

# 2005 Statistical Highlights

## Four Exciting Developments

By Frank LoPresti  
frank.lopresti@nyu.edu

This year has already produced four impressive developments in statistical tools: SPSS Output Management System; The University of California, Los Angeles (UCLA) Academic Technology Services' website; MIT's OpenCourseWare; and Minitab Version 14. These new tools are described below, with information on how to access them.

### SPSS OUTPUT MANAGEMENT SYSTEM

SPSS Output Management System (OMS) provides an output control function very much like a series of "if" commands; they are turned on and off during the session. Using OMS, you would first create several output files, each with its own format: HTML; XML; SPSS Dataset; or delimited text. You can then choose what SPSS statistical commands will trigger the writing of output to the various files you created.

For example, you could open a file called "anova\_out.sav" and, with OMS control, append the output to that file whenever an anova command is run. Since the file is formatted as an SPSS file, the statistics will be written as data, and the tables of Sums of Squares, F, etc., will all be saved as data. Say you have 20 years of weekly data and run similar anova commands on the 20 year groups; the file "anova\_out.sav" would have 20 lines of data with the statistics from each of the

anovas. The variables in this dataset would be statistics such as Sums of Squares or F—the various statistics created by the anovas.

Visit UCLA's OMS website, <http://www.ats.ucla.edu/stat/spss/faq/oms.htm>, for a variety of command syntax examples and other useful informa-

tion. OMS is included with SPSS version 12 and above; NYU has a site license for SPSS version 13 (and previous versions, if required). See <http://www.nyu.edu/its/spss.html> for information on acquiring SPSS from NYU Information Technology Services.

UCLA Academic Technology Services  
Stat Computing

Home Services Classes Contact Jobs Search

What statistical analysis should I use?

The following table shows general guidelines for choosing a statistical analysis. We emphasize that these are general guidelines and should not be construed as hard and fast rules. Usually your data could be analyzed in multiple ways, each of which could yield legitimate answers. The table below covers a number of common analyses and helps you choose among them based on the number of dependent variables (sometimes referred to as outcome variables), the nature of your independent variables (sometimes referred to as predictors). You also want to consider the nature of your dependent variable, namely whether it is an interval variable, ordinal or categorical variable, and whether it is normally distributed (see [What is the difference between categorical, ordinal and interval variables?](#) for more information on this). The table then shows one or more statistical tests commonly used given these types of variables (but not necessarily the only type of test that could be used) and links showing how to do such tests using SAS, Stata and SPSS.

| Number of Dependent Variables                      | Nature of Independent Variables | Nature of Dependent Variable(s) | Test(s)                     | How to SAS | How to Stata | How to SPSS |
|--|---------------------------------|---------------------------------|-----------------------------|------------|--------------|-------------|
| 0 IVs<br>(1 population)                            |                                 | interval & normal               | one-sample t-test           | SAS        | Stata        | SPSS        |
|  |                                 | ordinal or interval             | one-sample median           | SAS        | Stata        | SPSS        |
|  |                                 | categorical<br>(2 categories)   | binomial test               | SAS        | Stata        | SPSS        |
|  |                                 | categorical                     | Chi-square goodness-of-fit  | SAS        | Stata        | SPSS        |
| 1 IV with 2 levels<br>(independent groups)         |                                 | interval & normal               | 2 independent sample t-test | SAS        | Stata        | SPSS        |
|  |                                 | ordinal or interval             | Wilcoxon-Mann Whitney test  | SAS        | Stata        | SPSS        |
|  |                                 | categorical                     | Chi-square test             | SAS        | Stata        | SPSS        |
|  |                                 |                                 | Fisher's exact test         | SAS        | Stata        | SPSS        |
| 1 IV with 2 or more levels<br>(independent groups) |                                 | interval & normal               | one-way ANOVA               | SAS        | Stata        | SPSS        |
|  |                                 | ordinal or interval             | Kruskal Wallis              | SAS        | Stata        | SPSS        |
|  |                                 | categorical                     | Chi-square test             | SAS        | Stata        | SPSS        |
| 1 IV with 2 levels<br>(dependent/matched groups)   |                                 | interval & normal               | paired t-test               | SAS        | Stata        | SPSS        |
|  |                                 | ordinal or interval             | Wilcoxon signed ranks test  | SAS        | Stata        | SPSS        |
|  |                                 | categorical                     | McNemar                     | SAS        | Stata        | SPSS        |

Figure 1. UCLA's Academic Technology Services' website offers many useful resources, including this statistical analysis table.

## UCLA ACADEMIC TECHNOLOGY SERVICES' WEBSITE

UCLA's Academic Technology Services website cited above (<http://www.ats.ucla.edu/stat/>) is a rich, public resource for SPSS, SAS, and Stata users, with helpful documentation and links.<sup>1</sup> It also includes a wonderfully useful table entitled "What statistical analysis should I use?" (see figure 1). Using the number and type (categorical or not) of dependent and independent variables you have in your model, you can find the recommended statistical test, and links to examples of how to do the test with SPSS, SAS, or Stata. A discussion of the results and references for further study are also available.

Other resource pages, many of them cross-linked to the indices, include code fragments for advanced users, statistical papers and bibliographies, and organized sections of links dealing in great detail with statistical models. For example, in the "What statistical analysis should I use?" table mentioned above, you could click on the SPSS link for the Repeated Measures test. You could then download the sample data, follow the instructions for running the GLM (General Linear Model) command, then compare your results to those shown on the page. There is also a discussion of the results and links to other code and documentation. The site provides information on the prerequisite knowledge you would need, or, for more advanced statisticians, confirmatory details on the parameters calculated.

### MIT'S OPENCOURSEWARE

OpenCourseWare (<http://ocw.mit.edu/>) is "a free and open educational resource for faculty, students, and self-learners around the world" that provides MIT lecture notes, assignments and tests for many courses. Courses useful to quantitative researchers range from introductory to advanced statis-

The screenshot shows the MIT OpenCourseWare website. At the top, there are navigation links: COURSE LIST | ABOUT OCW | HELP | FEEDBACK. The main heading is MITOPENCOURSEWARE MASSACHUSETTS INSTITUTE OF TECHNOLOGY. Below this is a welcome message: "Welcome to MIT OpenCourseWare a free, open publication of MIT Course Materials. We invite you to view all the courses available at this time." On the left, there is a search bar and a list of available courses categorized by department, such as Aeronautics and Astronautics, Anthropology, Architecture, Biological Engineering Division, Biology, Brain and Cognitive Sciences, Chemical Engineering, Chemistry, Civil and Environmental Engineering, Comparative Media Studies, Earth, Atmospheric, and Planetary Sciences, Economics, Electrical Engineering and Computer Science, Engineering Systems Division, Foreign Languages and Literatures, and Health Sciences and Technology. The main content area features a "Welcome to MIT's OpenCourseWare:" section with a list of bullet points: "Is a publication of MIT course materials", "Does not require any registration", "Is not a degree-granting or certificate-granting activity", and "Does not provide access to MIT faculty". Below this is a "Learn more about MIT OCW..." link. There is also a section titled "Investing in Open Sharing" with a photo of Jon Gruber and text stating that he has donated \$1 million to the project. At the bottom, there are links for "Course List", "About OCW", "Help", and "Feedback".

Figure 2. MIT's OpenCourseWare website, <http://ocw.mit.edu/>, offers public access to hundreds of online MIT courses in statistics.

tics. At the elementary end of the spectrum, for example, is *Statistics for Applications*, a "broad treatment of statistics." At the more advanced end is *Introduction to Modeling and Simulation*, an "overview of modeling and simulation tools, as well as case studies in modeling and simulation."

MIT President Susan Hockfield states that "through MIT OpenCourseWare (OCW), educators and students everywhere can benefit from the academic activities of our faculty and join a global learning community in which knowledge and ideas are shared openly and freely for the benefit of all."<sup>2</sup>

OCW, which is funded jointly by the William and Flora Hewlett Foundation, the Andrew W. Mellon Foundation, and MIT, was launched

on April 4, 2001. Its mission is to expand access to information to students and teachers around the world, and to counter the trend towards the privatization of knowledge. The site currently offers more than 500 courses, and plans to include 1800 courses by 2008.

The MIT site, along with the UCLA statistics site discussed above, are good examples of the level of maturity the Internet has reached. These sites are not merely academic papers published on a university's statistics website. They were not simply found by Googling a topic of interest and getting one treatment of a subject out of context. They are organized archives, rich with context, which allow you to step backwards or forwards in a question to learn impor-

1. For example, the site points to very handy online SAS documentation at <http://v8doc.sas.com/sashtml/>.

2. <http://ocw.mit.edu/>

tant prerequisites or expert uses. Statistics scholars would be well served to visit these sites often.

### MINITAB RELEASE 14

Minitab, a well-respected 30-year-old statistical package, has just released a new version. It was originally developed at Penn State with contributions from many talented statisticians. Minitab has a feel similar to SPSS or Stata and is used at NYU for several introductory statistics courses. To take a tour of the new release, which offers increased graphics and graphing capabilities and a customizable interface, visit the Minitab website at <http://www.minitab.com/products/minitab/14/> (see figure 3).

One immediate advantage of Minitab over SPSS, Stata, and SAS is the cost. A semester rental of the student version is available on the Internet for only \$30 and is sometimes included with textbooks.<sup>3</sup> By comparison, NYU students can purchase:

- Stata for \$130 (<http://www.nyu.edu/gsas/dept/politics/datalab/>)
- SPSS student version for \$200 (<http://www.bookstores.nyu.edu/computer/>)

- SAS for \$75 per year after an initial fee of \$175 the first year (<http://www.nyu.edu/its/sas.html>)

All of these packages are also available for use at the ITS Third Avenue North computer lab (<http://www.nyu.edu/its/labs/third.html>). NYU presently has a site license that allows Minitab to be installed at no cost on University-owned computers. Minitab does not allow for the easy transfer of datasets between statistical packages, so additional software that facilitates this process (such as StatTransfer) should be installed on the computer to complement Minitab.

Minitab does have certain limitations that prevent it from being as useful as SPSS, Stata, and SAS for packaging, managing and documenting data.<sup>4</sup> Nonetheless, the fact that it is easy to learn Minitab and transfer that knowledge to other programs makes it a good choice for introductory courses.

\*

*If you have questions about any of the resources described here, please send e-mail to: frank.lopresti@nyu.edu.*

*Frank LoPresti heads the ITS ACS Social Sciences, Statistics & Mapping Group.*

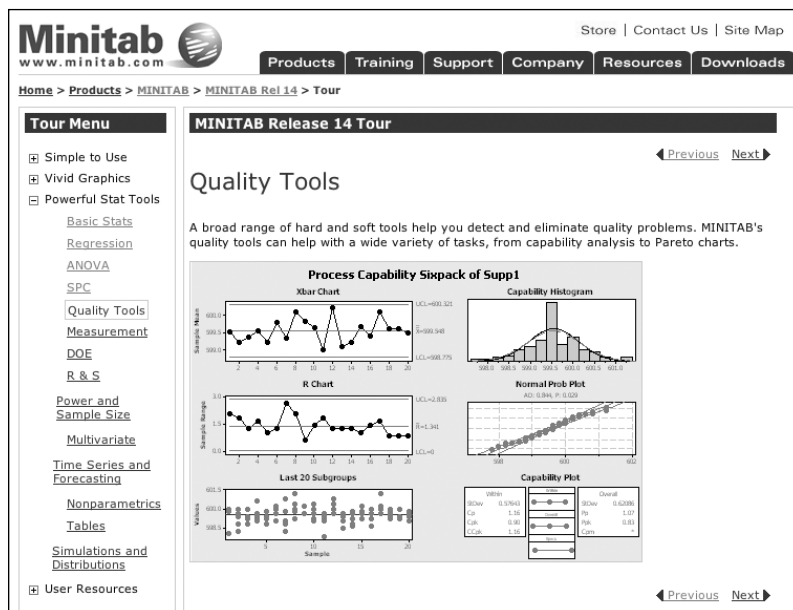


Figure 3. Tour Minitab 14 at <http://www.minitab.com/products/minitab/14/>.

3. <http://www.minitab.com/education/semesterrental/>

4. For example, the lack of labels for individual values of a variable (i.e., 0 = "Male", 1 = "Female").

## Did You Know?

### ITS Statistics Forum

ITS offers a free Statistics and GIS e-mail forum, which distributes occasional news and information of interest to the statistics community. To subscribe, simply send a blank e-mail message from your preferred e-mail address to: [subscribe-statistics@forums.nyu.edu](mailto:subscribe-statistics@forums.nyu.edu). Or, you can subscribe to this and other NYU Forums through NYUHome. To do so, simply log into NYUHome at <http://home.nyu.edu>, then, in the Forums channel (within the Home Tab) click the "Subscribe to an NYU Forum" link. On the page that opens, click the checkbox next to "Statistics," then click "Subscribe" at the bottom of the page.

### Statistics Classes & Clinics

Each semester, ITS offers a variety of free statistics and mappings classes for the NYU community. The current listing of classes is available on the ITS website at <http://www.nyu.edu/its/classes/>.

In addition, the ITS Social Science, Statistics & Mapping Group regularly hosts Friday GIS Clinics, where GIS professionals come to the ITS Third Avenue North computer lab to share their knowledge with NYU attendees. For more information about these clinics and the upcoming schedule, subscribe to the Statistics forum described above, or send e-mail to [frank.lopresti@nyu.edu](mailto:frank.lopresti@nyu.edu).