CSC 328/428 Data Analysis for Experimenters
Summer Session I – 2002

Instructor: Sharon Song, Ph.D.
Office: CST 611
Phone: CTI: (312) 362-8269 / Alternate: (847) 409-9308
Fax: (312) 362-6116
Email: ssong@cs.depaul.edu
Web: http://www.depaul.edu/~ssong1/Home.html

Required Text


Optional Texts


Topics

1. Analysis of variance and covariance in a variety of settings. Multiple comparisons. Use of PROC ANOVA versus PROC GLM.


3. Matrix notation and the general linear model.

The emphasis of the course will be on how to perform selected data analyses using the Statistical Analysis System (SAS®). SAS procedures (PROC’s) to be covered include ANOVA, FORMAT, GLM, IML, MEANS, NPARIWAY, PLOT, PRINT, RANK, REG, SORT, and UNIVARIATE. Statistical and mathematical issues will be covered that assist the student in understanding the models underlying the statistical PROC’s and in interpreting the output of those PROC’s.

Grading

A student’s grade will be based on his/her rank in the class using total points earned. Points will be assigned to homework problems and to the final examination.

Due dates for the assignments will be assigned after the associated material has been covered in class. Homework that needs to be reworked may receive a reduced score; students are
expected to request sufficient input on the assignments prior to the due dates. ALL HOMEWORK MUST BE HANDED IN TO THE INSTRUCTOR IN CLASS OR SENT BY REGULAR OR CAMPUS MAIL ON OR BEFORE THE ASSOCIATED DEADLINE. Please do not fax homework since there is no guarantee that faxes will be properly routed to me. Homework by electronic mail is acceptable in most cases.

Students in COL/Distance Learning sections have different deadlines. They should contact me for details.

Grading

There will be 5 data analysis projects worth a total of 120 points. There will also be a final examination worth 40 points or 25% of the total 160 points. Revisions to the assignments and to the point allocation are possible and will be addressed on an individual basis.

Final grades will be based on percentages of total points on the following scale:

<table>
<thead>
<tr>
<th>Percentage Range</th>
<th>Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>95%-100%</td>
<td>A</td>
</tr>
<tr>
<td>90%-95%</td>
<td>A-</td>
</tr>
<tr>
<td>87.5%-90%</td>
<td>B+</td>
</tr>
<tr>
<td>82.5%-87.5%</td>
<td>B</td>
</tr>
<tr>
<td>80%-82.5%</td>
<td>B-</td>
</tr>
</tbody>
</table>

or on one for which the final grade may be higher that this scale indicates. Scores below 80% indicate unacceptable performance and actual grades will be determined on a case-by-case basis.

Homework Assignments

It is possible to analyze data from other source than the suggested data sets, but consult the instructor first.

1. Analyze a one-way analysis of variance data set with at least 5 treatments/levels. Possible data sets include Ott exercises 13.5, 13.8, 13.25, 14.5, 15.29 (analyze data for the first 5 or more medications ignoring the investigator factor). 19 points.

2. Analyze a regression data set with at least 3 predictor variables. Possible data sets include Ott Table 12.1 (predict volume); Ott exercises 12.3 (predict crime), 12.48 (predict price), 12.55 (predict risk), and 12.59 (predict change in pulse). 29 points.

3. Analyze a 2-factor analysis of variance data set with 2 levels for each factor. Possible data sets include any 2-factor data set in the Ott textbook (adjust to two levels per factor: 15.25 with pH at 4.0 and 5.0 and calcium at 100
and 200; 15.29 investigator at 1 or 2 and medication at A or B; 15.36 with level of reinforcement at none or verbal and time of isolation at 20 and 40; 15.42 or 15.45 with laboratory at 1 or 2 and time at 1 or 3 months with response of either mg/ml of active ingredient or of pH).

19 points.

4. Analyze a 2-factor analysis of variance data set with 2 levels for one factor and at least 5 levels for the other factor. Possible data sets include any 2-factor data set in Ott with one factor restricted if necessary to at least 5 levels and the other factor restricted if necessary to 2 levels (15.3, 15.12, 15.19, 15.29, 15.33, 15.49).

29 points.

5. Conduct an analysis of covariance. Possible data sets will be provided in class. 24 points.

Final Examination

The final examination will be a combination of a take home and a written examination. The instructor will provide students with a programming assignment, which they implement and run prior to the final examination. Students then answer questions during the final on that code, its output, and related data analysis concepts using any notes they choose. The written portion of the exam will take place during the tenth class on Thursday July 18, 2002.