



## CSCU Announcements - June 2014

1. **Congratulations to the graduates**
2. **CSCU Summer Schedule**
3. **Summer 2014 CSCU Workshops**
4. **Generating Automatic Reports**
5. **Fitting Mixed Models in R: *lme4* versus *nlme***
6. **JMP 11 Introduces New Features**
7. **pqR: Pretty Quick R**
8. **Mixed Model Multinomial Logistic Regression in Stata 13**

### 1. Congratulations to the graduates

The statistical consultants at CSCU extend their congratulation to all the graduates. We have enjoyed working with you and hope that you will keep on using and expanding the statistical knowledge you acquired at Cornell. We wish you all the best in your future endeavors.

### 2. CSCU Summer Schedule

CSCU will be available for walk-in consulting and will schedule appointments throughout the summer as usual.

- **Walk-in Consulting:** Staff statisticians will be available for questions that take ten minutes or less daily between 11:00-11:30am and 1:30PM and 2:00PM in B07, B09, B11 and/or B13 Savage Hall.
- **Appointments:** Scheduling of appointments is encouraged for matters that require more than 10 minutes time. To schedule an appointment contact a staff statistician or visit: <http://www.cscu.cornell.edu/about/appointment.php>

### 3. Summer 2014 CSCU Workshops

Besides offering the Software Carpentry Bootcamp next week, CSCU will also be offering the following workshops this summer:

<b>Date/Time</b>	<b>Workshop</b>	<b>Location</b>
Tue 6/10 & Thu 6/12 9:00-11:00 AM	<b>Basic Data Analysis Workshop</b> Two Part Workshop	Stone Computer Room Mann Library
Wed 6/11 2:30-4:30 PM	<b>Introductory Statistical Analysis Using R</b>	Stone Computer Room Mann Library
Mon 6/16 1:00-3:00 PM	<b>Refresher on Interpreting Linear Regression Parameters</b>	200 Savage
Tue 6/17 10:00 AM -Noon Thu 6/19 10:00 AM –Noon Thu 6/19 2:00 PM-3:30 PM Fri 6/20 10:00-11:30 AM	<b>Introduction to Multilevel Modeling</b> Three Part Workshop	100 Savage Hall 100 Savage Hall Stone Computer Room Mann Library
Mon 6/23 9:00-11:00 AM	<b>Univariate Repeated Measures versus MANOVA</b>	200 Savage
Wed 6/25 2:30-4:30 PM	<b>Intermediate Statistical Analysis Using R</b>	Stone Computer Room Mann Library

For more information and to register for these workshops go to:

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

### 4. Generating Automatic Reports

Several of our clients have been interested in creating routine identical reports as data becomes updated from the field. These reports are usually very similar, and require a fair amount of cutting and pasting to a document, which is tedious and sometimes error-prone.

This problem can be addressed by dynamic report generation, which is becoming more accessible in standard statistics packages. This involves combining standard statistical syntax with plain text formatting syntax (Markdown) to produce a single document with tables and figures embedded among the textual content of the paper or report.

Here are some pages with information on how to do this in R and Stata:

**R:** <https://support.rstudio.com/hc/en-us/articles/200552086-Using-R-Markdown>  
(it's not markdown, but "rtf" is a package that can produce simple RTF files: <http://cran.r-project.org/web/packages/rtf/index.html>)

**Stata:** <http://hopstat.wordpress.com/2014/01/11/stata-markdown-2/comment-page-1/>

**Other:** <http://yihui.name/knitr/demo/engines/>

Markdown can then be converted into various other formats (.docx, .pdf, .html) using Pandoc:  
<http://johnmacfarlane.net/pandoc/getting-started.html>

For help setting up a workflow that involves dynamic report generation, do not hesitate to make an appointment with CSCU.

## 5. Fitting Mixed Models in R: *lme4* versus *nlme*

There are currently two packages commonly used to fit mixed effects models in R. The **nlme** package has been available for many years and can be used to fit linear mixed models with nested random effects. The **lme4** package is a slightly newer package, developed by one of the authors of **nlme**, which can be used to fit linear mixed models with more flexible random effects structures as well as generalized linear mixed models. However, there are also several other differences in the capabilities of the two packages which may make one package preferable in certain situations. The table below provides an overview of some of the differences between **lme4** and **nlme**.

	<b>lme4</b>	<b>nlme</b>
Model specification	Fixed and random effects specified in a single model statement.	Fixed and random effects specified separately.
Random effects	Easily handles both crossed and nested random effects.	Optimized for nested random effects.
Error distribution	Capable of fitting both normal error linear mixed models and generalized linear mixed models.	Can only fit normal error linear mixed models.
Correlation structure	Can only accommodate diagonal and unstructured correlation structures.	More flexible modeling of correlation structures possible. Options include diagonal, unstructured, compound symmetric and AR(1).
p-values	Not provided by default. Must use profile likelihood tests or parametric bootstrap tests for fixed effects, or rely on other packages to obtain approximate F-tests and p-values.	Provided by default. F-tests and p-values for fixed effects are computed using the classical decomposition of degrees of freedom.
Speed	Faster and more reliable for large data sets.	Not recommended for large data sets.

## 6. JMP 11 Introduces New Features

The latest release of JMP (version 11) extends its capabilities in several key ways. First, by adding a Mixed Model personality in Fit Model, one can now specify fixed, random and/or repeated effects. The addition of the “repeated” option allows for more flexible estimation of the covariance structure (e.g. autoregressive, unstructured, spatial).

Second, JMP 11 now includes a “Generalized Regression” personality in Fit Model. This refers to a number of model selection techniques known as regularized or penalized regression, including Ridge regression, the Lasso, and Elastic Net. These methods are useful when fitting regression models with a large number of explanatory variables by imposing a penalty on large fluctuations on the estimated parameters, which can help avoid multicollinearity and over-fitting. These regularization techniques can be applied to both linear and generalized linear (e.g. Poisson, ZIP, Negative Binomial, ZNB, and Gamma) regression models.

<http://www.jmp.com/software/pro/new-in-jmp-pro.shtml>

## **7. pqR: Pretty Quick R**

Do you use R for data analysis but find the computational speed to be unacceptably slow? pqR is a new version of the R interpreter. It is based on R-2.15.0 and distributed by the R Core Team, but improves speed, implements new features and fixes some bugs. One notable improvement is that on systems with multiple processors or processor cores, pqR is able to do some numeric computations in parallel with each other, and with other operations of the interpreter.

<http://www.pqr-project.org/>

## **8. Mixed Model Multinomial Logistic Regression in Stata 13**

A multinomial logistic regression model with random effects has always been challenging to implement. Often when faced with a multilevel multinomial logistic regression model with  $k$  levels in the response, researchers will opt to run  $k-1$  logistic regressions models. With the new Stata version 13 however running one multinomial multilevel logistic regression has been made easier. The complication in the implementation of one multinomial model with random effect arises partly due to the fact that it is difficult to model separate random effects for each of the levels. This problem has been addressed with the new Stata generalized structural equation modeling command `gsem`. It allows with one easy command to model a multinomial logistic regression model with multiple independent or correlated random effects. For an example see:

<http://www.stata.com/manuals13/semexample41g.pdf>