relative importance,*
> *followed by Frost, Population, and Income, in that order.*

"If you torture the data long enough it will eventually confess."
-Ronald Harry Coase -

Reply With Quote

09-17-2012 10:53 PM #2

spunky

King of all Drama

m.v.c.

Re: Relative weights (people's thoughts)

Everything you love, now on your TV. For $35.

Location: vancouver, canada
Posts: 1,640
Thanks: 118
Thanked 380 Times in 309 Posts

oh god... dont even get me staaaaarted there...:P

i began going into the whole debate of relative importance of predictors in multiple regression through my advisor who helped to formalize, alongside with a colleague of his, Pratt's measure of relative importance (which is also in dr. grömping's relaimpo package) for small sample sizes. actually, one of the papers i presented on this last AERA was a simulation that a friend of mine and i designed where we documented the performance of a whole bunch of these measures in particular cases, such as suppression and more than two or three predictors.

what i got from this debate is:

a) no one really agrees on what importance is
b) the way people define importance impacts the measure that is used to measure it, so when you marry a measure you're implicitly marrying their definition of importance
c) there doesnt seem to be a way out of this so that everyone is satisfied because importance is not a parameter that can be estimated in a population. it's just something people came up with.

my advisor, as you can imagine, was not happy with it because i felt like he sort of wanted his measure of importance to be the best one among the other ones. budescu & azen's dominance analysis is by far the dominant approach right now. johnson & lebreton's relative weights are presented as a shortcut to it because dominance analysis requires all possible combinations of regression equations for a given set of predictors, so you end up having $2^n - 1$ regression equations where n is the number of predictors and the calculations can becomes really cumbersome. what do i think about johnson & lebreton's approach? it is kinda weird, to be honest. the partition of the R-squared is done over those things that look like principal components (but which are not) because they need the orthogonality property among the predictors so the variance explained can be purely additive. however, at least in my opinion, i'm not really intersted in the importance of convenient, orthognal transofrmations of the predictors but on the DARN PREDICTORS THEMSELVES. relative weights also ignores suppression effects which usually provide a lot of very rich information if you're attempting to build pseudocausal models models through the use of regression weights (notice the 'pseudo' in front of the adjective 'casual' please :P)

dominance analysis seems promising in my opinion but i think my advisor has a point when he mentioned that you're averaging variance explained over a whole bunch of regression models that you simply arent intersted in. i thought that was sort of neat but i do see his point. besides, the computational aspect is cumbersome and i dont think it can be extended (at least not easily) to the generalized linear model. they did so for logisitc regression but they stopped right there.

last but not least is the axiomatic Pratt approach (my advisor's championed method) which is beautiful mathematically (it uses the geometry of regression to provide an efficient partition of R-squared through orthognal projections) but it does have the nasty habit of turning negative... and explainig negative importance is well... i dunno. challenging. HOWEVER, the beauty of its axioms is that it can easily be extended to anything linear. a PhD dissertation from the lab i'm a member of won the AERA quantitative dissertation something something award for extending this method to both exploratory and confirmatory factor analysis, so suddenly the issue of crossloadings among items becomes much more easy to understand.

and those are just a few methods. if i remember correctly, there is another article from Grömping where she used regression trees to define importance (particularly in the case where you have more predictors than sample units) and relaimpo has a whole bunch of other stuff there. including the always faithful and forever favourite of all the social science people: the semi-partial correlation.

my advice after a semester dwelling on this debate of relative importance of predictors in multiple regression? read all the theory behind each one, specifically how they define importance and see if their definition makes sense to you from a theoretical standpoint. choose wisely and caveat emptor

for all your psychometric needs!
https://psychometroscar.wordpress.com/about/

Reply With Quote

The Following 2 Users Say Thank You to spunky For This Useful Post:

GretaGarbo (09-18-2012), trinker (09-17-2012)

+ Reply to Thread

Like 0 Tweet 0 +1 0

AdChoices ▶ Linear Regression ▶ Standard Weights ▶ Analysis ▶ The Relative

« Predict variation | Question on Finding the Design Matrix »

AdChoices ▶ The Relative ▶ A Relative ▶ Regression Models ▶ Weight Weights

Advertise on Talk Stats

All times are GMT -5. The time now is 11:38 AM.

Powered by vBulletin™ Version 4.1.3
Copyright © 2014 vBulletin Solutions, Inc. All rights reserved.

Copyright © 2005-2014, talkstats.com