



Stats Happening - November 2011

- 1. New workshop offered this week: “Overview of Multiple Comparison Methods”.**
- 2. Estimating, comparing and testing specific effects in Structural Equation Modeling using phantom models.**
- 3. Linear Mixed models and Generalized Linear Mixed models.**
- 4. Reading proprietary data formats into R from other statistical tools.**
- 5. Fisher's combined probability test.**
- 6. User-friendly power and sample size calculations using SAS.**
- 7. Many useful upgrades with the latest version 12 of STATA.**

1. New workshop offered this week: “Overview of Multiple Comparison Methods”.

There is still time to sign up for the new workshop offered this week: “Overview of Multiple Comparison Methods”. In this workshop we will demonstrate when the problem of multiple comparisons occurs, and how inference regarding the Type I error is affected. We will briefly describe the traditional methods for correcting for multiple testing. We will then describe modern methods which are more suitable for large-scale experiments. This workshop will be offered Wednesday November 16, 12:30 PM - 2:00 PM in Savage Hall 100. For more information or to register, see our web site at

<http://www.cscu.cornell.edu/workshops/registration.php>

2. Estimating, comparing and testing specific effects in Structural Equation Modeling using phantom models.

When working with SEM models you might have encountered situations where your SEM software would not allow you to assess and compare specific effects. For example, when the effect of a variable on an outcome variable is mediated by several groups of variable it can be challenging to decompose and compare the specific indirect effects. Several approaches have been proposed to address this type of problem but they all have limitations. A new approach presented by Siegfried Macho and Thomas Ledermann offers several advantages: it is easy and flexible to use even with complex models, it can be implemented with all SEM software, and it allows bootstrapping of confidence intervals for specific effects. The following article will give you all the details as well as excellent guidance if you wish to implement the phantom model approach: Estimating, Testing, and Comparing Specific effects in structural equation models: The phantom model approach. *Psychological Methods*, 16, 34-43. (<http://www.unifr.ch/psycho/site/units/allpsy/team/Macho/pub>).

3. Linear Mixed models and Generalized Linear Mixed models.

Most of us have many unanswered questions when it comes to Linear Mixed Models and Generalized Linear Mixed Models. The following page might help you find some of the answers you have been looking for. Although the site is mostly intended for ecologists and R users, other researchers might find the explanations and the numerous references also very useful. See: <http://glmm.wikidot.com/faq>

4. Reading proprietary data formats into R from other statistical tools.

When collaborating with people who use different statistical software, or when importing files generated by others, it is often the case that the data is stored in a proprietary format, and without the corresponding software it is not possible to read and use the data. The "foreign" library in R contains useful tools to read data from other statistical systems, including STATA, SPSS, and SAS, or save data in these formats. For instance, to read a file saved by an SPSS user, use the function `read.spss()`. For more information, see the user manual at <http://cran.r-project.org/doc/manuals/R-data.html#Importing-from-other-statistical-systems>

5. Fisher's combined probability test.

This test is used to combine the results from several independent tests bearing upon the same overall hypothesis. Fisher's method can be useful in meta-analysis, where the goal is to combine independent p-values for a certain null hypothesis from multiple analyses. For example, suppose that the effect of X on the response, Y, was assessed by k independent studies, and the obtained p-values are p_1, \dots, p_k . Then, to check if the effect of X on Y is significant, combining all the evidence from the k studies, compute the statistic

$$X^2 = -2 \sum_{i=1}^k \log_e(p_i)$$

When all k null hypotheses are true (no effect of X on Y in any of the k experiments), then X^2 has a Chi-square distribution with $2k$ degrees of freedom. Using this fact, one can test whether the combined evidence allows rejecting the null. When combining dependent p-values Brown's method or Kost's method can be used instead.

6. User-friendly power and sample size calculations using SAS.

SAS 9.2 comes with a very user-friendly power and sample size calculator with a graphic user interface. To begin, choose the following menu selection in Windows: Start > Programs > SAS > SAS Power and Sample Size > SAS Power and Sample Size 3.1. A new window will be opened where you need to click on File > New to get started. You will have the choice of performing sample and power calculations for many different types of analyses, e.g. one- and two- sample t-tests, tests for proportions, general linear models, various equivalence tests and two sample survival rank test. The next screen shows several tabs which help the user choose the necessary parameters for the calculations. When all parameters have been entered, click on the "Calculate" button. The results can be displayed in a table, graphically and a narrative format.

7. Many useful upgrades with the latest version 12 of STATA.

The new version 12 of Stata was released about 4 months ago and offers many useful upgrades. Some of the new additions that you might enjoy are:

- Capability to import and export Excel files
- Automatic memory management and upgraded interface
- Estimation of multilevel models with complex survey data
- A new Structural Equation Modeling command with extensive capabilities
- A new rocreg command to model ROC curves controlling for covariates
- Ability to fit multivariate GARCH models
- Upgrades in the multiple imputation commands include chained equations (allowing imputing variables of different types), imputation by group, conditional imputation (imputation within group when group itself might be imputed) and estimating multilevel models with multiple imputation
- New contrasts, pairwise comparisons and margins command and plots
- Estimation using Full Information Maximum Likelihood when data are missing at random
- New tpoisson and tnegbin commands to model truncated poisson or negative binomial models.