

Nuts and Bolts

Stepwise Regression

Inés M. Anchondo DrPH, RD, LD, CSP, MPH



Stepwise modeling is a multiple regression procedure to analyze the association between a dependent variable (Y) and large set of independent variables (Xs). The point is to choose the best independent set of variables to create a model that has the best predictive ability. The model that explains 100% of the variance.

The step by step procedure for stepwise modeling involves:

- Identify an initial model that includes the regression intercept (unless there is no intercept).
- Enter or remove, depending on the method forward or backward, one independent variable (X) at a time.
- Evaluate if the independent variable (X) entered or removed is the best predictor variable e.g.

statistically significant and with the highest R^2 (explaining more of the variance in the dependent variable (Y)).

There are two stepwise model building methods: *forward* and *backward*. In the *forward* method the initial model includes only the intercept. Then, one by one, each independent variable is entered into the model. The independent variable that has the highest R^2 and it is statistically significant ($p = <0.25$)¹ is considered for inclusion in the model. This stops when none of the remaining variables are significant. In the *backward* method the initial model includes all the independent variables. Then, one by one, each variable is removed from the model. The variables that remain in the model are the ones with the highest R^2 and desired statistical significance level.

It is important to keep in mind that stepwise regression is an 'exploratory tool to identify potentially important predictors'.² To create the best model it is necessary to validate the model by testing it with 'new' sample data (data that was held from the initial sample). Validating the model involves applying the model to the held-out sample data to see what the model predictive accuracy is.³

References

1. Mickey RM, Greenland S. The impact of confounder selection criteria on effect estimation. *Am J Epidemiol.* 1989;129(1):125-37.
2. New View of Statistics: Multiple Linear Regression. Stepwise Regression. <http://www.sportisci.org/resource/stats/multiple.html> Published 2010. Accessed July 22, 2010.
3. S-news. Washington University in St. Louis listserv. Stepwise regression. <http://www.biostat.wustl.edu/archives/html/s-news/2000-11/msg00184.html>. Published November 18, 2000. Accessed July 22, 2010.

Member Spotlight

continued from page 13

How has your affiliation with ADA impacted your career progression?

I first joined ADA when I was a dietetic student, and I have continued and valued my membership ever since. ADA brings together an incredible diversity of information and resources that support registered dietitians in all kinds of careers and through all phases of their careers. I especially appreciate the opportunities through ADA for networking, collaboration, and lifelong learning.

If someone were to ask you to explain "why research is important to the field of dietetics" what would you say?

Everything we do is based on research – the recommendations we make to the public, the advice we give clients and patients, the practice guidelines we follow. Dietitians are in a great position to identify important questions that need to be addressed, and then to conduct studies that will add to what we know.