McMaster University

Department of Mathematics and Statistics

STATISTICS 3A03: Applied Regression Analysis with SAS

Fall 2017

Instructor	Angelo J. Canty
Office	Hamilton Hall 209
Phone	(905) 525-9140 Ext 27079
e-mail	cantya@mcmaster.ca
Course Web Site	http://www.math.mcmaster.ca/canty/teaching/stat3a03
Time of Lectures	Monday 9:30 – 10:20 Tuesday 10:30 – 11:20 Thursday 9:30 – 10:20
Lecture Location	BSB B135
Office Hours	Monday 10:30–11:30 Tuesday 11:30-12:30 Thursday 10:30–11:30 or at other times by prior appointment.
Labs	L1 Friday 11:30-12:20 in BSB 249 L2 Tuesday 11:30-12:20 in BSB 244 L3 Thursday 8:30–9:20 in BSB 249
Required Text	*Regression Analysis By Example* (Fifth Edition) by Samprit Chatterjee & Ali S. Hadi. Wiley (2012)
Optional Text (Recommended)	*Applied Linear Regression* (Fourth Edition) by Sanford Weisberg. Wiley (2014)

Grading Scheme :

The final grade will be based on the following components

1. 20% for the best 4 out of 5 assignments.
2. 20% for each of two term tests.
3. 40% for the final exam.

Notes :

- Assignments must be submitted in the appropriate mailbox outside by 5pm on the due dates. Late assignments will not be accepted. Tentative due dates for the assignments are Sept. 28, Oct. 19, Nov. 2, Nov. 16 and Nov. 30. Assignments will be posted on the web site at least 10 days before the due date.

- The term tests will be held during class time. Tentative dates are October 24 and November 28. The tests are currently scheduled to take place at the Canadian Martyrs Testing Centre. Details of the location and entrance for this testing centre are at http://www.cas.mcmaster.ca/ leduc/canMartyrTestFaciInfo.pdf. Further information will be announced in class and on the web site at least one week before the tests.

- The final exam will take place during the December examination period and will cover the entire course. Students should check the official exam schedule for the time and location.

- No materials other than a McMaster standard calculator and writing implements may be brought into any test or exam. A sheet of useful formulae will be provided in the exam and will be posted on the website in advance of the test/exam.

- Any student who misses work during the term should complete the online McMaster Student Absence Form (MSAF) within one week of the missed work. Second or subsequent absences must be approved by the associate dean of your faculty. For more details see the website https://www.mcmaster.ca/msaf/.
  Students using the MSAF should also email the professor after submitting their request to ensure it was properly received. In the case of a missed term test, the marks for that test will be distributed over the other test (+5%), assignments (+5%) and the final exam (+10%). In the case of one or more missed assignments, the assignment mark will be calculated using **all** other assignments not covered by an MSAF or otherwise approved. **No other modifications of the marking scheme will be considered.**

- Lecture notes and notes for the SAS labs will be posted on the website in advance of the lectures/labs. All datasets used for assignments and labs will also be posted on the website as plain text files for download by the students as they are required.

- Consult the website frequently as important notices such as corrections to assignments or notes as well as information regarding tests etc. will be posted there first.

- SAS labs are a mandatory component of the course and your understanding of both SAS code and output are examinable materials. Please ensure you go to the lab to which you have been assigned as the computer labs have limited capacity. Labs will start in the second week of term (September 11-15).

Academic Dishonesty

Academic dishonesty consists of misrepresentation by deception or by other fraudulent means and 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 kinds of academic dishonesty please refer to the Academic Integrity Policy, specifically Appendix 3, located at
http://www.mcmaster.ca/policy/Students-AcademicStudies/AcademicIntegrity.pdf

The following illustrates only three forms of academic dishonesty:

- **Plagiarism, e.g. the submission of work that is not one's own or for which other credit has been obtained.**
- **Improper collaboration in group work.**
- **Copying or using unauthorized aids in tests and examinations**

Brief Outline of the Course

This course will introduce the students to the theory of regression and its implementation using the SAS language. Note that the required textbook is rather applied in nature and does not fully cover the required theory. The optional textbook will be primarily used for the more theoretical parts of the course. The course consists of 3 lectures per week and also one computer lab. The lab is mandatory as this is where the students will learn and be able to use the SAS system. Assignments will require the use of SAS.

The course will commence with a review of some of the important distributions that will come up in this course; primarily the Gaussian (normal), Student's t and chi-squared distributions. We will then examine simple linear regression in which one response is related to a single predictor variable. Most students will have seen this in Stat 2MB3 but in this course we shall delve further into the theoretical basis for simple linear regression. We shall then extend to multiple regression in which there are more than one continuous predictor variables. This part of the course will also involve familiarity with matrix notation and operations.

We will then consider more advanced aspects of regression, the first of these will be diagnostics to assess problems with the underlying assumptions and fitted model. We

will follow this by considering the issue of qualitative predictor variables and the concept of interactions. Transformations and weighted least squares are two methods that can overcome some issues in regression and these will be considered next. Finally we shall look at methods to decide on which predictor variables should be included in a final model.

At the end of the course, students should have a reasonable amount of familiarity with the SAS system and be able to use it for data entry and model fitting and checking. They should also understand the theory behind linear models and be able to use them in settings such as regression, analysis of variance and analysis of covariance.

Approximate Timetable

Note that this timetable is subject to change depending on how the class progresses and is intended as a guide only. Major deviations (such as topics being omitted) will be announced in class and on the web site. The textbook sections indicated are the sections of the required text most closely related to what we shall cover but actual coverage may not correspond exactly to the textbook.

Dates	**Topic**	**Textbook**
Sept 5 – 7	Review and Introduction to Regression	Chapter 1.
Sept 11 – 21	Simple Linear Regression	Chapter 2.
Sept 25 – Oct 6	Multiple Linear Regression	Chapter 3.
Oct 16 – 26	Regression Diagnostics	Chapter 4.
Oct 24	**Term Test 1**	—
Oct 30 – Nov 9	Qualitative Variables in Regression	Chapter 5.
Nov 13 – 16	Transformations	Chapter 6.
Nov 20 – 23	Weighted Least Squares.	Chapter 7.
Nov 28	**Term Test 2**.	—
Nov 27 – Dec 6	Variable Selection	Chapter 11.

Important Notice:

The instructor and university reserve the right to modify elements of the course during the term. The university may change the dates and deadlines for any or all courses in extreme circumstances. If either type of modification becomes necessary, reasonable notice and communication with the students will be given with explanation and the opportunity to comment on changes. It is the responsibility of the student to check their **McMaster** email and the course website regularly during the term and to note any changes.