



## Stats Happening - May 2013

1. **Mediation Modeling Tools**
2. **Missing Data with SAS**
3. **SAS Variable Selection**
4. **Specification Error Checking in Stata**
5. **Estimation of Nonlinear Models**
6. **Significance: Statistics Making Sense**

### 1. Mediation Modeling Tools

#### PROCESS

A new and user-friendly tool has just come out that might prove useful for researchers using path analysis in **SPSS** or **SAS**. If you have been using mediation you might have been challenged by having several mediators or by having to incorporate moderation (interactions) in your mediation analysis. Mediated-moderation or moderated-mediation, unified under the term “conditional process modeling”, as well as models with several mediators either in parallel or in sequence are now much easier to assess using Andrew Hayes’ new tool “PROCESS”. PROCESS is a freely available SPSS or SAS macro that will fit a vast variety of models using OLS for continuous responses or maximum likelihood for dichotomous responses. Bootstrap confidence intervals for conditional and unconditional indirect effects as well as heteroscedasticity-consistent standard errors are available for inference about paths coefficients. Due to come out this month is Andrew Hayes’ latest book titled: “Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach”. His book will give an overview of mediation and moderation and their analytical integration in the form of “conditional process analysis”. For further information see:

<http://www.afhayes.com/introduction-to-mediation-moderation-and-conditional-process-analysis.html#process>

<http://afhayes.com/introduction-to-mediation-moderation-and-conditional-process-analysis.html>

#### The Mediation and Lavaan Packages in R

The "mediation" package in **R** provides a powerful tool for causal mediation analysis. It supports a wide range of models for fitting the mediator and the outcome, and can compute confidence intervals for mediation effects (direct and indirect) via bootstrapping. The “lavaan” (latent variable analysis) R package is another useful tool for testing for mediation in the context of a structural equation modeling

(SEM) framework. The package also fits a wide variety of other latent variable models including confirmatory factor analysis and latent growth curve models.

## **2. Missing Data with SAS**

SAS now offers additional choices to deal with missing data. The updated PROC CALIS allows estimation using the full information maximum likelihood method. This method allows estimation with missing values avoiding the deletion of incomplete observations. In many situations this offers an excellent alternative to multiple imputation. In addition the multiple imputation procedure (PROC MI) now offers fully conditional specification methods for multiple imputation. The FCS statement specifies a multivariate imputation by fully conditional specification (FCS) methods also called chained equation multiple imputation.

## **3. SAS Variable Selection**

If you are faced with a variable selection problem in a general linear model framework we recommend you look at the PROC GLMSELECT procedure in SAS. This procedure offers several important advantages over the variable selection procedures offered in PROC REG. It will not only handle categorical variables and interaction terms but also provides a wide variety of methods and fit criteria such as the LASSO and LAR method. All this is accompanied by useful tabular and graphical display of selection search.

## **4. Specification Error Checking in Stata**

One assumption we make in regression analysis is that we have included all of the relevant predictor variables in our model. When we have not, we call this a specification error. This can happen in several ways, e.g. a missing interaction term, or a missing higher order polynomial of a predictor already in the model. Stata has tools to help identify and remedy such specification errors. The 'linktest' command checks if there are any additional predictors that are statistically significant beyond chance, including interaction terms. The 'boxtid' command (short for Box-Tidwell model) tests for higher order polynomials using power transformations. Note that these tools for detecting specification errors are applicable to linear regression models, but also many other statistical modeling techniques, e.g. logistic regression, Cox regression.

## **5. Estimation of Nonlinear Models**

If you need to estimate nonlinear regression models, the JMP software might just be the tool you were looking for. The flexibility and the breadth of application of the nonlinear command in JMP are to be noted. Not only does it offer a library containing many often used equations but it will allow you to estimate any custom made nonlinear equations. Several estimation methods are available. The user is also offered the ability to provide appropriate starting values or to change the parameter estimates using a very practical sliding bar that interactively shows how the changes of the parameter influence the shape of the curve. For more information

see: [http://www.jmp.com/support/help/Fit\\_a\\_Custom\\_Model.shtml](http://www.jmp.com/support/help/Fit_a_Custom_Model.shtml)

## 6. Significance: Statistics Making Sense

*Significance* is a bimonthly magazine whose major content consists of articles on topics of statistical interest presented at a level suited to a general audience. The articles are accessible, thought-provoking and entertaining. It presents a statistical view of what is going on in the world in an understandable and engaging way. It is not a research journal and articles are not peer reviewed. Check out the most recent issue. We liked the stories “Playing with data” and “Are you sitting comfortably?” and the challenge to readers “Phoney data”. The Cornell community can access *Significance* for free through Wiley online while on campus:

[http://onlinelibrary.wiley.com/journal/10.1111/\(ISSN\)1740-9713](http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1740-9713)