Chemical Engineering 4C3/6C3, Winter 2011
Statistics for Engineering

Instructor Kevin Dunn, dunnkg@mcmaster.ca (no office on campus)

Teaching assistant Ian Washington, Room JHE-369, ext 24031, washinid@mcmaster.ca

Class time and location Hamilton Hall 109. Monday, Wednesday and Thursday, 10:30 to 11:20

1 About the course

Official description Linear regression analysis in matrix form, non-linear regression, multi-response estimation, design of experiments including factorial and optimal designs. Special emphasis on methods appropriate to engineering problems.

What you must be able to demonstrate by the end of the course

- Understand that all data has variability: we want to separate that variability into information (knowledge) and error (unknown structure, noise, randomness).
- Understand and use process monitoring charts and confidence limits.
- Least-squares models: how to fit and especially how to interpret it, understand the confidence limits, understand the model's limitations.
- Be able to design your own experimental program and then also interpret the experimental data.
- Understand the principles of latent variable methods for engineering data.

Prerequisites A basic course in statistics that covers probability, means, variances, confidence intervals and linear regression. However, all these topics are covered again in this course, focusing on their practical application to engineering problems.

Course materials The course website will be permanently available: http://stats4eng.connectmv.com

Course materials, assignments and solutions will be available from the website. Course announcements will be posted to the main page of the website and to the course's Twitter feed, http://twitter.com/stats4eng - students are expected to check the website at least 3 times per week.

Required textbook There is no official course textbook. We will be using the instructor's own material from his book, Process Improvement using Data. The book was written specifically for this course, and each section will be available as a PDF from the course website. It is your responsibility to print out these notes and bring them to class. The *Titles Bookstore* will have a limited number of printed copies, available for the cost of just printing the PDF, if you prefer not to print it yourself.

Recommended readings If you would like to buy one book to supplement the course material, I highly recommend the first one, Box Hunter and Hunter.

1. G.E.P. Box, J.S. Hunter, and W.G. Hunter, *Statistics for Experimenters - Design, Innovation and Discovery*, 2nd edition, Wiley. ISBN: 978-0471718130.
2. N.R. Draper and H. Smith *Applied Regression Analysis*, Wiley.
3. D.C. Montgomery and G.C. Runger, *Applied Statistics and Probability for Engineers*.

Other reference texts are listed on the course website and are generally available in Thode Library.

Course outline The course is divided into 6 main sections, taught over 12 weeks, 3 classes per week.

1. *Visualizing data*: creating high-density, efficient graphics that highlight the data honestly.
2. *Univariate data analysis*: Probability distributions and confidence intervals
3. *Process monitoring*, aka statistical process control (SPC), for tracking process behaviour.

4. *Least squares regression modelling*: correlation, covariance, ordinary and multiple least squares models. Enrichment topics will be covered, time permitting.
5. *Design and analysis of experimental data* and response surface methods for continual process improvement and optimization.
6. Introduction to *latent variable modelling*: a general overview of latent variable models and their use in (chemical) engineering processes.

Several enrichment topics are covered throughout the course: robust methods, cross-validation for model assessment, nonparametric methods, real-time application of the above methods, correlation and causality, and missing data handling.

2 Grading

To assess your understanding of the course materials, the grading for the course will be:

Component	Fraction	Notes
Assignments	20%	Expect around 7 assignments; can be completed individually, or in groups of 3 or less (4C3), or in groups of 2 or by yourself (6C3).
Midterm exam 1	10%	A 2 hour written exam, on 17 February, before the midterm break.
Midterm exam 2	25%	A take-home exam, using the course software over a 5-day period. Also includes an experiment that you have to do yourself (ahead of time) and analyze. Due on 24 March.
Final exam	45%	A written exam, lasting 3 hours.

6C3 students will have extra questions on all assignments and exams.

Policies regarding grading

- We encourage you to complete the assignments in groups of no more than 3 members. The 6C3 students may also work in groups of at most 2 members.
- You, and your group, will receive the greatest benefit if you each do **all** the questions yourselves. Arrange to meet and review your solutions, discussing various approaches.
- Assemble a **single submission** for the group - the TA will not grade loose sheets handed in after the first submission. All group submissions must clearly show the names of the group members.
- You are defeating the purpose of the group-based assignment if you simply divide the assignment into sections, one for each group member. This is definitely not recommended, because you are loosing out on the learning opportunity of seeing your mistakes and the group members' mistakes, and learning from them.
- No sharing of any work may be done between groups for assignments and take-home exams. This includes handwritten documents and electronic files of any type. This will be strictly enforced. Please ensure that you have read the University's academic integrity policy (reproduced below).
- This is a large class of about 85 students, so late hand-ins interfere with the TA's ability to efficiently grade your assignments. Late assignments will be penalized by deducting 20% per day for every late day. A grade of zero will be given for submissions handed in after the solutions are posted (usually within 2 days of assignment hand-in).
- Emergencies and such arise, so each person has 2 "late day" credits for assignments. So you can hand in one assignment 2 days late, or 2 assignments each one day late, without penalty.
- Grading of assignments and tests will include contributions for clarity and organization of presentation.
- No make-ups will be given for assignments.
- Any textbook(s) and course notes will be allowed during tests and exams.

- No make-ups will be given for mid-term tests. If due to some legitimate reason (in case of medical issues, please provide a medical certificate) the student cannot participate in a test **and** provides a missed-work form, the contribution of that test to his/her overall grade will be added to the final examination weighting; for unexcused absences, a zero grade will be given.
- All assignments will be graded, and the mean of the best $N - 1$ assignments used to calculate the assignment grade. You should expect $N \approx 7$, and the assignments will be frequent at the start of the course.
- Any calculator may be used during the tests and exams.
- The final percentage grades will be converted to letter grades using the Registrar's recommended procedure.
- Adjustment to the final grades may be done at the discretion of the instructor.
- The take-home exam tests your ability to use computer software to help complete the question. The 4th year chemical engineering lab has the course software installed in the event that you do not have access to a computer. You may complete this exam in groups of 3 students or less, while 6C3 students may complete the exam on their own or with one other person.
- The final exam will be cumulative, based on the entire semester's material.

3 Important notes

Class participation Please bring a calculator to every class.

Course software Use of a computer is required in the course. The R-language (http://www.r-project.org/) will be used, and is a freely available software package that runs on Linux, Mac and Windows computers. The software is available in the 4th year Chemical Engineering computer labs. Minitab (you can rent a 6-month version very cheaply), MATLAB, or Python may be used as well; you should not use Microsoft Excel. Where time permits, the TA and the instructor will post solutions in these languages. More details are posted on the course website.

Out-of-class access Since the course instructor does not have an office on campus, office hours will be before and after the class time, or arranged by appointment.

The TA for this course can be contacted by email - please see his addresses above. Only send email from your McMaster email address - we cannot respond to personal email addresses.

Disclaimer The above outline **may be modified**, as circumstances change, with agreement from the class.

4 Academic integrity

You are expected to exhibit honesty and use ethical behaviour in all aspects of the learning process. Academic credentials you earn are rooted in principles of honesty and academic integrity.

Academic dishonesty is to knowingly act or fail to act in a way that results or could result in unearned academic credit or advantage. This behaviour can result in serious consequences, e.g. the grade of zero on an assignment, loss of credit with a notation on the transcript (notation reads: "Grade of F assigned for academic dishonesty"), and/or suspension or expulsion from the university.

It is your responsibility to understand what constitutes academic dishonesty. For information on the various types of academic dishonesty please refer to the Academic Integrity Policy, located at http://www.mcmaster.ca/academicintegrity

The following illustrates only three forms of academic dishonesty:

1. Plagiarism, e.g. the submission of work that is not one's own or for which other credit has been obtained.
2. Improper collaboration in group work: this point is particularly important and will be strongly penalized in this course.
3. Copying or using unauthorized aids in tests and examinations.

5 Important dates

A list of *tentative* dates:

Date	Description
3 January 2011	Overview class: review of course content and administrative issues
5 January	1. *Data visualization* section starts
10 January	Assignment 1 due
10 January	2. *Univariate data analysis* section starts
17 January	Assignment 2 due
24 January	3. *Process monitoring* section starts
24 January	Assignment 3 due
3 February	4. *Least squares modelling* section starts
3 February	Assignment 4 due
14 February	Assignment 5 due
17 February	**Written midterm**
21-25 February	Break
2 March	Assignment 6 due
3 March	5. *Design and analysis of experimental data* section starts
10 March	Assignment 7 due
24 March	**Take-home exam due** (includes an experimental project)
28 March	6. *Latent variable methods* section starts
4 April	Final review class