3

Worked Example for Dominance Analysis

This example uses the ASA software integrated into Excel or SPSS (www.asastat.com). ASA is, in part, a point-and-click interface to R but analyses can be conducted from within SPSS or Excel. All data are hypothetical. We assume you have read the primer on dominance analysis.

We first conduct a multiple regression analysis using ordinary least squares (OLS) regression predicting the annual salary of professors (variable called *salary*) from (1) the number of years since getting a Ph.D. (variable called *timephd*), (2) the number of publications (variable called *pubs*), (3) gender (variable called *dfemale*, scored 1 = female, male = 0), and (4) the number of times the person's research has been cited by others (variable called *citations*). We use the multiple regression program in ASA called "Multiple regression from raw data" that is in the folder "Multiple Regression: General > Multiple Regression Analysis." After presenting the results for this analysis, we perform a dominance analysis of the predictors.

The ASA software routinely reports confidence intervals for key parameters in statistical models. There are different ways of presenting confidence intervals. One strategy is to report them directly. Another strategy is to report them as margins of error, much like the margins of error you see for political polls on television or in print media. In this case, one calculates the half width of the confidence interval and reports it in "plus or minus" format. For example, in a political poll, you might be told that the percent of people endorsing a candidate is 50% ±5%. In this case, the confidence interval is 45% to 55%. This is an efficient way of summarizing the interval. In some cases, confidence intervals are asymmetric. When this occurs, some researchers will report the lower and upper margin of error separately. Alternatively, the researcher might calculate the absolute difference between the lower limit and the parameter estimate as well as the absolute difference between upper limit of the interval minus the parameter estimate and then report whichever difference is larger using the ± format. Some analysts prefer the use of credible intervals in Bayesian analytic frameworks instead of confidence intervals for characterizing margins of error (see Curran, 2005).

The first section of the output for the multiple regression program provides information about overall model fit:

```
MODEL RESULTS

Model df: 4, 745
Model F: 181.5651
Model p value: 0.000000
Model R: 0.702589
Model R squared: 0.493631
Standard error of estimate (see): 6771.2427
95% confidence interval for see: 6421.9120 to 7097.7320
Lower and upper margin of error for see: -349.3311, 326.4897
```

The squared multiple correlation of 0.49 indicates that the predictors, considered as a collective, account for about 49% of the variation in salary. This is statistically significant ($F(4, 745) = 181.57$, $p < 0.05$). The standard error of estimate is an index of how far off predicted salaries are from observed salaries; it is an index of prediction error. On average, the predicted salaries were "off" by $6,771 ±$349 from the observed salaries.[1]

The program does not report a confidence interval for the squared multiple correlation but we can obtain it from the program "CI and margin of error for a squared multiple correlation" in the suite "Multiple Regression - General > R Squared and Effect Size." Here is the (abridged) output from this program:

```
RESULTS

Squared R: 0.493631
95% confidence interval: .44442 to .53339
Lower and upper margin of error: -.04922, .03976
```

Thus, the squared R is 0.49 ±0.05. (Although the absolute values of the lower and upper margins of error are asymmetrical, they are close enough to use just the larger of the two to simplify presentation).

Here is the output for the intercept:

```
Intercept

   Value of intercept: 33062.3501
   95% confidence interval: 31128.0202 to 34996.6799
   Standard error: 985.3178
   Margin of Error: +/- 1934.330
   t value: 33.5550
   p value: .000000
```

The intercept is the predicted mean salary when all the predictors are zero, i.e., for professors who have 0 years of experience, no publications, no citations, and who are

[1] The margin of error (MOE) for this statistic is based on a percentile bootstrap with 1000 replicates. The absolute value of the lower MOE is fairly close to the upper MOE (given the large metric on which salary is measured), so we characterize the average error in prediction using the larger of the two to simplify presentation.

male. It is \$33,062 ± \$1,934. The intercept maps onto a profile of a professor that is not very realistic, so it is not of much interest in the present case.

Here is the output for the predictor of time since one's Ph.D.:

```
Predictor: TIMEPHD

   Value of coefficient: 826.9518
   95% confidence interval: 675.4822 to 978.4215
   Standard error: 77.1563
   Margin of error: +/- 151.470
   t value: 10.7179
   p value: .000000
```

The value of 826.95 indicates that for every additional year of experience, the mean salary is predicted to increase by \$827 ± \$151, holding constant the other predictors. The coefficient is statistically significant ($t(745) = 10.72$, $p < 0.05$). In expressing the results of the significance test, we use the overall degrees of freedom from the model reported earlier to characterize the degrees of freedom for the t value. Similar interpretations of the regression coefficients apply to the number of publications and the number of citations, so we do not repeat them here.

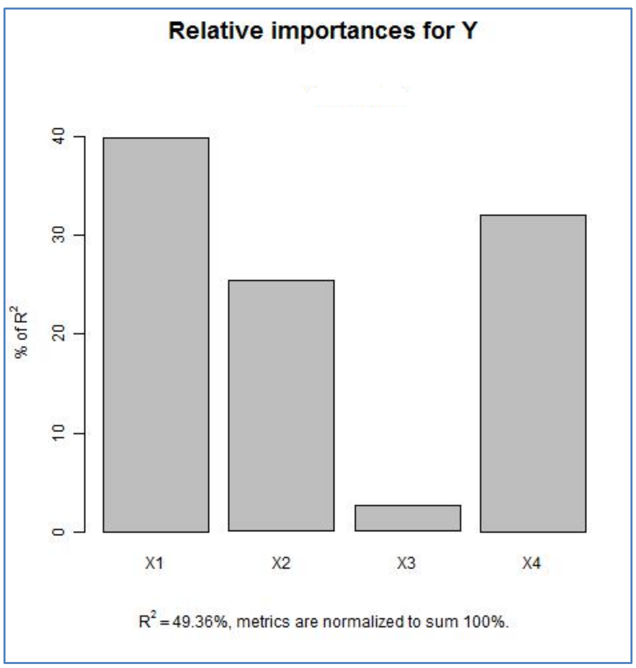
Here is the output for the dummy variable predictor, dfemale:

```
Predictor: DFEMALE

   Value of coefficient: -852.7739
   95% confidence interval: -1845.8008 to 140.2530
   Standard error: 505.8326
   Margin of error: +/- 993.027
   t value: -1.6859
   p value: .092237
```

Because this is a dummy variable, the regression coefficient is a mean difference between the group scored 1 on the variable (females) and the reference group (males). On average, females are paid \$852 ± \$993 less than males, holding constant all other predictors in the equation. The reason we know that females are paid less than males is because the coefficient is negative, which implies a larger mean was subtracted from a smaller mean. The mean difference is not statistically significant ($t(745) = 1.69$, *ns*), so the true population mean difference could conceivably be 0 with the difference we are observing just reflecting sampling error.

We would like to make statements about the relative importance of the different predictors in predicting salary. We use dominance analysis to do so. The dominance index for a predictor is its average unique contribution to the squared R across all possible subsets of predictors. To review material from the primer, let the letters A, B, C and D represent each of the four predictors, respectively. To index the "dominance" of predictor A, we calculate the increase in the R square that A yields over B, that A yields over C, that A yields over D, that A yields over B and C together, that A yields over B and D together, that A yields over C and D together, and that A yields over B, C, and D together. The average of these increases is the index of general dominance. These values can be rescaled or "normalized" to sum to 100 in such a way that they reflect the predictors' relative percent of contribution to the overall explained variance (the squared R) in salary. We used the program for relative importance of predictors in the "Multiple Regression – Relative Importance Analysis" suite to compute dominance indices (in conjunction with percentile bootstrapping to obtain margins of error for them; we also selected the option "Analysis of differences"). This program interfaces with the R package called relaimpo by Ulrike Groemping. Here is the output using the normalized version of importance (note: X1 = years since Ph.D., X2 = number of publications, X3 = gender, and X4 = number of citations), beginning with the generated graph:



INDEX: NORMALIZED GENERAL DOMINANCE (LMG)

Predictor	Importance	MOE (+/-)	Lower CI	Upper CI
TIMEPHD	39.8899	6.2714	34.1727	46.1613
PUBS	25.4102	4.9104	20.4998	30.2988
DFEMALE	2.6190	2.5786	1.1051	5.1976
CITATIONS	32.0809	6.2417	25.8392	38.2311.

Note that the importance indices in the table sum to 100. Time since the doctorate and the number of citations were the two most dominant predictors (importance indices of 39.9 and 32.1, respectively), followed by a slightly lower importance value for the number of publications (25.4). Gender of the professor was not much of a contributing factor to explained variance (importance index of 2.6).

Here is the output for the same dominance analysis but where the indices are not normalized, i.e., they are the dominance values in their own right (multiplied by 100):

```
INDEX: GENERAL DOMINANCE * 100 (LMG)


Predictor    Importance    MOE (+/-)    Lower CI     Upper CI

TIMEPHD      19.6909       3.4055       16.6223      23.0964
PUBS         12.5433       2.9367       9.9783       15.4800
DFEMALE      1.2928        1.2382       0.5577       2.5310
CITATIONS    15.8361       3.5767       12.3975      19.4128
```

The average (unique) contribution of time since the Ph.D. to the prediction of salary across all possible subsets of independent variables was 19.69% ± 3.4%. Note that the sum of the indices is 49.36 or, 0.493 when divided by100, which is the value of the squared R. We personally find the latter indices to be more intuitive than the former indices, but many researchers prefer the normalized indices.

As noted, we selected the option “Analysis of differences” in the relative importance analysis. This yields bootstrapped based significance tests of differences in the relative importance of predictors using the normalized metric . Here is the output:

```
Analysis of Differences

Contrast     Difference    MOE (+/-)    Lower CI     Upper CI

X1 - X2      14.4797       9.3262       5.1535       23.0556
X1 - X3      37.2709       6.2936       30.9773      43.2664
X1 - X4      7.8090        12.0126      -4.2035      18.1657
X2 - X3      22.7912       6.0935       16.6977      28.6748
X2 - X4      -6.6707       9.9736       -16.4672     3.3029
X3 - X4      -29.4619      7.1448       -36.4586     -22.3171
```

If the confidence interval does not contain zero, the difference is statistically significant at $p < 0.05$. If a test is significant for the normalized metric, then it also is significant for the non-normalized metric.

Here is how we might write-up these results for a report assuming we have already dealt with the issues of regression assumptions and that we have explained how we are defining margins of errors (e.g., “Margins of errors (MOEs) are calculated from 95% confidence intervals and are the absolute distance between the lower limit or upper limit of the interval minus the parameter estimate, whichever is larger, unless otherwise noted”):

“The squared multiple correlation for the model was 0.49 ±0.05, which was statistically significant ($F(4, 745) = 181.57$, $p < 0.05$). The standard error of estimate was $6,771 ±$349, with the MOE based on a percentile bootstrap with 1,000 bootstrap replicates. Table 1 presents the regression coefficients, their associated margins of error, and t values. Only gender was not statistically significant. The last two columns of Table 1 present the results of a dominance analysis to provide perspectives on the relative importance of predictors (Azen & Budescu, 2003, 2006). The dominance column reflects the average percent unique contribution to the prediction of salary across all possible subsets of predictors. The importance column in Table 1 rescales the dominance values to sum to 100 so they reflect the predictors’ relative percent contribution to the explained variance in salary. The rank ordering of predictor importance was (1) time from Ph.D., (2) number of citations, (3) number of publications, and (4) gender. The significance tests for relative importance were based on a percentile bootstrap with 1,000 replicates.

Table 1: Results of Regression Analysis

Predictor	Coefficient	t Value	Dominance	Importance
Time from Ph.D.	826.9 ±151.5	10.72*	0.19 ±0.03	39.9[a] ±6.3
Number of publications	136.8 ±151.5	5.97*	0.12 ±0.03	25.4[b] ±4.9
Gender	-852.8 ±993.0	1.69	0.01 ±0.01	2.6 ±2.6
Number of citations	178.8 ±30.8	11.41*	0.15 ±0.04	32.1[a,b] ±6.2
Intercept	33,062.4 ±1,934.3	-	-	-

(Notes: N = 750, * $p < 0.05$; For importance, predictors with common subscripts did not differ significantly from one another, $p > 0.05$)

REFERENCES

Azen, R. & Budescu, D. (2003). The dominance analysis approach for comparing predictors in multiple regression. *Psychological Methods*, 8, 129–148.

Azen, R. & Budescu, D. (2006). Comparing predictors in multivariate regression models: An extension of dominance analysis. *Journal of Educational and Behavioral Statistics*, 31, 157-180.

Curran, J. M. (2005). An introduction to Bayesian credible intervals for sampling error in DNA profiles. *Law, Probability and Risk*, 4, 115-126.