ALLERGY, SOFTWARE

# FORGET ABOUT MULTIPLE REGRESSION ANALYSIS

26.01.2016

When starting in epidemiology I had  only high school math skills. Nevertheless, I could usually find major associations by simple tables and plots. Then I learned about multiple regression analysis and used it in numerous research papers. Nevertheless I soon discovered that

> The results are often somewhere between meaningless and quite damaging.

whenever basic principles had been violated. Models have been poorly adjusted and results always over-interpreted. As described before

> Multicollinearity between explanatory variables should always be checked using variance inflation factors and/or matrix correlation plots. Despite the fact that automated stepwise procedures for fitting multiple regression were discredited years ago, they are still widely used and continue to produce overfitted models containing various spurious variables.
> A key issue seldom considered in depth is that of choice of explanatory variables (i.e., if the data does not exist, it might be better to actually gather some).
> Typically, the quality of a particular method of extrapolation is limited by the assumptions about the regression function made by the method.

So I feel some support now by reading on "The Crusade Against Multiple Regression Analysis" by the eminent Richard Nisbett.

I hope that in the future, if I’m successful in communicating with people about this, that there’ll be a kind of upfront warning in New York Times articles: These data are based on multiple regression analysis. This would be a sign that you probably shouldn’t read the article because you’re quite likely to get non-information or misinformation.

CC-BY-NC Science Surf accessed 15.02.2026