

Introduction to Multivariate Methods

Course Information Sheet (corrected)

Instructor Christopher Bingham,
 372 Ford Hall, Minneapolis Campus, phone: 612-625-1024
 146E Classroom Office Building, St. Paul Campus, phone: 612-625-7023
 Email: kb@umn.edu (*text mail only please*; HTML will not be read)
 Office hours: Monday, Wednesday, Friday 1:30 - 2:30 or by appointment

Teaching Assistant Edgar Maboudou, email: edgar@stat.umn.edu
 Office hours: in Ford 352 to be determined, phone 612-625-2858

Required Text

Applied Multivariate Statistical Analysis, 5th Edition by Richard A. Johnson and Dean W. Wichern, Prentice Hall, 2002

An Introduction to MacAnova, available on the Web

(<http://www.stat.umn.edu/macanova/documentationintro.html>)

Other books that may be helpful

Multivariate Statistical Methods, 3rd Edition

by Donald F. Morrison, McGraw-Hill, 1990

Linear Statistical Inference and its Applications, 2nd Edition

by C. Radhakrishna Rao, Wiley, 1973

An Introduction to Multivariate Statistical Analysis, 3rd Edition, by T. W.

Anderson, Wiley 2003 (new edition of a classic of mathematical statistics)

Course Web Site <http://www.stat.umn.edu/~kb/classes/5401>

This will have links to current assignments, data files, MacAnova downloads pages, and most handouts. There is also a link to a compilation of questions asked about MacAnova. Look here for late breaking announcements. Handouts may be distributed *only* on the web.

Reading

There will be specific reading assignments every week or so, some of them in handouts that will be distributed in class or posted on the web site. I will try to give these at least a week ahead. After the first week, reading should be done ahead of time. I will not necessarily lecture about everything in the reading, nor will everything I say in class be covered in the reading.

Don't hesitate to use email (text only; not HTML) to ask questions about the reading, lectures or homework. If the answer would be difficult to give by email, I may ask you to come to office hours. When a question might be of interest to the whole class, I will post edited copies of the question and answer on the class web page.

Written Assignments

There will be written assignments most weeks. These usually involve numerical analysis of data using a computer or numerical exploration of the properties of procedures. These form an important part of the course and will count substantially in the final grade. They will normally be due in class on Mondays (the first Assignment is due on Wednesday, September 14). Late assignments may be refused or, if graded, will lose points for lateness. The T.A. will try to get them back by Friday. Not all problems may be graded.

With each assignment, you must attach a signed affirmation of academic honesty, stating that the work is your own. Copies will be provided to you. *Homework received without such an affirmation will not be graded.* Copies will be distributed in class or can be downloaded from

<http://www.stat.umn.edu/~kb/classes/5401/files/HonestyAffirmation.pdf>

In grading the assignments, emphasis will be placed on obtaining correct and/or reasonable answers. The organization of ideas and clarity and conciseness of presentation is also important. What you turn in should be prepared on a word processor such as Microsoft Word, incorporating selections from computer output including graphs with your discussion. You should provide enough output so that we can see how you obtained your answers. It's often better to put in an appendix the output that documents how you got your answers (as opposed to the answers themselves). Homework formatted using Latex is also acceptable.

Computer output included in your work should be fully annotated. In particular, informative axis labels are required on all graphs and units of measurement should be given where known.

Raw computer output without annotation or explanation is not acceptable.

In a word processor document, any computer output *must* be in **equal width characters** as provided by a non-proportional font such as Courier or Monaco. Otherwise, columns of numbers will not line up and some graphic displays will be gibberish.

Homework may *not* be submitted by email. In a pinch, submit it by fax, with a cover sheet to indicate it goes to me or the TA. Also send an email that you have faxed your homework.

Examinations

Currently I am scheduling two exams, a midterm and a final. The **midterm** is tentatively scheduled for **Monday, October 24** and will be open book. The **final** will either be a two hour open book exam at the regularly scheduled time (**10:30 am - 12:30 pm Saturday, December 17**), or (more probably) will be a take home exam. If the final is a take home exam, it must be submitted in paper form; email will not be accepted.

Statistics 8401 project

If you are registered in Stat 8401, you are expected to complete a **project**, in addition to the other work in the course. A project normally is completed by learning about and mastering a multivariate statistical method that is not covered in the course.

Stat 5401/8401 Course Information Sheet

You will find data to which it can be applied and write a report summarizing the methodology and the results of an analysis. An alternative project would be to write a MacAnova macro package that provides an informative “black box” interface to one of the analyses such as Principal Components Analysis or Factor Analysis covered in the course.

In addition, you will make a 25 minute presentation to the class in late November or early December.

Please meet with me before October 1 to discuss possible topics. A written outline will be due by November 14, a first draft by November 29 and the final copy by December 19.

Grading

The final grade will be based on the homework and exams in approximately the proportions 30%-30%-40% (HW, MT, F) for Stat 5401 and 25%-25%-30%-20% (HW, MT, F, Project) for Stat 8401. For those taking a course S/N, “S” corresponds to “C” or better.

It is the policy of the School of Statistics and the College of Liberal Arts that an incomplete (grade of “I”) is to be given only for compelling reasons, and then only after a “contract” is signed specifying how the “I” will be made up with a firm deadline for completion. In particular, being about to get a low grade or not having attended class is not a valid reason for being granted an “I”.

Computer use

The principal computer program to be used is MacAnova, an interactive statistical analysis program. It is programmable so it has the capability of adding statistical analyses that are not built-in. There are versions for Macintosh, Windows, DOS and Linux. The current version level is 5.05. If you have a previous version, *you should upgrade* to the latest since handouts and other documentation may use features not available on earlier version. See below for information on obtaining MacAnova.

For Windows, Linux, and Macintosh OS X, there is a new interface in which commands are typed in one panel of a window and output goes to another panel. As yet, most documentation does not reflect this. In the Classic Macintosh OS version and previous Windows version, commands are typed and output printed in an editable text window.

On the Statistics workstations, you can start the new version of MacAnova by typing `macanovacpc` at the Linux prompt.

Even though you may have access to or experience with other programs that can do equivalent computations, *you should use MacAnova*. Specifically I have been asked in the past if it would be o.k. to do the homework using SAS or SPSS. Although I would encourage anyone with easy access to SAS, SPSS or other statistics program to attempt to duplicate MacAnova results, only MacAnova computer output may be turned in.

I use MacAnova because it allows what I call a “white box” approach to computing in which you directly use definitions and formulas to specify what is computed.

This can be contrast with a “black box” approach where you a complete analysis is done without any clue as to how it is done. As an example, in MacAnova, a white box approach to computing a sample mean is to type the command `xbar <- sum(x)/n`; a black box approach would be `xbar <- describe(x,mean:T)`.

In the Macintosh and Windows versions (but not Linux version) you can copy high resolution graphs to the “clipboard” for pasting into a word processor document. You can directly print graphs using **Print** or **Print Window** on the **File** menu. You can also save graphs in various formats usin **Save window as** on the **File** menu. All versions allow you to save PostScript versions of graphs using keyword `file` on plotting commands.

If you have difficulty including high resolution graphs in homework, you can make low resolution graphs composed entirely of printable characters. You do this by including keyword phrase `dumb:T` on plotting commands. See the handout *Making Plots Using MacAnova* (<http://www.stat.umn.edu/~kb/classes/5401/files/MVGraphics.pdf>) for information on making graphs in MacAnova. This will be revised in the next few days.

All versions implement the `spool()` command. This records everything you type and everything MacAnova prints, including low resolution but not higher resolution plots, in a file. Type `help(spool)` for detailed information or see *An Introduction to MacAnova* (see URL below). This is one way to preserve your work for including in homework. Type `help(spool)` in MacAnova for details.

To get started in MacAnova, you should read through *An Introduction to MacAnova* while running MacAnova. Try to reproduce the output in the many examples and understand why you get it.

There is extensive on-line or browsable help in MacAnova. In addition there is the *MacAnova User's Guide* (<http://www.stat.umn.edu/macanova/documentationug.html>). This does not include features added since version 4.07 but is otherwise quite complete. This is not a good document for learning MacAnova as a novice but may be useful for learning some of the details, especially how to write macros.

There will be handouts illustrating various aspects of the use of MacAnova.

Obtaining MacAnova

The normal place to obtain MacAnova is from the MacAnova download page <http://www.stat.umn.edu/macanova/download.html>.

When you get to a page from which you actually download MacAnova, please read the instructions in file `Install.msw.txt`, `Install.osx.txt` or `Install.txt` carefully before downloading. For both Windows and and Macintosh OS X, there is just one file you must download. The current version for Macintosh Classic is 4.13. This will be upgraded soon. You may also want to download from the same page the HTML formatted help which you can read in a web browser.

On a Windows computer, you should use the **right mouse button** to select each file and then select Download link from the popup menu. On a Macintosh, just clicking the link should work. On both types of computers, I find the desk top is a convenient download destination.

The file downloaded for the Windows version is an installer. See `Install.msw.txt` for details on how to use it. The default choices in the installer dialog boxes are probably OK for most people. The program itself and associated help and macro files are installed and shortcuts are added to the program menu.

For the Macintosh version the single file is a binhexed executable archive. If your browser is configured appropriately, the downloaded files should automatically be expanded with all the needed files in a folder named something like `MacAnova4.13R1` possibly inside another folder with a name like `MV413ppcAll.1 Folder`. If it's not automatically expanded, use Stuffit Expander to unpack it. You can leave folder `MacAnova4.13R1` on your desktop or move it someplace else where you can find it. All the files in `MacAnova4.13R1` should be kept together.

If you have difficulty downloading or installing MacAnova please let me know immediately.

Obtaining data files

Data files for use in the course are available on the web. The principal file is `JWData5.txt`, a file of all the data sets in Edition 5 Johnson & Wichern in a format readable by MacAnova command `read()` or `matread()` but *not* by `readdata()` or `readcols()`.

I strongly suggest that you use `JWData5.txt` rather than attempt to use the data files on the CDROM that comes with the text. A number of them cannot be read easily by MacAnova. There is a direct link to `JWData5.txt` on the class web page. Click on Johnson & Wichern Fifth Edition Data files in MacAnova format (<http://www.stat.umn.edu/~kb/classes/5401/files/data/JWData5.txt>) in your browser. This will display the file in a window. Then choose **Save As** on the **File** menu to write it to your disk.

From time to time there will be other files used.

You can directly access all data files through the web page. You may want to download all the data files now, although we may not use them all and may use others that will be installed in the future. You can browse all the data and special macro files at <http://www.stat.umn.edu/~kb/classes/5401/files/>.

Data file names generally end in `.txt`, macro files end in `.mac` or `.mac.txt`, and any help files end in `.hlp` and `.hlp.txt`. The `.mac.txt` and `.hlp.txt` files are the same as the files without `.txt`. They are there because experience has shown that certain browsers have difficulty correctly downloading text files that don't end in `.txt`.

Adobe Acrobat (PDF) versions of assignment sheets and other handouts will be placed in this same location.

For work in the labs, it is expected that you will have data sets on your own diskettes or download them while you work.

Misconduct

Students are *strongly encouraged* to work together on the homework provided all mutual collaborators acknowledge the collaboration on their papers (just put on the first page who you worked with). However, *the actual preparation of answer papers and computation must be done by each student individually, without copying from another*. If these rules are violated on a homework assignment, that assignment will receive a grade of 0. In addition, the incident will be reported to the student's college or department. As stated above, with each assignment, you must attach a signed affirmation of academic honesty, stating that the work is your own.

On examinations, whether in-class or take-home, there must be *no* sharing of work or communication with anyone except the instructor. Any violation will result in a 0 for the entire examination and will be reported to the offender's college and/or department.

All course materials are available in alternative formats upon request. Please contact Dana Tinsley, Applied Statistics, Ford 313, 612-625-8046.