Available online at www.sciencedirect.com

ScienceDirect

Health Professions Education ▮ (▮▮▮▮) ▮▮▮–▮▮▮

HEALTH PROFESSIONS EDUCATION

www.elsevier.com/locate/hpe

Analysis of Covariance (ANCOVA) vs. Moderated Regression (MODREG): Why the Interaction Matters

Jimmie Leppink[1]

School of Health Professions Education, Maastricht University, PO Box 616, 6200 MD Maastricht, The Netherlands

Received 31 January 2018; received in revised form 17 April 2018; accepted 18 April 2018

Abstract

Analysis of covariance (ANCOVA) is a commonly used statistical method in experimental and quasi-experimental studies. One of the fundamental assumptions underlying ANCOVA is that of no interaction between factor and covariate. Unfortunately, many researchers report the outcomes of ANCOVA but not the outcomes of a check on that non-interaction assumption. Through a comparison of ANCOVA (which assumes non-interaction) and moderated regression (MODREG, which allows for interaction) in a worked example, this article demonstrates that omitting the check of the non-interaction assumption comes at the risk of misestimating a treatment effect or other group difference of interest. If there is substantial interaction between factor and covariate, ANCOVA will result in conclusions of there being a group difference or no group difference whereas MODREG indicates that the magnitude of a group difference depends on the level of the covariate. Therefore, this article advises to first check and report on the interaction, to use that check to decide whether a model without interaction (ANCOVA) or with interaction (MODREG) is to be preferred, and to use ANCOVA only if the criteria outlined in this article indicate a preference towards the model without interaction. Moreover, omitted terms, such as the omitted interaction if one proceeds with ANCOVA, should be reported as well.

Keywords: Analysis of covariance; Moderated regression; Interaction; Treatment effects

1. Introduction

Analysis of covariance (ANCOVA)[1] is a widely used statistical method for analyzing quantitative data from experimental and quasi-experimental studies in a variety of fields, including education and psychology. For example, while in experimental studies the interest usually lies in differences between experimental treatment and control conditions (i.e., treatment effects), including a meaningful covariate – such as prior knowledge of a subject or even pretest performance in a study on the effects of instructional methods on learning outcomes as measured through posttest performance – may increase statistical power for group differences of interest.[2,3] Apart from being meaningful, to reduce the likelihood of obtaining findings that have no meaning beyond the study in question, expectations with regard to the effects of covariates should be formulated *before not after seeing the data*.[4] However, like any statistical method, ANCOVA is based on

E-mail address: jimmie.leppink@maastrichtuniversity.nl

[1]Jimmie Leppink, PhD, is currently Assistant Professor of Methodology and Statistics and Data Manager at the School of Health Professions Education, Maastricht University, the Netherlands.

Peer review under responsibility of AMEEMR: the Association for Medical Education in the Eastern Mediterranean Region

https://doi.org/10.1016/j.hpe.2018.04.001

assumptions,[1,5] and violations of these assumptions may have serious consequences for the outcomes and interpretations.

1.1. Assumptions

As elegantly formulated by Huitema,[1] the assumptions for the ANCOVA model "are relatively straightforward because it is simply another linear model" (p. 182).

Firstly, the residuals are assumed to be *independent*. Although this assumption may be realistic for instance in randomized controlled experiments where participants receive individual treatment and – throughout the experiment – do not interact in any way with other participants, when participants interact (e.g., group learning) or are measured repeatedly on the same variable(s) of interest (i.e., repeated measures) that assumption is usually violated. Interaction between participants and repeated measurements from the same participants are two phenomena that usually create some kind of a *dependence* of residuals and that dependence needs to be accounted for in the statistical analysis, for instance through multilevel analysis.[6,7]

Secondly, the residuals are assumed to have a *mean of zero* regardless of the grouping variable or the level of the covariate. This can be considered true when the relation between response variable and covariate within groups is *linear*, and in cases where the latter is not true, researchers should consider nonlinear alternatives to the linear model (i.e., it is possible to have means of zero for residual distributions when the relation between response variable and covariate is nonlinear and an appropriate nonlinear function is used).[1]

Thirdly, the residuals are assumed to be *normally distributed*. Inspecting the plotted residuals of the ANCOVA model and/or the normal probability plot provides a straightforward approach to checking this assumption.[1] Although moderate departures from normality in samples in the 20 s or larger generally do not constitute a cause of concern, more severe departures from normality may introduce substantial distortion [8] and hence need to be accounted for, for instance by using a model that allows for another type of distribution.[1]

Fourthly, the variance of the residuals is the same regardless of the grouping variable or covariate (i.e., *homoscedasticity*). Two common types of deviation from that assumption are (1) increasing (or decreasing) residual variance with increases in the level of the covariate but no difference between groups and (2) constant residual variance within but not between groups. The first type of deviation appears to not meaningfully affect the outcomes of ANCOVA, whereas the second type of deviation is mainly problematic when dealing with groups that differ in sample size (with larger differences being more problematic).[9] Huitema[1] discusses several alternatives to ANCOVA for such situations.

Fifthly, the grouping variable and covariate are assumed to be *fixed* and measured without error. For the grouping variable, this is straightforward. Whenever the interest lies in a comparison between specific groups, such as treatment conditions in a randomized controlled experiment, the categories of the grouping variable are fixed. The comparison is clear and there is no interest in generalizing to other groups not observed in the study. However, in cases where groups under comparison can be considered a *random* sample of a population of possible groups and the interest lies in generalizing the findings of the groups observed to other groups, the groups are in fact treated as *random* not fixed. An example of the latter is found in a random sample of say twenty health centers from a much larger population of health centers. In the latter case, ANCOVA does not work for it does not enable generalization to groups not observed; a multilevel model that treats the groups as random units in which individuals of interest (e.g., employees, patients) are nested constitutes a better approach.[6] That said, for the covariate, the situation is more complex. Assuming random sampling, covariate values observed in a sample in practice rarely cover all values of the covariate in the population but rather constitute a *random* sample of covariate values in the population. Moreover, given that in educational and psychological settings the covariate often results from a psychometric instrument, the covariate is often measured *with error*. Although ANCOVA was initially derived under the assumption of the covariate being fixed, ANCOVA with a random variable covariate measured without error is appropriate. [10] For instance, from a limited number of values observed on the covariate, researchers may use a linear model to generalize to values of the covariate that have not been observed in the sample but are within the range (i.e., between minimum and maximum) of values observed in the sample. However, the measurement error issue is more serious: "In experimental research, unreliable covariates lead to loss of power and a conservative statistical test through underadjustment of the error term"[11] (p.326). Besides, when ANCOVA is used to adjust means in nonrandomized studies, "the difference between the adjusted means is partly a function of the reliability of the

Please cite this article as: Leppink J. Analysis of Covariance (ANCOVA) vs. Moderated Regression (MODREG): Why the Interaction Matters. *Health Professions Education* (2018), https://doi.org/10.1016/j.hpe.2018.04.001