

Final Course Outline

Advanced Data Analysis

CHEM 498/610

Gregor Kos

September 12, 2018

1 Course description

CHEM 498/610 presents current methods of univariate, multivariate data analysis and experimental design. The introduction will cover the use of R and how prerequisite statistics knowledge is conducted in R. You will explore your own experimental data or data provided to obtain answers to research questions in a hands-on approach. Focus will be on the validation of models and a discussion of their limitations. The section on experimental design will introduce students to current methods using Modde software. Applications will be illustrated through discussion of the peer-reviewed literature, published methods and results.

2 Learning outcomes

After this course students will –

- understand data structures and be able to curate experimental data sets to prepare them for statistical analysis.
- know and apply basic data manipulation steps in R.
- know to apply univariate and multivariate statistical methods using Excel and R.
- be able to critically evaluate model performance and know about limitations of the models employed.
- have critically evaluated the scientific literature studies on the covered topics
- know and apply methods of experimental design.
- analyse and present results from their own or provided data as part of a comprehensive data analysis project.

3 Course administration

Please note that this course will be given for the first time. The instructors reserve the right to make changes to the syllabus and assessments as the course progresses and as needed without penalizing students.

Instructors:

Gregor Kos (GK, course coordinator), gregor.kos@concordia.ca

Cameron Skinner (CS), cameron.skinner@concordia.ca

Dajana Vuckovic (DV), dajana.vuckovic@concordia.ca

Offices: SP-275.27 (CS) / SP-275.29 (GK) / SP-275.31 (DV)

Office hours:

GK: Wed, 15:30–17:00 & Fri, 10:00–11:30 and by appointment

CS: Most afternoons, for extensive questions set an appointment by email

DV: Wed, 15:30–16:30 (Oct 24–Dec 3) and by appointment

Prerequisites: CHEM 312 or permission of the instructor

Class time: Mon & Wed, 11:45–13:00

Room: CC 320 LOY

Course website: <http://moodle.concordia.ca> (automatic enrolment)

Course text: **There will be no designated textbook for the course.**

Resources will be presented by the instructors

4 Software

You require a laptop computer with the following software packages installed (free of charge) for in-class work, assignments and projects –

- R (<https://cran.r-project.org>)
- RStudio (<https://www.rstudio.com>)
- MS Excel (<https://www.concordia.ca/it/services/office-365-education.html>).
- Modde Pro (<https://umetrics.com/product/modde-pro>)

Do not install Modde Pro until prompted by the instructor!

For platform or version-related questions, please contact the course coordinator.

5 Assessment

During the lecture two assignments will be posted on Moodle. Together with the individual term project the final mark will be calculated –

- Participation: 6%
- Assignments (GK) (2, take-home): 12%
- Assignment (DV) (1, take-home): 6%

- Assignments (CS) (3, take-home): 18%
- In-class quiz (DV): 8%
- Final project (centrally scheduled): 50%

Assignments handed in after the posted deadline will receive a 20% deduction. If they are more than a week late, you will receive a grade of zero for the assignment. **Please see the instructor before the deadline for extensions without deductions (doctor's note can be handed in after the deadline).**

All your assignments and project files are to be submitted electronically via Moodle. Please see the detailed submission guidelines on Moodle.

5.1 Assignments and quizzes

Each instructor will be posting assignments and run in-class quizzes to determine your progress with the course material.

Assignments GK

Two assignments will be posted with Assignment #1 covering the “Introduction to R” at the beginning and Assignment #6 covering “multivariate methods” towards the end of the course.

Assignments CS

Three (#2-#4) assignments will cover least square regression and non-parametric statistics.

Assignments and Quiz DV

One assignment (#5) and one in-class quiz will be given during the “Experimental Design” section of the course,.

5.2 Final project

The main objective of the course is to enable students to do data analysis on their own (or chosen) data set. The objective of the final project is, therefore, to show progress in the analysis of your own data and to apply methods learned in the course. For the final project define a data analysis objective and analyse your data to prove or disprove your hypothesis. As documentation submit –

- Your data in csv format
- Your R code with comments (must run on any machine)
- Your plots in png format
- Your final report consisting of a brief introduction with peer-reviewed references, a data analysis method section, your results and discussion with conclusions
 - Report length: 2000 words (undergraduate students); 3000 words (graduate students) ($\pm 5\%$)
 - Title page with the title of paper, your name, your student ID, the course name and year

- Printed on letter-sized paper
- Fonts & spacing: Times New Roman, 12 pt font, double spaced for body text; Times New Roman, 16 pt font for the title and section headers; no additional spacing
- Use the Harvard citation style; e.g., `guides.is.uwa.edu.au/harvard`
- Do a spell and grammar check before submission
- Only articles/books from the scientific, peer-reviewed literature are permitted; if in doubt, ask!
- You may use up to 3 references from the general interest media for motivational or introductory purposes

Detailed instructions will be provided by the instructor covering the topic your project will cover (univariate data analysis (CS), multivariate data analysis (GK), experimental design (DV)).

All final projects must be approved by the instructors before you start your work. Feel free to discuss your final project at any time during the course. Seek feedback early to avoid problems as the deadline approaches.

6 Important dates

CS will be announcing assignment (#2, #3 and #4) dates during the “Univariate Statistics” section.

Date	Event
5 Sep 2018	First class
26 Sep 2018	Assignment #1 (GK) posted
1 Oct 2018	<i>Quebec elections – no class</i>
3 Oct 2018	Deadline Assignment #1 (GK)
8 Oct 2018	<i>Thanksgiving – no class</i>
14 Oct 2018	Deadline to complete CHEM 101 quiz
5 Nov 2018	In-class quiz (DV)
7 Nov 2018	Assignment #5 (DV) posted
21 Nov 2018	Deadline Assignment 5 (DV)
21 Nov 2018	Assignment #6 (GK) posted
28 Nov 2018	Deadline Assignment #6 (GK)
3 Dec 2018	Last class (unless a make-up class is scheduled)
4 Dec 2018	Deadline to submit final project report

7 Attendance

I strongly suggest to attend classes regularly. To be excused from assessments (no supplemental), you must present a doctor's note or other suitable official excuse. Marks for a missed assessment with an official excuse will be added to the final project; without excuse a grade of zero will be given.

8 Course content

Chemistry CHEM 498/610 is an advanced data analysis course. After a brief introduction to R, the focus will be on statistical methods of data analysis (univariate and multivariate) and experimental design. The following preliminary list of lecture content provides an overview of the material covered during the course.

8.1 Lecture schedule overview

- 3 weeks: R introduction & basic (CHEM 312) statistics with R (GK)
- 3 weeks: Univariate statistics using R and Excel (CS)
- 3 weeks: Experimental design using Modde (DV)
- 3 weeks: Multivariate statistics using R (GK)

8.2 Detailed lecture schedule

Date	Content	Instructor
5 Sep	Course introduction	GK
10 & 12 Sep	What is R? Elements of R programming	GK
W1	Script writing, data and variable types	
17 & 19 Sep	Data curation, subsetting, missing data	GK
W2	Plotting, packages, getting help	
24 & 26 Sep	Statistical methods you know. Now	GK
W3	do it in R! Statistical tests & linear regression in R	
3 Oct	Univariate data analysis	CS
W4		
10 & 15 Oct	Univariate data analysis	CS
W5		
17 & 22 Oct	Univariate data analysis	CS
W6		
24 & 29 Oct	Experimental design	DV
W7		

Date	Content	Instructor
31 Oct & 5 Nov W8	Experimental design	DV
7 & 12 Nov W9	Experimental design	DV
14 & 19 Nov W10	Univariate vs. multivariate data analysis How do I know my model is good?	GK
21 & 26 Nov W11	Principal component analysis; widely used, but not the gold standard Data analysis methods 1; pros and cons	GK
28 Nov & 3 Dec W12	Data analysis methods 2; pros and cons Summary and what next? Teach yourself!	GK
4 Dec W13	Make-up class; scheduled as needed. Please watch Moodle announcements for schedule changes!	TBD

9 Mandatory Quiz and Seminar

As part of this course, you are required to

1. attend a Chemistry and Biochemistry Departmental Seminar on the academic conduct code and the appropriate use of information sources.
2. pass the online quiz associated with this seminar (the passing grade for the quiz is 100%). **The deadline to take and pass the quiz is Sunday, October 14, 2018 at 11:55 pm.**

Note: This is not the University's quiz you may have been asked to take when you first registered and logged into the myConcordia portal; the one you must take is similar, but graded by the Department of Chemistry and Biochemistry, and you cannot take it until after you have attended the seminar.

The aim of this seminar is to clarify the academic conduct code in terms of what practices will be considered unacceptable with regards to work submitted for grading in Chemistry and Biochemistry courses. You are only exempt from repeating the seminar and the quiz if you have done both in Fall 2013 or more recently [*], otherwise you are required to repeat both this term.

This short seminar (1 hour) will be held at the following times (note that **late-comers will not be admitted**):

Date	Time	Place
Mon, Sep 24	16:45-17:45	HC-155
Tue, Sep 25	16:45-17:45	CC-116
Tue, Sep 25	20:45-21:45	HB-130
Wed, Sep 26	16:45-17:45	HC-155
Wed, Sep 26	20:45-21:45	SPS-110
Thu, Sep 27	16:45-17:45	CC-115
Fri, Sep 28	16:45-17:45	HB-130

As space for each of the seminars is limited by the room size, please sign up to your preferred time as soon as possible (slots fill up quickly). Sign-up sheets are available two weeks in advance of the seminars outside SP 201.01 (Departmental office). Only sign up in available slots: rooms must not be filled over capacity!

If you do not complete this course requirement, your final grade for the course may be lowered by one full letter grade with an incomplete (INC) notation until such time as this requirement is completed. Please refer to the undergraduate calendar (section 16.3.6) for details on removal of an incomplete notation.

[*] You are exempt, if you can locate your ID in the pdf file on the

departmental website at <http://www.concordia.ca/content/dam/artsci/chemistry/docs/Compliance-list.pdf>.

10 Plagiarism and other forms of academic dishonesty

The Academic Code of Conduct can be found in section 17.10 of the academic calendar (<http://www.concordia.ca/academics/undergraduate/calendar/current/17-10.html>). Any form of unauthorized collaboration, cheating, copying or plagiarism found in this course will be reported and the appropriate sanctions applied. The mandatory seminar is a clear and fair opportunity to learn what our faculty regards as academic misconduct. Failure to take part in this learning opportunity and thus ignorance of these regulations is no excuse and will not result in a reduced sanction in any case where academic misconduct is observed.