LETTER

Misuses of Regression and ANCOVA in Educational Research

To the Editor: Education researchers sometimes use multiple regression and analysis of covariance (ANCOVA) inappropriately. Specifically, they sometimes use standardized regression coefficients (betas) as indices of predictor importance, even though this usage is unjustified for multiple regression. Furthermore, they sometimes use ANCOVA to statistically "control for" confounds when statistical control is not achievable with ANCOVA. The goal of this letter is to alert researchers to inappropriate usage of multiple regression and ANCOVA, and to introduce them to appropriate techniques for determining variable importance and achieving statistical control.

In simple regression, beta reflects the strength of association between the predictor and the outcome variable. When using multiple predictors, beta's interpretation is more complicated. Rather than representing the amount of change in the dependent variable given a one unit change in a predictor, beta now represents the amount of change in the dependent variable given a one unit change in a predictor, assuming that all other predictors are held constant. When predictors are correlated, however, a change in any predictor will cause a change in all other predictors, so they cannot be "held constant." Partial/semi-partial correlations, and changes in R^2 also assume independent predictors, and therefore suffer from the same limitation. None of these indices accurately reflect predictor importance when using multiple correlated predictors because they treat dependent predictors as uncorrelated. Using these indices can result in gross over-estimates of a predictor's importance when it has a relatively low correlation with the other predictors but is no more correlated with the outcome than the other predictors are. This could occur if pre-pharmacy biology GPA, pre-pharmacy chemistry GPA, and PCAT score were used to predict first year pharmacy students' (P1) GPA. The three variables should have similar correlations with P1 GPA, but pre-pharmacy biology GPA and pre-pharmacy chemistry GPA should be more correlated with one another than with PCAT score. In this case, the above-mentioned methods will indicate that PCAT is a substantially stronger – perhaps several hundred times stronger – predictor of P1 GPA than the other two predictors, even though its correlation with P1 GPA is not much greater.

Failing to account for dependencies among predictors can also result in misleading significance tests for beta.

Courville and Thompson re-analyzed 11 years of multiple regression publications from the *Journal of Applied Psychology* and found that *t*-tests for beta sometimes produced erroneous results because they did not account for shared variance among multiple predictors.[1]

Rather than relying on beta, changes in R^2, or semi-partial/partial correlations to determine predictor importance, researchers should use relative importance analysis, as it produces superior estimates of correlated predictors' importance in both simulation studies[2] and primary studies.[3] Relative importance analysis transforms the original predictors into orthogonal (uncorrelated) variables that are maximally related to the original predictors and then regresses the outcome variable on these new predictors. Since the transformed predictors are uncorrelated, the resultant standardized betas are directly comparable. Researchers wishing to use relative importance analysis should visit relativeimportance.davidson.edu for a user-friendly interface. Alternatively, dominance analysis,[4] random forests,[5] or boosted/bagged regression trees[6] also produce accurate estimates of predictor importance. Researchers have several superior methods from which to choose.

Dependencies between predictors can also complicate ANCOVA. If a covariate is influenced by the treatment, ANCOVA is an inappropriate technique for reducing confounding. For example, a researcher may want to determine whether active learning improves students' information retention more than traditional lecture, while controlling for students' motivation. If active learning also increases students' motivation, ANCOVA will not be able to control for motivation, because there is no way to partition the variance in information retention that motivation and teaching strategy share into "motivation variance" and "teaching strategy variance." In this situation, motivation will control for both itself and the treatment effect. Unless the covariates are measured before treatment and un-influenced by the treatment, ANCOVA cannot establish statistical control. Inappropriate statistical control is almost always a concern when participants are not randomly assigned, because it is difficult to determine whether a pre-treatment difference resulted from random error or from a true group difference. If there is a true group difference, ANCOVA will control for both the effect of group membership and the effect of the covariate, thereby biasing the estimate of group membership's effect.[7] This problem could occur when comparing cohorts of PharmDs. Given the trend of rising matriculation despite declining applications, it is likely that each cohort will be less qualified than the last. If this is the case, an ANCOVA on the effect of cohort on NAPLEX score while controlling for PCAT score would not be able to

separate the variance in NAPLEX that cohort and PCAT share into "PCAT variance" and "cohort variance." Indeed, PCAT may be a defining feature of cohort in the same way that anxiety is a defining feature of depression.

When confronted with non-random assignment, researchers should use propensity score matching,[8] propensity score weighting,[8] or doubly robust estimation[9] to address confounding. In propensity score matching, logistic regression or boosted/bagged classification trees are used to estimate each case's probability of being in the treatment group given a set of covariates. This probability estimate is called a propensity score. Treatment cases are then compared to control cases with similar propensity scores to estimate the treatment effect. In this way, each treatment case is compared to a control case (or control cases) that is as similar to it as possible in terms of the covariates, thereby reducing confounding. Cases can also be weighted by their propensity scores in order to achieve statistical control. Finally, researchers can use doubly robust estimation in which a multiple regression equation containing the treatment and the covariates is applied to the propensity score weighted data to adjust for any covariate imbalance that is missed by the weighting.

While techniques like doubly-robust estimation, propensity score weighting/matching, and relative importance analysis sometimes require extra time and specialized software, it is important that researchers use them when appropriate. Inappropriate usage of more common techniques may lead to unjustified conclusions that harm or fail to help students.

Samuel C. Karpen, PhD
Bill Gatton College of Pharmacy, East Tennessee State University, Johnson City, TN

REFERENCES

1. Courville T, Thompson B. Use of structure coefficients in published multiple regression articles: β is not enough. *Educ Psychol Meas.* 2001;61(2):229-248.
2. LeBreton JM, Ployhart RE, Ladd RT. A Monte Carlo comparison of relative importance methodologies. *Organ Res Methods.* 2004; 7(3):258-282.
3. LeBreton JM, Hargis MB, Griepentrog B, Oswald FL, Ployhart RE. A multidimensional approach for evaluating variables in organizational research and practice. *Pers Psychol.* 2007;60(2):475-498.
4. Azen R, Budescu DV. The dominance analysis approach for comparing predictors in multiple regression. *Psychol Methods.* 2003;82(2):129-148.
5. Breiman L. Random forests. *Machine Learn.* 2001;45(1):5-32.
6. Dietterich TG. An experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting, and randomization. *Machine Learn.* 2000;40(2):139-157.
7. Miller GA, Chapman JP. Misunderstanding analysis of covariance. *J Abnorm Psychol.* 2001;110(1):40-48.
8. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behav Res.* 2011;46(3):399-424.
9. Funk MJ, Westreich D, Wiesen C, Stürmer T, Brookhart MA, Davidian M. Doubly robust estimation of causal effects. *Am J Epidemiol.* 2011;173(7):761-767.