



## Stat Happenings #9: Statistics News You Can Use Cornell Statistical Consulting Unit May 2009

1. **Statistical workshops at Cornell Summer 2009**
2. **Noteworthy updates in statistical software available at Cornell**
3. **Recommended readings**
4. **Off-campus Statistical Workshops**

### 1. Statistical workshops at Cornell Summer 2009

- This summer CSCU will be offering again the popular *Basic Data Analysis* workshop. If you need to collect and analyze data for a research project and need help with your practical skills, this workshop can get you started. It will be offered in two parts, with the first part being offered Tuesday, June 16 from 10:00 AM - noon and the second part on Thursday, June 18 also from 10:00 AM noon in the Mann Library Stone Computing Room.

For more information and to register go to: <http://www.cscu.cornell.edu/workshops/schedule.php>

- The 24th **International Workshop on Statistical Modeling** will be hosted by Cornell University. The event will take place **July 20 to 24**, 2009. The workshop presents the use of statistical modeling in various areas of research and applied problems. More information on this workshop can be found at <http://www.statmod.org/>

A short course on "**Bayesian Methods for Model Comparison**", presented by Murray Aitkin, July 19, 2009 will be offered as part of the International Workshop on Statistical Modeling.

For more information and to register for the International Workshop on Statistical Modeling and the short course please contact Jim Booth, Professor and Chair, Department of Biological Statistics and Computational Biology at [jim.booth@cornell.edu](mailto:jim.booth@cornell.edu) .

## 2. Noteworthy updates in statistical software available at Cornell

CSCU strives to keep the Cornell community abreast of new developments in the field of applied statistics, so researchers can employ the latest methodologies. Following is a review of new statistical tools made available at Cornell with the latest software updates:

**a.** The latest release from SAS, **SAS 9.2**, is now available at Cornell. Here is a sampling of new features available in SAS 9.2:

- For survival analysis, Proc PHREG includes now a `class` statement, previously available only with the experimental procedure TPHREG.
- For fitting generalized linear models, Proc GLIMMIX, originally available as a download only in SAS 9.1.3, is part of SAS 9.2. It allows fitting additional covariance structures, including heterogeneous AR(1), heterogeneous compound symmetry, linear structures, heterogeneous Toeplitz, and penalized B-spline.
- Bayesian analysis for generalized linear models is incorporated into Proc Genmod, Lifereg and Phreg, using a Bayesian statement.
- An experimental procedure, HPMixed, is available for fitting mixed models with a large number of fixed and random effects, with shorter computer processing times than the usual Proc Mixed.

For more information on new SAS 9.2 features go to: <http://support.sas.com/rnd/app/da/new/dastat92.html>.

- If you are running Microsoft Windows XP Home Edition or Microsoft Windows Vista Home Edition on your PC, you need to know that SAS 9.2 will not be supported on these platforms. SAS 9.2 will be supported on Microsoft Windows Vista: Enterprise, Business and Ultimate Editions. It will also work with Microsoft Windows XP Professional - Service Pack 2 or later.
- Cornell continues to support the current version, SAS 9.1.3. For more information and for ordering please go to: <http://www.cusoftware.cornell.edu/cusoftware/purchase/sas.cfm>.
- Our advice for those who need SAS Enterprise Miner for their analyses: continue to use SAS 9.1.3 version, as SAS Enterprise Miner module is NOT included in the new SAS 9.2.

**b.** **STATA** version **10** has expanded its capabilities for analyzing multilevel random effects models for binary and count responses. A new estimation routine, *xtmelogit*, has been added for analyzing random intercept logistic models, in addition to *xtlogit* command available in STATA 9. This new *xtmelogit* command can be used to fit 3-level and higher logistic regression models.

Similarly, for mixed-effects models for count responses a new command, *xtmepoisson* is available. Both commands have a similar syntax and output. More at: <http://www.stata.com/stata10/mixedmodels.html>. For purchasing information on STATA 10 at Cornell please see our website at [www.cscu.cornell.edu](http://www.cscu.cornell.edu).

c. **SPSS 17** is now available at Cornell (the annual license runs through March 31, 2010). We advise you to download and apply the **SPSS Patch 17.0. 1**. For more information on the issues addressed by the patch and to download it go to:

[http://support.spss.com/ProductsExt/SPSS/Patches/SPSS%20Server/17.0.1/windows/17.0.1\\_Readme.html](http://support.spss.com/ProductsExt/SPSS/Patches/SPSS%20Server/17.0.1/windows/17.0.1_Readme.html)

As a new analysis tool, multiple imputation for analyzing data sets with missing values, is now available with the add-on module Missing Value Analysis. This module is included in the SPSS Statistics license fee from [CU Software Licensing Services](#).

For those who use SPSS for statistical analysis and need advanced statistical routines not yet available in SPSS, the R extension commands available in SPSS may be of interest. Several R extensions commands can be run in SPSS, including Tobit truncated regression, quantile regression, and Rasch models. The SPSS extension commands act like built-in SPSS commands. They are add-on modules that can be obtained from the SPSS Developer Central [www.spss.com/devcentral](http://www.spss.com/devcentral) and need the Python and/or R plug-in modules. Most of these extension commands work with both SPSS 17 and older versions.

Other notable new features include a syntax editor that allows highlighting commands, subcommands, and keywords using different colors, making your syntax easier to read and debug. Errors appear in the output window but are also highlighted in the syntax editor.

The Codebook module allows for automatically creating a summary of your data, including variable names, value labels, and missing values.

The output files in SPSS 17 have a new file format, “.spv”. The previous “.spo” output files can still be viewed and edited with SPSS Legacy Viewer, available after installing SPSS 17.

d. The new version of JMP software, **JMP 8**, will soon be available at Cornell. A SAS toolbar is included in JMP 8 to access SAS data sets and features directly. JMP 8 works with both SAS 9.1.3 and SAS 9.2. JMP 8 has a new interactive feature, a graph builder platform for interactive data exploration and visualization. To order JMP at Cornell see <http://amber.cit.cornell.edu/citweb/software/swl/jmp/order.cgi>

For an introduction to using JMP for your statistical analyses, check out the book: *JMP Start Statistics: A Guide to Statistics and Data Analysis Using JMP*, Sall J et al., Fourth Edition. A preview of this book is available at:

<http://books.google.com/books?id=M3Mr0THgdd8C&pg=PR14&lpg=PR14&dq=Jmp+7+tutorial&source=bl&ots=y9v000U-7S&sig=7K4Vu1TmMZdMohk-5fuykzUyv4w#PPR10,M1>

e. Did you experience problems working a Reliability/Survival graph in Minitab 15? If you use Minitab 15 for statistical analysis, the installation of the latest updates is recommended. The update, **Minitab 15.1.3.**, contains fixes and usability enhancements to Minitab 15. To install the Minitab 15.1.3 fix, which will fix problems such as the one mentioned above, start Minitab 15 and select **Help -> Check for Updates**.

For a Minitab tutorial go to: <http://www.minitab.com/support/docs/rel15/MeetMinitab.pdf> or use the tutorials available in the Minitab help system.

We would also like to remind you that **CISER** offers data and computing workshops, including a number of statistical software programs, such as SAS, SPSS, and Stata. See their website for more information: <http://ciser.cornell.edu/beta/workshops/>

**3. Recommended reading:** *The abuse of Power: The Pervasive Fallacy of power Calculations for Data Analysis*, Hoenig J.M.

Will you be reporting a statistically not significant result in your study? If so, can this result be interpreted as not significant due to a treatment effect that is almost 0 or the result of low statistical power in determining the effect?

For a debate on this question, read *The abuse of Power: The Pervasive Fallacy of power Calculations for Data Analysis*, Hoenig J.M., *The American Statistician*, Feb 2001, 55, 1, and it can be found at <http://web.vims.edu/fish/faculty/pdfs/hoenig2.pdf> .

The article makes the case of using confidence intervals instead of (post-hoc) power calculations for the interpretation of non significant results. If computed confidence interval for the parameter of interest is relatively narrow and includes the null value, this is evidence that the effect is almost 0 or not significant.

#### 4. Off-campus Statistical Workshops

Do you want to take advantage of the summer to learn about longitudinal data analysis or Bayesian modeling? Would you like to get an introduction to structural equation models and latent variables? Below is a selection of off-campus statistical workshops offered this summer that you might find useful.

- The 2009 Joint Statistical Meeting held by the American Statistical Association is in Washington DC and offers a variety of continuing education classes. Check out their website at <http://www.amstat.org/meetings/jsm/2009/index.cfm?fuseaction=workshops>.
- The University of Michigan has a summer program that offers a large range of statistical workshops. For topics covered and more information check out their website: <http://www.icpsr.umich.edu/sumprog/2009/schedule.html#fourweek>

Also from the University of Michigan, check out these summer workshops: <http://www.umich.edu/~cscar/workshops/>

- Learning the **R programming** language? The online short course, Introduction to R., will be offered May 8 to May 29 2009. For more information go to: <http://statisticsonline.info/intro2R.htm>
- The summer statistics workshops offered by the Department of Educational Psychology at Texas A&M University include a number of statistical applications in quantitative methods in the social

sciences. Topics include hierarchical modeling and nonparametric statistics. Please see <http://epsy.tamu.edu/articles/ssw> for more information.

- A categorical data analysis workshop is offered July 13-17, Philadelphia, PA by Paul Allison. It covers topics such as binary logistic regression, ordered logistic regression, multinomial logit, GEE estimation, multilevel logistic regression, Poisson and negative binomial regression, loglinear models, discrete choice models and longitudinal data. For more information and other courses available go to <http://www.statisticalhorizons.com>.