



Cornell University Cornell Statistical Consulting Unit

Stats Happening – October 2016

- 1) **CSCU Fall Schedule**
 - 2) **Fall 2016 CSCU Workshops**
 - 3) **Ten Simple Rules for Effective Statistical Practice**
 - 4) **Should Post-hoc Sample Size Calculations Be Performed?**
 - 5) **Specifying Mixed Models in Various Statistical Software Packages**
 - 6) **The Many Uses of Stata's gsem Command**
 - 7) **Recommended R Packages**
 - 8) **New Features in JMP 13**
 - 9) **What We Are Listening to: Statistics Podcasts**
-

1) **CSCU Fall Schedule**

CSCU will be available for walk-in consulting and scheduled appointments throughout the Fall semester as usual.

- Walk-in Consulting: Staff statisticians will be available daily for questions that take ten minutes or less at the following times and locations:

Monday-Friday 11:00am – 11:30am: B11 and B18 Savage Hall

Monday-Friday 12:00pm – 1:00pm: Mann Library Consulting Area (1st Floor)

Monday-Friday 1:30pm – 2:00pm: B09 and B13 Savage Hall

- Appointments: Scheduling of appointments is encouraged for matters that require more than 10 minutes of time. To schedule an appointment, contact a staff statistician or visit: <http://www.cscu.cornell.edu/about/appointment.php>.

2) Fall 2016 CSCU Workshops

The following CSCU workshops will be offered later this semester.

Workshop	Date and Time
Getting Started with Data Analysis	October 20 th , 6:00-8:00pm
Visualizing Data using ggplot2 in R	October 28 th , 2:00-4:00pm
Nonparametric Statistical Methods	November 1 st , 2:30-4:30pm
Introduction to Modeling Count Data	November 4 th , 1:00-3:00pm
Model Selection and Multimodel Inference	November 10 th , 1:30-3:00pm
Missing Data (a two part workshop)	November 17 th and November 18 th 1:30-3:00pm

For more information about our the workshops, and to register, please visit <https://cscu.cornell.edu/workshops/schedule.php>.

3) Ten Simple Rules for Effective Statistical Practice

The PLOS journal of Computational Biology has released a list of rules for effective statistical practice. We believe these rules are useful to keep in mind at all stages of a research study. These rules include:

- i. Statistical methods should enable data to answer scientific questions
- ii. Signals always come with noise
- iii. Plan ahead, really ahead
- iv. Worry about data quality
- v. Statistical analysis is more than a set of computations
- vi. Keep it simple
- vii. Provide assessments of variability
- viii. Check your assumptions
- ix. When possible, replicate
- x. Make your analysis reproducible

The full journal article can be accessed by visiting:

<http://journals.plos.org/ploscompbiol/article?id=10.1371%2Fjournal.pcbi.1004961>.

4) Should Post-hoc Sample Size Calculations Be Performed?

In many fields, pre-study sample size calculations are mandatory. By specifying a clinically relevant effect size, the significance level, and power, sample sizes can be calculated for various statistical tests. These calculations help researchers design studies which are both effective and efficient. However, when a treatment effect is found insignificant, researchers often ask, “How many subjects would I have needed to declare this effect significant?” This is not a meaningful question to ask, as the observed effect may not be clinically relevant. Using the observed effect, you obtain a power that changes in step with the p-value for the observed data. For more information on this issue, check out the following articles: <http://www.ncbi.nlm.nih.gov/pubmed/18416448> and http://www.vims.edu/people/hoenig_jm/pubs/hoenig2.pdf.

5) New CSCU Handout: Specifying Linear Mixed Models in Various Statistical Programs

Linear Mixed Models are useful statistical tools when analyzing a response variable which was measured on the same experimental units over multiple time points, or analyzing data that is clustered due to the design of the study. Every statistical software program has its own default settings for statistical methods, and being unaware of these settings can produce misleading results. The following document provides code and directions to reproduce the same results in R, Stata, SPSS, SAS, and JMP: http://cscu.cornell.edu/news/Handouts/Mixed_Models.pdf.

6) The Many Uses of Stata’s gsem Command

Stata’s **sem** command executes structural equation models to determine mediation and moderation effects on continuous response variables. The **sem** command does not support multilevel analyses. However, the **gsem** (generalized structural equation model) command allows users to relax the assumption that each response variable be continuous; users can select from a list of families that each response variable follows including Gaussian, Binomial, Multinomial, Poisson, and Negative Binomial. Also, **gsem** allows users to specify multilevel models. Currently, the only way to estimate a multilevel multinomial logistic regression in Stata is through the **gsem** command. However, the one drawback of the **gsem** command is that it lacks the ability to test indirect effects.

7) Recommended R Packages

For new and frequent users of R, the following recently published article on the R-Studio Support page provides a list of recommended packages not included in base R. These packages can make data manipulation and visualization easier; some of the recommended packages are necessary to perform advanced statistical methods, such as mixed models and survival analysis. The only package not on this list that we at CSCU use frequently is the *lsmeans* package. The full list of the recommended packages can be accessed by visiting:

<https://support.rstudio.com/hc/en-us/articles/201057987-Quick-list-of-useful-R-packages>.

8) New Features in JMP 13

There are several new useful tools in the new release of JMP 13. New to the software is the text explorer, which helps users visualize and analyze text data. Users can also create a web report containing graphs, numerical output, and notes. Data tables can be joined using JMP's new Query Builder function.

9) What We Are Listening to: Podcasts on Statistics

No, these podcasts are not monotone presentations of mathematical proofs on the central limit theorem. Instead, these podcasts center on data scientists and researchers explaining the results in fields ranging from economics and political science to botany and physics. Increase your numerical and statistical literacy by checking out these podcasts:

- Freakonomics
- Hidden Brain (NPR)
- More or Less (BBC4)
- Not So Standard Deviations
- Partially Derivative
- Stats + Stories (Miami University)
- What's The Point (FiveThirtyEight)