

Predicting the Quality of Your Statistical Regression Models

Bruce Ratner, Ph.D.

If you would like a *prediction of the quality* of your statistical linear regression models, then [email](#) me your answers to the following questions, related to the essential characteristics of a good model. I will reply with my prediction of the quality of your model. [1]

1. Your typical model has predictor variables:
 - a. less than 6
 - b. 6 to 12
 - c. 13 to 20 or
 - d. greater than 20 predictor variables
2. You attempt to transform predictor variables (e.g., $1/x$, $\log x$):
 - a. always
 - b. sometimes, or
 - c. never
3. Your method of variable selection: [2]
 - a. forward selection
 - b. backward elimination
 - c. stepwise, or
 - d. other, please specify
4. i) How do you determine your final model? Namely, you have narrowed the number, say, 2 – 4 of candidate models, from which to pick the best model. What *criteria* do you use to declare the best model among the 2 – 4 models?
 - ii) Do you include in declaring the best model testing for a linear-relationship between dependent variable and each predictor variable in the model?
5. How do you validate the best model chosen in question #4?

1– Bruce Ratner's Pithy Bio

I am founder and President of DM STAT-1 Consulting. I have made the company the ensample for Statistical Modeling & Analysis and Data Mining in the DM Industry. I have a Ph.D. in mathematical statistics with prime (17) years of experience in consulting for virtually [all sectors of the DM Industry](#). I have built virtually [every type of model](#), statistical and non-statistical (machine learning) with the inseeing digging of data with my data mining tool (GenIQ Model). I have published two books on database marketing statistics and the reigning top-seller since its release in 2003, [Statistical Modeling and Analysis for Database Marketing: Effective Techniques for Mining Big Data](#). As well, I have published over 100 peer-reviewed articles for national and international journals. I am an often-invited speaker at public and private industry events.

2 – Must Reading – [Variable Selection in Regression: Few Statisticians Know They Produce Poor Models](#)