Commonality Analysis: Demonstration of an SPSS Solution for Regression Analysis

Kim Nimon, Ph. D.
University of North Texas
College of Information
3940 N. Elm, Suite G150
Denton, TX 76207
kim.nimon@unt.edu

Mariya Gavrilova
University of North Texas
College of Information
3940 N. Elm, Suite G150
Denton, TX 76207
mariya.gavrilova@unt.edu

ABSTRACT

Multiple regression is a widely used technique to study complex interrelationships among people, information, and technology. In the face of multicollinearity, researchers encounter challenges when interpreting multiple linear regression results. Although standardized function and structure coefficients provide insight into the latent variable ($\hat{y}$) produced, they fall short when researchers want to fully report regression effects. Regression commonality analysis provides a level of interpretation of regression effects that cannot be revealed by only examining function and structure coefficients. Importantly, commonality analysis provides a full accounting of regression effects which identifies the loci and effects of suppression and multicollinearity. Conducting regression commonality analysis without the aid of software is laborious and may be untenable, depending on the number of predictor variables. A software solution in SPSS is presented for the multiple regression case and demonstrated for use in evaluating predictor importance.

Categories and Subject Descriptors

G.3 [Probability and Statistics]: Correlation and regression analysis

General Terms

Theory, Measurement

Keywords

Commonality analysis, multicollinearity, suppression

1. COMMONALITY ANALYSIS

Developed in the 1960s as a method of partitioning variance (R^2) [4],[5],[6],[7], commonality analysis provides a method to determine the variance accounted for by respective predictor variables [9],[11]. Commonality analysis partitions a regression effect into unique and common effects. Unique effects identify how much variance is unique to an observed variable, and common effects identify how much variance is common to groups of variables. The number of equations required for a commonality analysis is 2^k-1 components, where k is the number of predictor variables in the regression analysis. The sum of unique and common effects equals the total variance in the dependent variable explained by the predictor variables. For a detailed discussion of commonality analysis, readers are encouraged to consult [8].

2. ILLUSTRATIVE EXAMPLE

To illustrate the benefit of commonalty analysis, an example is provided from Yao, Rice, and Wallis [12]. Yao, Rice, and Wallis examined how need for privacy (PrivNeed), self-efficacy (SelfEff), beliefs in privacy rights (PrivRight), Internet use diversity (DivUse), Internet use fluency (FlueUse) related to online privacy concerns (PrivConc). They used a two step hierarchical regression analysis where Internet use diversity and Internet use fluency were first regressed on the dependent variable. In the second step, they entered the three psychological and belief variables. Although they indicated they wanted to examine the unique effects of each of the hypothesized factors, their analyses actually indicated how much variance the three psychological and belief variables contributed to variability in concerns about online privacy after controlling for the effects of Internet use diversity and Internet use fluency. We conducted commonality analysis based on the correlation data from their study to demonstrate its analytic capability and to answer the researchers' identified research question.

3. SOFTWARE DEMONSTRATION

To perform the regression commonality analysis, we used an SPSS script that was developed based on the R code published by Nimon, Lewis, Kanes, and Hayes [8] that conducts commonality analysis for any number of predictor variables. The SPSS script file can be obtained at not cost by contacting the lead author. As depicted in Figures 1 – 4, the script file prompts the user for the: (1) SPSS data file, (2) output filename prefix, (3) dependent variable, and (4) independent variables. Due to limitations in the SPSS MATRIX command, all variables names must be eight characters or fewer.

Using the information supplied, the script generates two SPSS data files – CommonalityMatrix.sav and CCByVariable.sav where both file names are prepended with the output file name prefix. CommonalityMatrix.sav contains the unique and common commonality coefficients as well as the percent of variance in the regression effect that each coefficient contributes. The individual entries in the table can be used to determine how much variance is explained by each effect as well as which coefficients contribute most to the regression effect. CCByVariable.sav provides another view of the commonality effects. The unique effect for each of the predictors is tabularized, as well as the total of all common effects for which the predictor is involved. The last column sums the unique and common effects. Dividing the variance sum by the regression effect yields the percent of variance explained by each variable, equivalent to a squared structure coefficient. The benefit of employing commonality analysis in conjunction with the analysis of squared structure coefficients is that the researcher can determine how much variance each variable uniquely contributes and how much each shares, if any, with every other variable in the regression [8].

Based on the example, Tables 1 and 2 respectively contain the contents of YaoCommonalityMatrix.sav and YaoCCByVariable.sav. Table 3 presents an example of how the commonality effects by variable can be displayed alongside traditional multiple regression output to add another layer of consideration when evaluating the importance of predictors.

4. COMMONALITY INTERPRETATION

In Yao, Rice, and Wallis [12], the majority of the regression effect was explained by variance that was unique to belief in privacy rights (61.63%), need for privacy (14.61%), and self-efficacy (3.06%). Internet use diversity and fluency contributed little unique variance to explaining differences in online privacy concerns. In total, the four predictors uniquely accounted for 80.290% of the regression effect. The remaining 19.710 was due to variance the sets of predictors shared in common with the dependent variable. The most noticeable common effect was between need for privacy and beliefs in privacy rights, which accounted for 9.55% of the regression effect.

The commonality coefficients further indicate the presence of negative commonality coefficients. Negative commonalities occur in the presence of suppressor effects when some of the independent variables affect each other in the opposite direction [10]. While Frederick [3] indicated that negative commonalities should be interpreted as zero, others have disagreed [1], [2], [10]. Negative commonality coefficients indicate the amount of variance in the regression effect that is confounded by a set of predictor variables. In the case of suppression, negative commonality coefficients identify the increase in power associated with the suppressor effect. The commonalty data in Table 1 indicate that the regression effect was confounded by 7 out of the 15 predictor variable combinations involving self-efficacy. Suppression accounted for 3.049% of the regression effect.

The data in Table 3 demonstrate the benefits of fully reporting regression effects. In one table, researchers can simultaneously consider beta weights, structure coefficients, unique effects, and common effects when evaluating the importance of predictors. For example, while Internet use fluency might be considered an unimportant predictor due to its insignificant beta weight, its squared structure coefficient indicates that it explains a moderate amount of the regression effect. The discrepancy between the significance of the variable's beta weight and its contribution to the regression effect can easily be explained as most of its effect is due to variance that it shares in common with other predictor(s). On the other hand, the data in Table 3 demonstrates that the agreement between the relative importance of beliefs in privacy rights based on its beta weight and structure coefficient is due to the magnitude of unique variance that the variable contributes to the regression effect.

5. CONCLUSION

From a didactic perspective, commonality analysis clarifies the roles that multicollinearity and suppression play in the relationship between standardized function and squared structure coefficients. In addition, it can be observed that commonality analysis subsumes the role of computing squared structure coefficients because the portion of the regression effect explained by each variable generated from the canonical commonality analysis is identical to the squared structure coefficient generated from multiple linear regression. From a theoretical perspective, regression commonality analysis can provide important insights into variable relationships.

6. REFERENCES

[1] Beaton, A. E. 1973, March. Commonality. (ERIC Document Reproduction Service No. ED111829)

[2] Capraro, R. M., and Capraro, M. M. 2001. Commonality analysis: Understanding variance contributions to overall canonical correlation effects of attitude toward mathematics on geometry achievement. Multiple Linear Regression Viewpoints, 27(2), 16-23.

[3] Frederick, B. N. 1999. Partitioning variance in the multivariate case: A step-by-step guide to canonical commonality analysis. In B. Thompson (Ed.), Advances in social science methodology, 5, 305-318. StaMford, CT: JAI Press.

[4] Mayeske, G. W., Cohen, W. M., Wisler, C. E., Okada, T., Beaton, A. E., Proshek, J. M., Weinfeld, F. D., and Tabler, K. A. 1969. *A study of our nation's schools.* Washington, DC: U.S. Department of Health, Education, and Welfare, Office of Education.

[5] Mood, A. M. 1969. Macro-analysis of the American educational system. Operations Research, 17, 770-784.

[6] Mood, A. M. 1971. Partitioning variance in multiple regression analyses as a tool for developing learning models. American Educational Research Journal, 8, 191-202.

[7] Newton, R. G., and Spurrell, D. J. 1967. A development of multiple regression for the analysis of routine data. Applied Statistics, 16, 51-64.

[8] Nimon, K., Lewis, M., Kane, R., and Haynes, R. M. 2008. An R package to compute commonality coefficients in the multiple regression case: An introduction to the package and a practical example. Behavior Research Methods, 40, 457-466. DOI= 10.3758/BRM.40.2.457

[9] Onwuegbuzie, A. J., and Daniel, L. G. 2003, February 19. Typology of analytical and interpretational errors in quantitative and qualitative educational research. Current Issues in Educatio*n* [On-line], 6(2). Available: http://cie.ed.asu.edu/volume6/number2/

[10] Pedhazur, E. J. 1997. Multiple regression in behavioral research: Explanation and prediction (3rd ed.). Ft. Worth, TX: Harcourt Brace.

[11] Rowell, R. K. 1996. Partitioning predicted variance into constituent parts: How to conduct regression commonality analysis. Advances in Social Science Methodology, 4, 33-43.

[12] Yao, M. Z., Rice, R. E., and Wallis, K. 2007. Predicting user concerns about online privacy. Journal of the American Society for Information Science and Technology, 58, 710-722. DOI= 10.1002/asi.20530

Table 1. YaoCommonalityMatrix

Variables	Coefficient	%Total
Unique to PrivNeed	0.030	14.605
Unique to SelfEff	0.006	3.059
Unique to PrivRght	0.127	61.529
Unique to DivUse	0.001	0.444
Unique to FlueUse	0.001	0.654
Common to PrivNeed SelfEff	0.008	3.804
Common to PrivNeed PrivRght	0.020	9.555
Common to SelfEff PrivRght	-0.001	-0.390
Common to PrivNeed DivUse	0.000	0.116
Common to SelfEff DivUse	-0.001	-0.247
Common to PrivRght DivUse	0.002	0.914
Common to PrivNeed FlueUse	0.001	0.343
Common to SelfEff FlueUse	-0.001	-0.263
Common to PrivRght FlueUse	0.006	2.756
Common to DivUse FlueUse	0.001	0.505
Common to PrivNeed SelfEff PrivRght	0.002	1.029
Common to PrivNeed SelfEff DivUse	0.000	-0.139
Common to PrivNeed PrivRght DivUse	0.001	0.254
Common to SelfEff PrivRght DivUse	0.000	-0.189
Common to PrivNeed SelfEff FlueUse	0.000	-0.182
Common to PrivNeed PrivRght FlueUse	0.002	1.024
Common to SelfEff PrivRght FlueUse	-0.001	-0.334
Common to PrivNeed DivUse FlueUse	0.000	0.204
Common to SelfEff DivUse FlueUse	-0.001	-0.219
Common to PrivRght DivUse FlueUse	0.004	1.672
Common to PrivNeed SelfEff PrivRght DivUse	0.000	-0.203
Common to PrivNeed SelfEff PrivRght FlueUse	-0.001	-0.287
Common to PrivNeed SelfEff DivUse FlueUse	0.000	-0.162
Common to PrivNeed PrivRght DivUse FlueUse	0.001	0.583
Common to SelfEff PrivRght DivUse FlueUse	0.000	-0.155
Common to PrivNeed SelfEff PrivRght DivUse FlueUse	-0.001	-0.279
Total	0.207	100.000

Table 2. YaoCCbyVariable

Variable	Unique	Common	Total
PrivNeed	0.030	0.032	0.063
SelfEff	0.006	0.004	0.010
PrivRght	0.127	0.033	0.160
DivUse	0.001	0.006	0.006
FlueUse	0.001	0.011	0.012

Table 3. Regression Results for Yao, Rice, and Wallis (2007) Data Predicting Online Privacy Concerns

Predictor	*R*	R^2	R^2_{adj}	*β*	*p*	*Unique*	*Common*	*Total*	*% of R2*
	0.454	0.206	0.197						
PrivNeed				0.179	<0.001	0.030	0.032	0.063	0.303
SelfEff				-0.083	0.073	0.006	0.004	0.010	0.049
PrivRght				0.366	<.001	0.127	0.033	0.160	0.777
DivUse				0.033	0.493	0.001	0.006	0.006	0.031
FlueUse				0.040	0.406	0.001	0.011	0.012	0.059

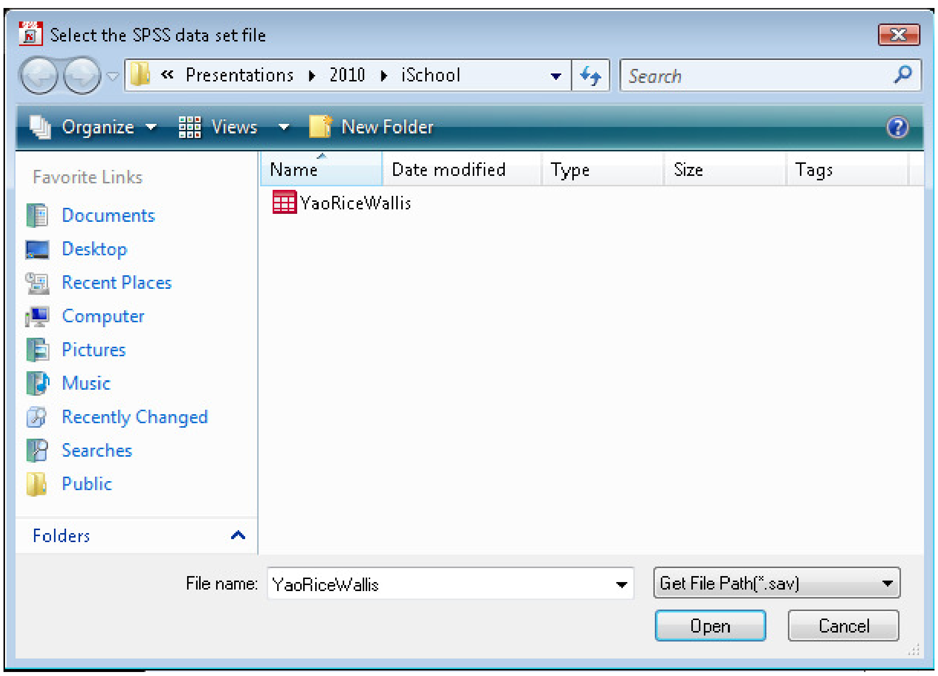

Figure 1. Screen Snapshot of Regression Commonality SPSS Script User Input – Step 1.

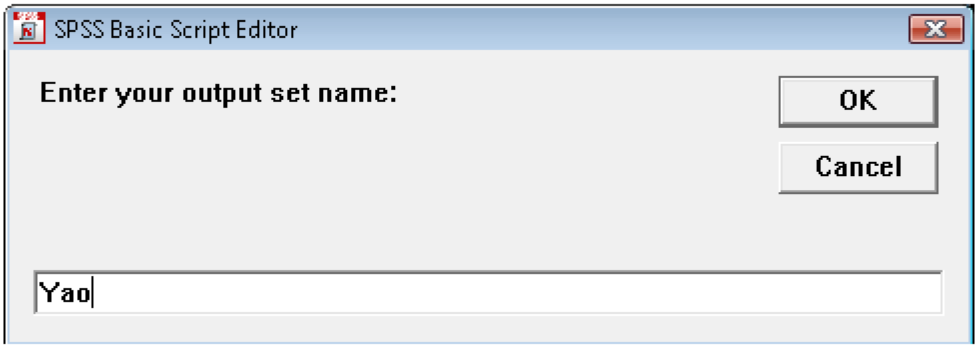

Figure 2. Screen Snapshot of Regression Commonality SPSS Script User Input – Step 2.

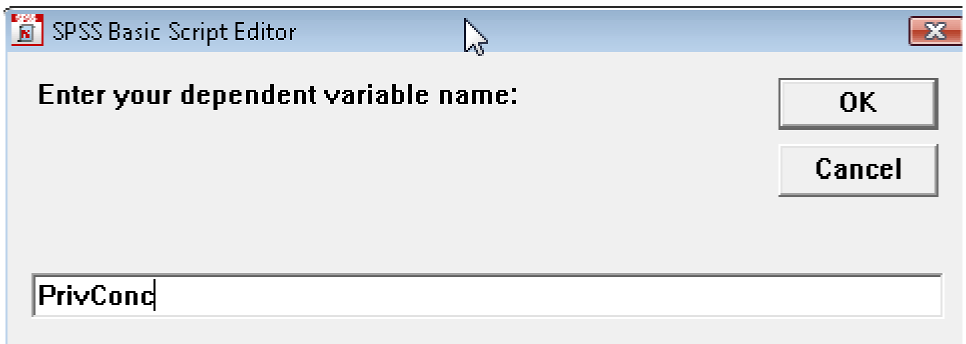

Figure 3. Screen Snapshot of Regression Commonality SPSS Script User Input – Step 3.

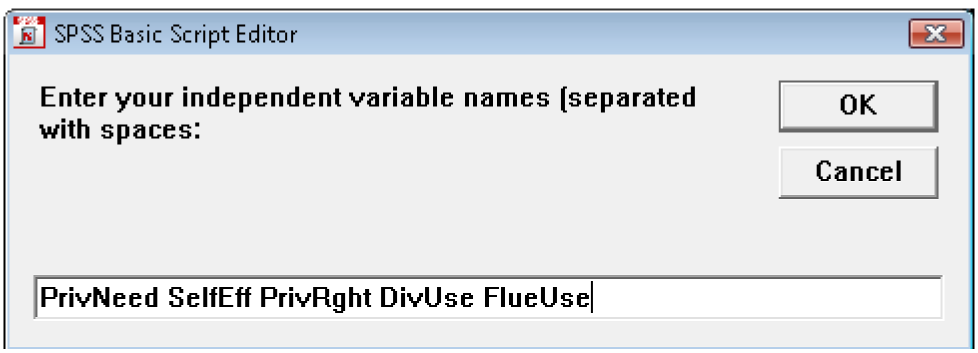

Figure 4. Screen Snapshot of Regression Commonality SPSS Script User Input – Step 4.