



Stats Happening - November 2013

1. **Web Resources for Data Visualization**
2. **Vuong Test for Comparing Non-Nested Models**
3. **Noteworthy Upgrades in Stata 13**
4. **Tobit and Quantile Regression in SPSS**
5. **Saving Model Information in SAS with Proc PLM**
6. **Follow CSCU on Twitter**

1. Web Resources for Data Visualization

There are an increasing number of exciting web-based tools and resources for visualizing and displaying data:

- [Tableau Public](#) is a free tool that allows you to upload your data and experiment with different forms of data display and visualization. It's very easy to use, can support interactive features and will even suggest the most appropriate chart for the data you provide. You can embed your finished chart or graph onto your webpage for users to explore your data themselves. Everything you make is stored publicly on the web, however. If local storage is important to you, there is a desktop version that comes at a price, but there are options for student and educational pricing.

Examples: Used frequently by [The Seattle Times](#), [Newscientist](#), [the Huffington Post](#), [CBS Sports](#), and [countless others](#).

- If you love analyzing your data in R, you'll love [Shiny](#). It's a package from RStudio, and allows you to create interactive web applications to let your data, maps, or model outputs react to input parameters chosen by your webpage audience using sliders, drop-down menus, or buttons.

Examples: Learn how the [ordinary least squares regression estimator](#) is affected by various aspects of your data or explore some Harvard [Dialect Survey Maps](#).

- If you prefer to do your data manipulation and visualization in Excel, you'll no doubt be familiar with the chart defaults. And like many things Microsoft, the built-in defaults are often not the best way to show what you want. [ExcelCharts.com](#) is a blog by [Jorge Camoes](#) that focuses on good chart design principles and explains how to use them in Excel. Some courses and tutorials

are premium content, but there are great blog posts and examples available for free. Recently, a blog post featured a [classification of standard Excel chart types](#) and when to use them.

- Did your data chart or visualization not turn out the way you expected? Did you mis-code or mis-specify something? Take a step back and critique it like a piece of art. Like what you see? Submit it to the new blog [accidental aRt](#) – when data visualization goes beautifully wrong. This blog takes [submissions](#) from researchers from around the world, and is curated by CSCU's own [Erika Mudrak](#) along with information manager [Kara Woo](#).

2. Vuong Test for Comparing Non-Nested Models

The likelihood ratio test is a powerful, widely available statistical test for comparing the fit of two nested models. But what if the models that you would like to compare are not nested? The Vuong test can be used to compare models which are strictly non-nested (for example, have different error distributions or the same error distribution but different functional forms). One example of a strictly non-nested comparison would be comparing a Poisson (or negative binomial) model to a zero-inflated Poisson (or zero-inflated negative binomial) model. The statistic tests the null hypothesis that the two models are equally close to the actual model, against the alternative that one model is closer. However, it cannot determine whether the "closer" model is the true model. The Vuong test for strictly non-nested models can be performed in STATA, using the Vuong macro in SAS, and using the `pscl` package in R. An example of how to perform the Vuong test in R to compare a Poisson model to a zero-inflated Poisson model can be seen here: <http://www.ats.ucla.edu/stat/r/dae/zipoisson.htm>

3. Noteworthy Upgrades in Stata 13

Stata version 13 offers several noteworthy improvements:

- Maybe the most impressive of the enhancements is the one made to their set of commands for structural equation modeling (SEM). The most recent version of Stata will allow you to estimate SEM models with not only normally distributed response variables but also an array of generalized linear responses (binary or ordered logistic, count data, etc.). Multilevel or hierarchical SEM models can now also be estimated including random slopes and intercepts. All this can be done using several different estimation methods.
- Another noteworthy enhancement that has been made in version 13 is the capability to now run multilevel and panel data models with a probit, negative binomial, ordered probit, ordered logistic and multinomial logistic responses.
- Many commands for causal inference have been consolidated into the *treatment effects* (*teffects*) command. This includes inverse probability weights, propensity score matching, doubly robust methods, and much more.

For a more thorough exploration of what is new visit the [Stata 13 webpage](#).

4. Tobit and Quantile Regression in SPSS

Two new analysis methods are now available to SPSS users: Tobit Regression and Quantile Regression. Tobit regression is used when the dependent variable is censored, i.e. when cases with a value at or beyond some threshold all take on the value of that threshold. The true value might be equal to the threshold, but it might also be higher (right censored) or lower (left censored) than the threshold. Quantile Regression estimates the conditional median or 1 more other conditional quantiles for a linear model. This is contrast to ordinary linear squares (OLS) regression which estimates the conditional mean.

These methods are implemented as part of SPSS extension bundles which run the analysis via R hidden in the background. To install these (and many other) extensions within the SPSS, download the [extension files](#) and an appropriate version of [R](#) (free, open-source). The Tobit Regression and Quantile Regression options will then appear within the SPSS drop-down menus and can be run like any other command.

5. Saving Model Information in SAS with Proc PLM

A procedure that is often overlooked in SAS is the rather new Proc PLM. Proc PLM is used in conjunction with a STORE statement that is available in many procedures and saves model information as a SAS item store. Proc PLM then takes these item stores as input and produces many postfitting statements such as testing hypotheses, computing confidence intervals, producing prediction plots and scoring a new data set, all using regular SAS statements such as lsmeans and test. This procedure is especially useful when running computationally intensive models as you could save the postfitting inferences for a later time without having to refit your model. The statistical procedures that offer a STORE statement and support Proc PLM are: GENMOD, GLIMMIX, GLM, LOGISTIC, MIXED, ORTHOREG, PHREG, SURVEYLOGISTIC, SURVEYPHREG, and SURVEYREG. This is very easy to use as the following example demonstrates:

```
proc glm data=ph_experiment;  
class ph;  
model response=ph;  
store model1;  
run;  
proc PLM source=model1;  
lsmeans ph;  
run;
```

6. Follow CSCU on Twitter

If you enjoy reading our Stats Happening you might also consider following [CSCU on Twitter](#) as we highlight similar statistics related information.