

# Chemical Engineering 4C3/6C3, Winter 2015

## Statistics for Engineering

---

### Instructor

Kevin Dunn, [kevin.dunn@mcmaster.ca](mailto:kevin.dunn@mcmaster.ca) (BSB, room B105)

### Teaching assistants

Masoud Kheradmandi, Room JHE-370, [stats4eng@gmail.com](mailto:stats4eng@gmail.com)

Hadi Shahnazari, Room JHE-370A, [stats4eng@gmail.com](mailto:stats4eng@gmail.com)

### Class time and location

BSB, room 119. Tuesday and Friday (Wednesday class is cancelled - see below for why), 09:30 to 10:20.

First class: Tuesday, 06 January 2015.

### Disclaimer

This outline **may be modified**, by the instructor or university, as circumstances change.

#### **Please read carefully**

The 4C3 course is going to be very different to any other course you've ever taken at McMaster. It will be very different to how the course has been offered in the past. Any information you have heard from prior students will likely not be valid for 2015.

Please read the course outline carefully, as you may decide this course is not for you. This outline is long, because it provides details about how this course is different from other courses you have taken.

## 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: separate that variability into information (knowledge) and error (unknown structure, noise, randomness).
- Interpret confidence intervals and univariate data statistics (mean, median, histograms, significant differences).
- Use and react appropriately to process monitoring charts.
- Fit and interpret a least squares model, the confidence limits and the model limitations.
- Design your own experimental program and then interpret experimental data.
- Recognize the need for 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 briefly covered in this course, focusing on their practical application to engineering problems.

If you have done your undergraduate courses at McMaster, then the `STATS 3Y3` course will be sufficient.

## Course materials and website

The course website in prior years was at <http://learnche.mcmaster.ca/4C3>. You may continue to use this website as a resource after you graduate.

**However, for 2015 the official course website is at <https://mcmaster.coursera.org/>.** Please note that this is not an Avenue course website. Coursera is an external company that hosts over 1000 online courses (called MOOCs, or [massive open online courses](#)). MOOCs are an excellent way to keep your skills up to date as a form of self-directed learning. Many employers now recognize these credentials, and it is a great way to differentiate yourself on a resume. Parts of the 4C3/6C3 course were hosted as a MOOC on [Coursera in the summer of 2014](#).

You have to register to use the 4C3/6C3 website (it is not automatically done):

1. Visit this website: <http://yint.org/register>
2. Use any name you like; it does not have to be your real name. For example, you could use `jpstar876`. See below for the reason why. You can, and we hope you will, use your actual name.
3. Use your @mcmaster.ca email address to sign up (ignore the message about .cn emails)
4. Chose any password you like.
5. Hit “Sign up”; you might see an error message afterwards [HTTP 500]. That is OK, ignore it.
6. You will receive an email from Coursera to confirm your email address. Please click on the link in your email.
7. Once you have signed up, and **confirmed your email**, please email the course TAs at [stats4eng@gmail.com](mailto:stats4eng@gmail.com) with (a) your made-up username and (2) your student number.
8. It may take up to 24 hours for your account to be activated at <https://mcmaster.coursera.org>

Once activated, you will be able to access all course materials, assignments and solutions will be available from this new 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. If you want to automatically receive all course notifications, please subscribe to the course Twitter feed at [@stats4eng](#).

## Course textbooks

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.

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 structure

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, sequential methods.

## 2 How this course is different

A typical semester for a typical course at McMaster has 39 classes of teaching. In the 4C3/6C3 course in 2015 this will be reduced to a very small amount of time where Kevin gives lectures.

**So what has happened to the lecture time?** Let's see:

On average, lectures are about one half teaching of new concepts, and one half for questions, pauses, diversions and disruptions, starting late, ending early, exercises, worked problems and guest lectures. So 39 lectures of  $39 \times 50 = 1950$  minutes, of which about 50% is actual teaching, or 975 minutes, so  $16.25 \approx 17$  hours of actual teaching per term. These teaching hours have become videos, that you will learn from.

**When will you watch these videos?**

One lecture a week has been cancelled, which frees up 13 hours for you. The Wednesday morning class is removed. I know you may not actually watch the videos in that time. In fact, I don't expect you to. I expect that you will use that extra hour to work on lab reports, assignments, or catch up on sleep, errands, or whatever activities you decide. This time-shifts a free hour for you to watch the video later on, for example in the evening, or during your commute home on the bus.

**Why videos?**

The videos are carefully written and edited version of the class lectures, cutting out any distractions, pauses and errors. They are recorded in a professional studio. And because they have been edited and externally reviewed, they are of better quality than a regular lecture. In fact, it takes 3 to 5 times more effort, time and lots of money (over \$50,000 for this course) to produce online videos. The videos have also been closed-captioned, which really helps if English is not your first language.

In some instances, the videos have been used by over 8000 students last summer, in 2014. They have been improved based on their feedback, and continue to be improved, just like regular lectures.

Video lectures, of course, give you tremendous flexibility in scheduling your own time, but you have to self-disciplined and able to manage your time. This is a skill you have almost certainly developed at this point in your career. So we know that 4C3/6C3 is a suitable course to use these video lectures.

**What should I do when watching a video?**

Like a regular lecture, you should focus your attention on the video and the content. Create an environment free from distractions (as if you were in a regular classroom); turn off the TV, your cellphone and music.

The advantage of the video is of course that you pause when required, rewind and review parts a second time. If you use the video player in the course website (mobile, desktop or tablet versions) you can even watch at faster speeds to skip over parts you are already familiar with. For example, if you are reviewing a video for the final exam, you can watch it a  $2\times$  the regular speed.

**What do I have to do on the course forums?**

The course forums are a place for you to interact with your colleagues. A small percentage of the course grade, 5%, counts for participation. Each forum post you make (either a reply to someone else, or a question that you ask), counts 1% for each post. You are encouraged to make more than 5 posts. We have done this by allowing you to register on the course website with any name you like (e.g. jpststar876). High-quality posts are expected, as described in more detail on the course website.

People in the class often feel more comfortable asking a question in written form, where they have time to think about the question. It is also a great way for people who don't speak English as a first language to participate in the course.

Please use the forum to help out your friends in a way that is totally legal. (Do not directly post solutions here though).

### So what will happen in class time?

In class time we will be doing almost only exercises, case studies, mini-projects, quizzes, and collaborative group work. Every class will have something to hand-in at the end. Sometimes these will be individually-based, mostly they will be group-based. Sometimes you will require a computer in your group, mostly you can complete the work in class without one. Make sure one person in your group has a computer with the course software installed.

Preparing for the class time by watching the video and doing the assigned reading is critical to achieving a good grade in the course. Once you have done the assigned work before class you must take the pre-class quiz.

Occasionally a regular lecture will be given in class time, but this will be an exception, or a guest lecture. Because the classes are so interactive, they will not be recorded on video, since it will just look like a noisy tutorial session.

### Finally, is this course for me?

There is a significant chunk of time you will have to invest outside of classes. However, the amount of time is no more than a regular 3-credit course, i.e. you should be spending about 3 to 4 hours per week outside of classroom time (4C3), or about 5 hours for 6C3 students.

If you answer NO to any of these questions, you might want to consider dropping 4C3/6C3:

- Do you like being online or on a computer/tablet/smart phone?
- Are you willing to participate in group discussions and activities?
- Are you self-motivated and self-disciplined?
- Do you have the initiative to keep up with tasks, quizzes, and regular (almost daily) deadlines?

## 3 Grading

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

Component	Fraction	Notes
Pre-class quizzes	10%	Quizzes that are to be done before class time to demonstrate that you know the material
In-class activities	15%	Every class will have some short deliverable piece of work that is completed alone, or in a group of 3 people.
Assignments	10%	Expect around 6 or 7 assignments; can be completed individually (mandatory for 6C3), or in groups of 3 or on your own (4C3).
Midterm exam	10%	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 01 April.
Online forum participation	5%	Either posting questions or replying to others; 1% per post; must be of good quality to count.
Final exam	40%	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. The threshold to consider passing the course is higher, a B- grade, for 6C3 students.

**Very important note: Achieving a grade of below 50% in the final exam will automatically imply failure in the course.** The reason is simple: if you know your material, it is not difficult to obtain at least 50% on the final exam.

Please also note that students in their final year have failed courses: that you are about to graduate is not a reason to stop giving 100% effort before you reach the finishing line.

Note the above grade breakdown spreads your grades over multiple pieces of work. The advantage of this is to lower the stakes and reduce the pressure on you, for example if you are having a bad week, it will only affect a small part of your grade, as opposed to a heavily weighted midterm or final exam.

#### *Pre-class quizzes* [10%]

Pre-class quizzes are posted 12 to 24 hours prior to the next class. These must be completed prior to the next class; the website automatically turns off access to the quiz at 7:30am, and the class is at 09:30am. This feature cannot be deactivated or delayed. Solutions will be displayed at 07:30am. Obtaining a score of 80% or greater on these easy pre-class quizzes entitles you to having the worst assignment grade dropped.

#### *In-class activities* [15%]

- In-class activities are varied, but there will be roughly 20 activities. These activities are completed in your group and only group members present in class can put their names on the work done. The activities will generally be graded out of 3 points [0=fail, 1=33%, 2=67%, 3=100%], so you should do enough work to convince the TA or myself your group deserves a passing grade. Short answers and minimal work will be a guaranteed fail.
- We understand people fall sick, or have to miss class for job interviews or other reasons. So missing 10 to 15% of classes will not penalize your grade (i.e. about 3 activities can be missed). Should you score 80% or greater in this category you will have the worst assignment grade dropped.

This can be combined with the prior criterion: in other words, 80% or greater in both the pre-class quizzes **and** the in-class activities implies your *two* lowest assignment grades will be dropped. To put this in context, it can move your grade up by about a single grade letter, e.g. from a B+ to a A-.

- Please bring a calculator. Later in the course, a laptop in your group will be essential. You will form small groups that you will work in in every class. I will not create these groups for you; I expect you will find 2 other people to form a group of 3 (4C3 students), or one other person (6C3) to form a group of 2.

You will regularly work with these colleagues, and encourage each other to come to class. You will not be allowed to randomly change groups during the course. You must email me 2 days prior to when you want the change to take effect.

#### *Assignments* [10%]

- We strongly encourage you to complete the assignments in groups of no more than 3 members if you are in 4C3. For 6C3 students the group size is 2 people (you, and one other person). 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 100 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). Note: a "day" in this outline refers to any day of the week, not only working days.
- 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.
- All assignments will be graded, and the mean of the average of all  $N$  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.

### *Midterm and final exam [50%]*

- Electronic textbooks and resources are, unfortunately, not permitted at the midterm or final exam, but may be used during class time and for your assignments.
- The midterm and the final exam will be closed-notes; no materials are allowed. A page of formulas will be provided to you. The class will jointly decide what appears on this page 5 to 10 days prior to the test or exam, and this page will be available on the course website for you, so you are aware of what is provided.
- Only the standard McMaster calculator may be used during tests and exams.
- The final exam will be cumulative, based on the entire semester's material.

### *Missed work*

No make-ups will be given for missed assignments, class work, tests, quizzes or midterms. Assessments for which an MSAF is received may have an alternative piece of work prescribed, at the instructor's discretion, or the weight transferred to the final exam.

### *Grading policies*

- 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.
- 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.

## **4 Important notes**

### **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 TAs and the instructor will post solutions in these languages. More details are posted on the course website.

### **Out-of-class access**

The instructor has his availability [posted on his website](#). Try to send email from your McMaster account - email from personal accounts are sometimes discarded by spam filters. Your instructor filters his email, so emails from @mcmaster.ca addresses receive higher priority than personal addresses.

The TA's for this course can be contacted by email - please use their email addresses above.

## **5 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.

## 6 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:

<b>Date</b>	<b>Description</b>
6 January 2015	Overview class: review of course content and administrative issues
09 February	<b>Written midterm</b> (tentative)
16-22 February	Midterm break
27 March	<i>Kipling</i>
01 April	<b>Experimental project due</b>
08 April	Review class
10 to 30 April	Final exams