

Dominance Analysis

Regression can be used to solve two very different questions. First it can be used for prediction. Second, it can be used for inference. Dominance analysis can be useful when determining predictor importance if the independent predictor variables are correlated.

Note that the overall R^2 is the sum of the importance coefficients. Consequently, the recommended metric is the *lmg* metric and this will always produce a non-negative value. However, adding the “rela = TRUE” will normalize the sum of the *lmg* to 1.0. In this example it is easy to see that the percent with a college education has the greatest importance when predicting income in 2009.

```
library(relimpo);
states <- read.csv("dominance-data.dat", sep="\t", header=T);
row.names(states) <- states$STATE;
states <- subset(states, select=c(-STATE));
calc.relimp(states,
             type = c("lmg", "last", "first", "betasq", "pratt", "genizi", "car") )

crf <- calc.relimp(INCOME2009~COLLEGEDEGREE+BLACKPER+POPULATION+POP65,states,
                  type = c("lmg", "last", "first", "betasq", "pratt"), rela = TRUE );
crf;
linmod <- lm(INCOME2009~COLLEGEDEGREE+BLACKPER+POPULATION+POP65,states);
crlm <- calc.relimp(linmod,
                   type = c("lmg", "last", "first", "betasq", "pratt", "genizi", "car"), rela = TRUE
plot(crlm);

cor(states)

calc.relimp(states,
             type = c("lmg", "last", "first", "betasq", "pratt"), rela = FALSE,
             always = "COLLEGEDEGREE" );
```

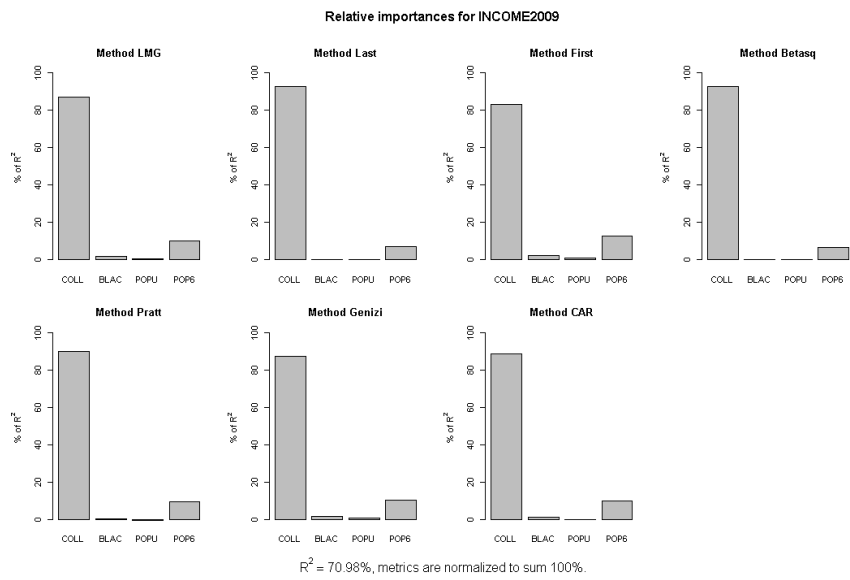


Figure 1: Relative Importance for 2009 Income