Chemical Engineering 4C3/6C3, Winter 2014
Statistics for Engineering

Instructor

Kevin Dunn, kevin.dunn@mcmaster.ca (BSB, room B105)

Teaching assistants

Yanan Cao, Room JHE-369, ext 24013, caoy4@mcmaster.ca

Shailesh Patel, Room JHE-369, ext 24031, patelsr@mcmaster.ca

Class time and location

MDCL, room 1110. Monday, Wednesday and Thursday, 16:30 to 17:20. First class: Monday, 06 January.

Disclaimer

This outline **may be modified**, as circumstances change.

1 About the course

Official description

Univariate statistics and process monitoring. Linear regression. Experiments: full and fractional factorial designs. Introduction to latent variable methods and other current statistical tools. Applications to relevant engineering problems. Interpretation of computer-based output.

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).
- Interpret confidence intervals and univariate data statistics (mean, median, histograms, significant differences).
- Understand and use process monitoring charts.
- Least-squares models: how to fit and especially how to interpret them, understand the confidence limits and model limitations.
- Be able to design your own experimental program and then also interpret experimental data.
- Understand the principles of latent variable methods for engineering data.

Note: there are no tutorials scheduled for 4C3/6C3. There is a however a significant chunk of time you will have to invest outside of classes to learn new computer software. Web-based tutorials will be provided for you to complete this self-directed learning in your own time.

Prerequisites

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

Course materials

The course website will be permanently available at http://learnche.mcmaster.ca/4C3. You may use this as a resource even after you graduate.

Course materials, assignments and solutions will be available from the website. Course announcements will only be posted to the main page of the website - 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 will be available as a PDF from the course website. It is your responsibility to print out these notes and bring them to class.

Recommended readings

If you would like to buy one book to supplement the course material, I highly recommend the Box Hunter and Hunter book for its practical engineering perspectives on data analysis.

> G.E.P. Box, J.S. Hunter, and W.G. Hunter, *Statistics for Experimenters - Design, Innovation and Discovery*, 2nd edition, Wiley. ISBN: 978-0471718130.

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.

1. *Visualizing data*: creating high-density, efficient graphics that highlight the data honestly.
2. *Univariate data analysis*: Probability distributions and confidence intervals
3. *Least squares regression modelling*: correlation, covariance, ordinary and multiple least squares models. Enrichment topics will be covered, time permitting.
4. *Design and analysis of experimental data* and response surface methods for continual process improvement and optimization.
5. *Process monitoring*, aka statistical process control (SPC), for monitoring process behaviour.
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, missing data handling, Bayesian methods, nonlinear regression.

2 Grading

To assess your understanding of the course materials, the grading for the course will be assessed as described below.

Component	Fraction	Notes
Assignments	20%	Expect around 6 or 7 assignments; can be completed individually (mandatory for 6C3), or in groups of 2 or on your own (4C3).
Weekly tests	13%	Short weekly tests covering the material taught the previous week, as well as assigned readings.
Midterm exam	12%	A written exam
Experimental report	10%	An experimental report that you have to do with your group and analyze the data from. Due electronically on 31 March.
Final exam	45%	A written exam, lasting 3 hours.

6C3 students will have extra questions on all assignments, tests and exams. 6C3 students will also have extra readings and work assigned to their project.

Policies regarding the course and regarding grading

- *Very important note*: **Achieving a grade of below 50% in the final exam will automatically imply failure in the course, with a grade of F, no matter what your other grades in tests and assignments are.**

- Weekly tests: These tests may be answered from any device with internet access. These tests are *completed individually*. There will be multiple choice, short answer and long answer questions on the tests, with questions randomly generated per student. Many questions will be automatically graded and the results returned to you after completion of the test window. Some questions will require you to upload electronic files and images.

  The testing window is a period of time (approximately a 30-hour window). Once you start the test on your device you will only have an hour to complete it. The intention is that you study the material taught in class the week prior and then take the test. Questions will occasionally be from work covered several weeks prior to the test. Solutions will be revealed at the end of the window.

  The reason for using short, frequent tests is that there is ample evidence in the learning literature that the testing effect and that spaced repetition improve your retention and understanding of the material. This is especially true in 4C3/6C3 where the course structure has been carefully planned to be cumulative. Click on the links for some basic background reading.

  The other advantage of smaller tests are that they they lower the stakes and reduce the pressure on you, for example if you are having a bad week, it will only affect a small test of around 1%, as opposed to a midterm or final exam which has much greater stakes.

- The midterm *is optional* and there is no make-up for it. If you choose not to write the midterm, or cannot write it due to illness or other reasons, then the usual approach will be followed: the percentage contribution from the midterm will be added to the final examination weighting.

- The weekly tests and assignments will also test your ability to use computer software to help complete the questions. The 4th year chemical engineering lab has the course software installed in the event that you do not have access to a computer.

- We strongly encourage you to complete the assignments in groups of no more than 2 members. The 6C3 students **must** complete assignments on their own.

- 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's 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 losing out on the learning opportunity of seeing your mistakes and the group member's mistakes, and learning from them.

- No sharing of any work may be done between groups for assignments. 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 (part of which is reproduced below).

- This is a large class of about 90 students, so late hand-ins interfere with the TA's ability to efficiently grade your assignments. Late assignments will be penalized by deducting **30%** 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 all work in this course will include contributions for clarity and organization of presentation. This is an important professional skill that you have now successfully developed since second year.

- No make-ups will be given for missed assignments. No make-ups will be given for missed weekly tests.

- Any paper-based materials (textbooks, notes, *etc*) are allowed during tests and exams. Electronic textbooks and resources are, unfortunately, not permitted at the final exam, but may be used at any other time during the course, including weekly tests.

- Any calculator may be used during the tests and exams.

- 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, slowing down at the end.
- 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 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, Apple 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

The instructor has office hours posted on his website, or other times may be arranged by email.

The TA's for this course can be contacted by email - please see their addresses above. Try to send email from your McMaster account - email from personal accounts are sometimes discarded by spam filters.

Disclaimer: The above outline **may be modified**, as circumstances change.

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 is below. Some changes will occur as the course progresses. Please check the course website at least 3 times per week for updates:

Date	**Description**
6 January 2014	Overview class: review of course content and administrative issues
15 January	Assignment 1 due
22 January	Assignment 2 due
29 January	Assignment 3 due
05 February	Assignment 4 due
12 February	**Written midterm**
17-23 February	Midterm break
3 March	Assignment 5 due
17 March	Assignment 6 due
28 March	*Kipling*
31 March	**Experimental project due**
07 April	Assignment 7 due
07 April	Review class
10 April	Exams start