**Multiple Regression with SAS[[1]](#footnote-1)©**

 You have already done a trivariate (two predictors, one criterion) multiple regression with SAS (back in PSYC 6430), let us now try a multiple regression with four predictors.

 Please refer to “[Presenting the Results of a Multiple Regression Analysis](http://core.ecu.edu/psyc/wuenschk/MV/multReg/MultReg-WriteUp.pdf),” in which I presented results from research that resulted in a model for predicting graduate students’ Grade Point Average from four predictor variables (GPA, GREQ, GREV, MAT, and AR). The data from this research are in the file "Multreg.dat" on my [StatData page](http://core.ecu.edu/psyc/wuenschk/StatData/StatData.htm), and the program is the file "MReg.sas' on my [SAS programs page](http://core.ecu.edu/psyc/wuenschk/SAS/SAS-Programs.htm). Download both and run the program.

Look at the output. Remember that the slopes are in raw scores units, so it is dangerous to compare slopes of variables with different metrics. For example, when you look at the slope for GRE\_Q, .004, you might conclude that it has only a tiny effect -- but that slope is in GPA points increase per one point increase in GRE\_Q. A one point increase in GPA is an enormous increase, a one point increase in GRE\_Q is a tiny increase. When you want to compare slopes, do it with the standardized slopes, where the units are standard deviation increase in criterion per one standard deviation increase in predictor. When you look at the standardized slopes, you see that GRE\_Q has the greatest unique contribution (***β*** = .324) of all predictors.

Note that the single best predictor is AR, but that AR has the smallest unique contribution in the context of the other predictors. If you look at the intercorrelation matrix you will see why this is so: AR is highly correlated with each of the other predictors, so, when the other predictors are included in the model, AR becomes redundant.

 The output statement is used to create a data set called "hats." This data set will include all of the variables in the model statement plus the one variable we created, GPA\_hat, which is GPA scores predicted from our model. "Hat" refers to the caret symbol, ^, which we use to designate an estimator, and which looks like a hat.

 Note that when we correlated GPA with GPA\_hat we obtained the multiple correlation coefficient, ***R***.

**Multicollinearity**

 Multicollinearity exists when a predictor can be nearly perfectly predict from a weighted linear combination of the other predictors – that is, when is very large. In that case, the partial statistics would be unstable, that is, they would tend to vary wildly among samples drawn from the same population. The usual solution here is to drop variables from the model to eliminate the problem with multicollinearity. The **tolerance** statistic is computed as . So multicollinearity is present when a tolerance is very low. The **variance inflation factor (VIF)** is computed as 1/tolerance, so high values of VIF indicate a problem. So, how high must VIF get before we get worried? Some say 10, some say 5, and a few say 2.5.

**Singular Intercorrelation Matrix**

 As you should recall from [a previous lesson](http://core.ecu.edu/psyc/wuenschk/MV/multReg/MultReg-Matrix.docx), to conduct a multiple regression you need to invert the matrix of correlations among the predictor variables. If any one of those variables can be perfectly predicted from the others, then it is impossible to invert that matrix, and the multiple regression analysis crashes. For an example of how this can happen, see [Suggestion of Moron Reviewer](http://core.ecu.edu/psyc/wuenschk/SAS/MReg_Singular.docx).

**Effect Size Estimates**

 Both partial and semipartial effect size estimates are produced from the SAS program here. I’ll be talking more about these in the near future.

Links

* [Output for This Lesson](http://core.ecu.edu/psyc/wuenschk/SAS/MReg-Output.docx)
* [More Lessons on Multiple Regression](http://core.ecu.edu/psyc/wuenschk/StatsLessons.htm#MultReg)
* [Return to My SAS Lessons Page](http://core.ecu.edu/psyc/wuenschk/SAS/SAS-MV.htm)
* [Using Proc GLM to Put Confidence Intervals Around Effect Size Estimates](file:///C%3A%5CUsers%5CVati%5CDocuments%5C_XYZZY%5C_Stats%5CSAS%5CMultreg_Confidence-Intervals.htm)
* [When Can You Safely Ignore Multicollinearity?](http://www.statisticalhorizons.com/multicollinearity)
* [Variance Inflation Factor (VIF)](http://www.how2stats.net/2011/09/variance-inflation-factor-vif.html)

[Karl L. Wuensch](http://core.ecu.edu/psyc/WuenschK/KLW.htm), Dept. of Psychology, East Carolina University, Greenville, NC.
January, 2020.

1. © Copyright 2019, Karl L. Wuensch - All rights reserved. [↑](#footnote-ref-1)