Cornell University
Cornell Statistical Consulting Unit

Stats Happening – November 2015

1) **CSCU Winter Session Workshop on Multilevel Modeling**
2) **Model Specification in R**
3) **R^2 for Generalized Linear Mixed Effects Models**
4) **Options When Ordinal Logistic Regression Assumptions Are Violated**
5) **Equivalence Testing in SPSS**
6) **New in Stata 14: Bayesian Analyses and Item Response Theory**
7) **Organizing R Model Output Using the New `broom' Package**
8) **Multiple Correspondence Analysis in JMP**

1) **CSCU Winter Session Workshop on Multilevel Modeling: Jan 25th & 26th 2015**

Multilevel models (also commonly referred to as mixed models or hierarchical models) can be used when the standard regression model assumption of independent observations is not met. Dependent observations may be the result of a specific experimental design (e.g., split plot designs), data collected simultaneously on different observational units of analysis (e.g., households and their individual members), measurements taken on the same observational units over several time periods (longitudinal studies), or measurements taken at different geographic locations (spatial data). The purpose of this workshop is to introduce the concepts behind the analysis of multilevel data in addition to relevant statistical models and estimation techniques for analyzing multilevel data. Numerous examples will be presented during the workshop to enable participants to recognize multilevel data in their own research, be able to apply the appropriate statistical methods, and to interpret the results. The workshop is intended for participants who have had the equivalent of two semesters of statistics and some previous experience with ANOVA and linear regression. The workshop will consist of a classroom lecture followed by a hands-on session in a computer. The classroom lectures will be on Monday, January 25th and Tuesday, January 26th from 10:00am - 12:00pm in Savage 100. The hands-on session will be held on Tuesday, January 26th from 2:00pm - 4:00 pm in Mann Library. Register by completing the form on our website:
https://cscu.cornell.edu/workshops/registration.php.

2) Model Specification in R

Whether you are new to R or an experienced user, model specification in R can be a tricky matter. We have compiled a list of useful model specification commands that spans analysis of variance, multiple regression, and mixed models using the R functions lm and glm in the base package, and the mixed model functions lmer and glmer in the lme4 package. A handout is available on our website: https://cscu.cornell.edu/news/Handouts/R_Model_Specification.pdf.
The nlme package can also be used to run mixed models, however the specification of the random effects for the two packages is different. For more information on the differences between the lme4 and nlme packages, see our 2014 Stats Happening: https://cscu.cornell.edu/news/StatHappenings/sthapsp14.pdf.

3) R^2 for Generalized Linear Mixed Effects Models

Often researchers using mixed/multilevel models or generalized linear mixed models are interested in reporting an R^2 value as is commonly done with linear models. In the linear model setting, the R^2 value represents the amount of variability in the dependent variable explained by the model. In the case of mixed models, the amount of variability explained by the model can have several interpretations. Questions such as whether variability explained by just fixed effects (marginal R^2) or both the random and fixed effects (conditional R^2) is of most interest, and how to handle multiple levels of random effects must be addressed. Moreover, the R^2 value must be easy to implement and interpret. The following easy to read paper proposes methods to calculate both different types of R^2 values for mixed models and generalized linear mixed models: http://onlinelibrary.wiley.com.proxy.library.cornell.edu/doi/10.1111/j.2041-210x.2012.00261.x/pdf The proposed methods are easy to implement regardless of the statistical software package. Please contact CSCU if you would like assistance implementing these methods.

4) Options When Ordinal Logistic Regression Assumptions Are Violated

When a response variable has more than two categories, there are several methods available to extend the technique of logistic regression. The choice of method depends on whether the response variable is measured on an ordinal or nominal scale.
When the response is nominal, multinomial logistic regression can be used. This compares all levels of the response to a reference level through a set of logistic regression. When the levels of the response variable are ordered, ordinal logistic regression can be used. The major assumption with an ordinal logistic regression model is the assumption of proportional odds (sometimes called the assumption of parallel lines), meaning that the slopes relating the response to each covariate are constant across all cumulative logits of the response. When this assumption is violated researchers usually turn to multinomial logistic regression. However, another possible solution is generalized ordinal regression. This approach fits a partial proportional odds model which relaxes the parallel lines assumption for those variables where

the violation is observed and calculates different estimates for each cumulative level, all while keeping a constant slope for variables that do not violate the proportional odds assumption. This model can be implemented in Stata using the ado file gologit2. For more information on gologit2's options and interpretation of the output, see to the Stata manual: http://www.stata.com/meeting/4nasug/gologit2.pdf.

5) Equivalence Testing in SPSS

When your research objective is to show that two groups are similar enough for practical purposes, an equivalence test (rather than a standard test for differences between means) may be an appropriate statistical tool. For more information on equivalence testing, check out our newsletter #85: https://cscu.cornell.edu/news/statnews/stnews85.pdf.

Although most commonly used statistical software can be used to implement equivalence testing, there is no option in the base installation of SPSS that will perform equivalence testing. However, you can download an SPSS macro that will do the calculations for independent sample equivalence tests and paired samples equivalence tests: http://www.medianeuroscience.org/equivalence_testing.

6) New in Stata 14: Bayesian Analyses and Item Response Theory

Stata 14 offers several very useful upgrades. Two notable new additions are the following:
- Bayesian regression, hypothesis testing, and many other Bayesian analyses can now easily be conducted in Stata 14, which comes equipped with 12 built-in likelihood models and 22 built-in prior distributions.
- Researchers interested in creating instruments to measure various abilities, traits, or other characteristics that cannot be measured directly, might be pleased to learn that Item Response Theory (IRT) models can now be implemented in Stata.

For more information about new features in Stata 14, see: http://www.stata.com/new-in-stata/.

7) Organizing R Model Output Using the New `broom' Package

The new `broom' package for R can be used to convert output from the lm(), nls(), glm(), and t.test() functions into a nice, usable format. The following three functions in the broom library are particularly useful:
- tidy: Constructs a data frame that summarizes the model's statistical findings. This includes coefficients and p-values for each term in a regression.
- augment: Add columns to the original data that was modeled. This includes predictions, residuals, and cluster assignments.
- glance: Construct a concise one-row summary of the model. This typically contains values such as R^2, adjusted R^2, and residual standard error.

8) Multiple Correspondence Analysis in JMP

Multiple Correspondence Analysis can now be performed in JMP 12 using the Consumer Research menu. Multiple correspondence analysis (MCA) is a multivariate analysis method which can be used to identify and explain systematic patterns of variations with categorical data on the same observational units. MCA can be used to construct useful plots describing the relationships between the categorical variables. MCA is a technique that may be useful to researchers in many fields including the plant sciences, genetics, and the social sciences.