Journal of Memory and Language 72 (2014) 37–48

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

What residualizing predictors in regression analyses does (and what it does *not* do)[☆]

Lee H. Wurm[a,*], Sebastiano A. Fisicaro[b]

[a] Department of Psychology and Program in Linguistics, Wayne State University, Detroit, MI 48202, USA
[b] Department of Psychology, Wayne State University, Detroit, MI 48202, USA

ARTICLE INFO

Article history:
Received 24 May 2013
revision received 12 December 2013
Available online 23 January 2014

Keywords:
Regression analysis
Mixed-effects modeling
Statistical analysis
Collinearity
Residualization
Orthogonalization

ABSTRACT

Psycholinguists are making increasing use of regression analyses and mixed-effects modeling. In an attempt to deal with concerns about collinearity, a number of researchers orthogonalize predictor variables by residualizing (i.e., by regressing one predictor onto another, and using the residuals as a stand-in for the original predictor). In the current study, the effects of residualizing predictor variables are demonstrated and discussed using ordinary least-squares regression and mixed-effects models. Some of these effects are almost certainly not what the researcher intended and are probably highly undesirable. Most importantly, what residualizing does *not* do is change the result for the residualized variable, which many researchers probably will find surprising. Further, some analyses with residualized variables cannot be meaningfully interpreted. Hence, residualizing is not a useful remedy for collinearity.

Introduction

In psycholinguistics there has been a move toward regression studies, which offer several advantages over traditional factorial designs. Baayen, Wurm, and Aycock (2007), for example, used mixed-effects modeling[1] to examine auditory and visual lexical decision and naming times. They found a number of curvilinear effects that are difficult to detect with factorial designs. Even more interesting, the authors found sequential dependencies in the response times, such that response latency on a given trial could be predicted by latencies on the previous four trials. This sequential dependency, which cannot be assessed in a factorial design, ultimately exhibited more explanatory power than nearly all of the other predictors that were examined.

A second advantage of regression designs is pragmatic. With the increased complexity of many theoretical models, it becomes impractical to isolate a difference on one predictor while adequately equating stimulus materials on the growing number of other variables known to affect psycholinguistic processing. Baayen et al. (2007) examined 18 predictor variables. The influential megastudy of Balota, Cortese, Sergent-Marshall, Spieler, and Yap (2004) examined 19. A factorial design matching on all but one or two of the variables in situations like these is virtually inconceivable, and so a large number of potentially interesting studies simply could not be done. The Balota et al. (2004) study is interesting for the additional reason that they included as stimuli virtually all single-syllable monomorphemic words in English. An exhaustive study such as this cannot be done in a factorial manner, because the words in the language are naturally correlated on a number of variables of theoretical interest.

☆ Portions of this study were presented at the 54th annual meeting of the Psychonomic Society in Toronto, Ontario (November 14–17, 2013).

* Corresponding author. Address: Department of Psychology and Program in Linguistics, Wayne State University, 5057 Woodward, 7th floor, Detroit, MI 48202, USA. Fax: +1 313 577 7636.

E-mail address: lee.wurm@wayne.edu (L.H. Wurm).

[1] The distinction between ordinary least-squares regression and the kind of mixed-effects model described by Baayen, Davidson, and Bates (2008) is unimportant for the current study. Both techniques are used here, and the same findings and conclusions hold.

0749-596X/$ - see front matter
http://dx.doi.org/10.1016/j.jml.2013.12.003

Many researchers express concern about the extent to which these natural correlations between predictors might lead to collinearity and computational problems. For example, Tabachnick and Fidell (2007) assert that, with predictor intercorrelations of .90 and above, there are statistical difficulties in the precision of estimation of regression coefficients (citing Fox, 1991). Further, Cohen, Cohen, West, and Aiken (2003) state that the estimates of the coefficients will be "very unreliable" and "of little or no use" (p. 390). In addition, Darlington (1990) emphasizes the loss of statistical power of tests on the individual regression slopes.

However, Friedman and Wall (2005) assert and demonstrate that improvements in algorithms and computer accuracy have eliminated the computational difficulties. The current study lends additional support to their claim. Further, Friedman and Wall (2005), along with others, also note that collinearity *per se* is not necessarily bad. For example, if a researcher's goal is simply to maximize explained variance, collinearity can be ignored (Darlington, 1990; Tabachnick & Fidell, 2007). The goal of most psycholinguistic applications of regression, though, is to evaluate the effects of several of the individual predictor variables. The potential interpretational problems caused by collinearity here can be thorny, even if the computational problems are not.

Because of concerns like this, some researchers have attempted to deal with collinearity by residualizing one of the correlated predictor variables. To do this, one runs a preliminary regression analysis using one of the predictor variables to predict the other (e.g., using X_2 to predict X_1). The residuals from this analysis constitute a new predictor variable, $X_{1\text{resid}}$, that is used in subsequent analyses in lieu of X_1. $X_{1\text{resid}}$ is guaranteed to be uncorrelated with X_2, providing an apparent solution to the problem of collinearity. Thus, residualizing seems like a useful and appropriate technique.

Psycholinguists have offered several justifications for residualizing. A review of some of those justifications is instructive, as it illustrates a considerable range of beliefs, some erroneous, about what residualizing accomplishes[2]:

> "To avoid problems with increased multicollinearity, we included the residuals…in our mixed-effects model…These residuals are thus corrected for the influence of all variables correlated with the original familiarity and meaningfulness measures" (Lemhöfer et al., 2008, p. 23)

> To dissociate the effect of one predictor from another and demonstrate that the effect of one predictor does not explain the effect of the other (Green, Kraemer, Fugelsang, Gray, & Dunbar, 2012, pp. 267–268)

> To help rule out the possibility that the effect of one predictor masks the effect of another (Kuperman, Bertram, & Baayen, 2010, p. 89)

> "…to assess the effect of " [a predictor] (Jaeger, 2010, p. 33)

> "…to ensure a true effect of" [a predictor] (Cohen-Goldberg, 2012, pp. 191–192)

> "…to allow for assessment of the respective contributions of each predictor" (Ambridge, Pine, & Rowland, 2012, p. 267)

> "…to determine the unique contribution of" [a predictor] (Cohen-Goldberg, 2012, p. 188)

> To provide "…a reliable estimate of the unique variance explained by each" [predictor] (Ambridge et al., 2012, p. 268)

> To pit predictors against one another and determine whether one explains variance that the other cannot (Ambridge et al., 2012, p. 268)

> "…to reliably assess effect *directions* for collinear predictors" and to be able to simultaneously assess "…the independent effects of multiple hypothesized mechanisms" (Jaeger, 2010, p. 30; emphasis in original)

> to test the effect of one predictor beyond the properties of two other predictors (Jaeger, 2010, p. 33)

> "Orthogonalisation of such variables is crucial for the accuracy of predictions of multiple regression models. Teasing collinear variables apart is also advisable for analytical clarity, as it affords better assessment of the independent contributions of predictors to the model's estimate of the dependent variable" (Kuperman, Bertram, & Baayen, 2008, p. 1098).

Most researchers do not specify precisely what would trigger the strategy. Cohen-Goldberg (2012) said it was done when a predictor "…was collinear with one or more control variables…" (p. 188). Jaeger and Snider (2013) did it "since the two predictor variables were correlated" (p. 63). Kahn and Arnold (2012) residualized "Because of high correlations" between the predictor variables (p. 317). This last case is interesting for the additional fact that the residualization was restricted to variables that were included only for purposes of statistical control. The individual effects of these variables were not of interest – the goal was simply to be able to assure readers that the analysis had controlled for them. Below, we show that residualizing accomplishes literally nothing in this case. Further, examination of the cut-off values that are reported reveals a lack of consensus about when one should residualize: Kuperman et al. (2008) residualized whenever a zero-order correlation between predictors exceeded 0.50, whereas Bürki and Gaskell (2012) used 0.30 as a cut-off.

Use of this strategy in psycholinguistics is a relatively recent phenomenon. The earliest example we have identified is Baayen, Feldman, and Schreuder (2006). The scope of what Baayen et al. (2006) did was restricted, and the reasons for it were principled and clearly articulated. They wanted to determine if a subjectively-rated version of word frequency offered anything beyond various objective measures. They partialed the objective measures from the subjective measure, and added the residuals to a model they had already specified as more or less complete. They did mention collinearity in this context, but it was not their

[2] One reviewer wondered if perhaps researchers were guilty of imprecise writing, rather than misunderstanding residualization. Evidence is presented later that there is genuine misunderstanding in at least some of these cases.