

FHS-HRM753: Regression Analysis

02 May – 15 June 2022

Content/Rationale

Coordinator: Dr. Shrikant I. Bangdiwala
Office: C2-210, DBCVSRI Bldg, 20 Copeland Ave. (behind Hamilton General Hospital)
Desk phone: 905-297-3479, ext. 40340
E-mail: bangdiws@mcmaster.ca (preferred method of contact)

Instructors: Dr. Shrikant I. Bangdiwala
Dr. Wei Q. Deng - dengwq@mcmaster.ca - May 11, 16, 18
Dr. Joycelyne Ewusie - jewusie@stjosham.on.ca - May 25

Teaching Assistant: Ricky Lali
E-mail: lalir@mcmaster.ca

Course Objectives:

This is a second level course in statistics concentrating on several modeling techniques. There will be four main regression topic areas:

- 1) simple and multiple linear regression of continuous outcomes
- 2) logistic and Poisson regression for binary and count outcomes
- 3) proportional hazards regression for time to event outcomes
- 4) regression (multilevel) models for correlated outcomes

We will focus on general model building approaches applicable to all regression models, as well as on the specific assumptions and goodness of fit methods for each type of regression model.

Students will be required to carry out analyses on various datasets that will be provided to the students, using one or more of these techniques; this work will form a major basis for course evaluation. Students will also be expected to analyse a dataset of their own and prepare a 5-page report. Examples of uses of regression models in the literature will also be discussed in class.

Prerequisites: HRM 702, or equivalent. Students are expected to be comfortable with basic statistical principles and methods: sampling, probability distributions, descriptive analyses, statistical inference (confidence intervals, tests of hypotheses). A survey on background will be conducted.

Anti-requisite: HRM 723

Time and Place:

The standard time for the course is on Monday and Wednesday from 1:00pm to 4:00pm, held in-person in HSC 1J10 on the hospital first floor (purple section). The lectures will be held from 1:00-2:30pm. A question and answer period will follow (2:30 to 2:45pm). Break from 2:45-3:00pm. Tutorial sessions will take place from 3:00-4:00pm.

Format:

Tutorial sessions will involve discussions on how each person approached the assigned problem, both conceptually and with the computer software, and the interpretation of the results. Tutorial will be based on content from the preceding class. Each tutorial will have a small group lead and take-up the tutorial questions. Small groups will be assigned at the beginning of the course.

This is a hybrid of the problem-based learning (PBL) approach in which solving problems in a small group setting is used as a vehicle for learning. The lectures will provide the tools necessary to solve the problems and will be structured to allow time to address “matters arising” from previous weeks.

Resources:

Here are some ideas about resources that you could consult.

1) The lectures, datasets, exercises and literature examples. These are not intended as a blueprint to indicate exactly what is and is not included in the course; instead, they are one of several resources you can use to try to understand the concepts of the course. I approximately follow the topics as given in the lectures, but there will be flexibility to respond to the needs and interests of the group.

2) Textbooks.

Purchase of a textbook is **optional**. Many textbooks cover similar ground.

Recommended textbooks for:

- **linear and logistic regression and ANOVA** topics: "*Applied Regression Analysis and other Multivariable Methods*," Kleinbaum, Kupper, Nizam & Rosman, 5th ed., 2014 (KKNR).
- **logistic regression**: "*Applied Logistic Regression*," Hosmer & Lemeshow, 3rd ed., 2013 (HL)
- **survival analysis**: "*Applied Survival Analysis: Regression Modeling of Time-to-Event Data*," Hosmer, Lemeshow, & May, 2nd ed., 2008 (HLM)
- **multilevel analysis**: "Data Analysis Using Regression and Multilevel/Hierarchical Models", Gelman & Hill, Cambridge University Press, 1st ed., 2007 (GH)

Two great comprehensive textbooks are

- "*Regression Methods in Biostatistics: Linear, Logistic, Survival and Repeated Measures Models*," 2nd ed., by Vittinghoff, Glidden, Shiboski and McCulloch, Springer, 2012
- "*Regression Modeling Strategies, with Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*", by Harrell, F; Springer 2015

3) Software manuals and websites for programs such as SPSS, STATA, R or SAS also may have some background on many of the topics covered in this course.

4) Great site from University of California: <http://www.ats.ucla.edu/stat/>

None of these resources is sufficient by itself. Like the real world, you are expected to approach problems using multiple methods.

Computing:

Because of the diversity of computing environments available to students, and their previous

experience with various software programs, the current position of the HRM program is that we do not mandate the use of any specific package for this course. The exercises will be generally illustrated using STATA or SPSS and maybe R, but suitable modification will usually allow one of the other major packages (SAS, Minitab, ...) to do the same things.

If you anticipate having unusual analytic requirements that are not handled by routine packages, you may want to consider using a program such as S-Plus, or R; these require more programming skills, but do provide greater customisation and flexibility in the form of output you can generate.

If you are already expert with a suitable statistics program, don't switch. It is **your** responsibility to make sure you have access to adequate statistical software as soon as possible after the start of the course. It is desirable to have a statistical package installed on your computer.

Student Evaluation:

Evaluation will be based on

- 30% - midterm assignment/examination – 5 pages maximum
- 40% - final assignment/examination – 8 pages maximum
- 10% - completion of tutorials submissions
- 15% - leading tutorial
- 5% - attendance of lectures and tutorials

The **mid-term exam** will cover regression models presented in class and will involve answering research questions from a dataset provided (5-page maximum + Appendix of computer program & output). The **final exam** will cover the entire course and will be similar in style to the mid-term (8-pages maximum + Appendix). **All exams** are to be completed by the student independently, without any discussion between students and without any contact involving individuals that would be considered a biostatistical resource.

Schedule: Here is the approximate schedule of topics

Session	Date	Instructor	Topic
1	May 02	SIB	Introduction – studying relationships
2	04	SIB	Simple linear regression
3	09	SIB	Multiple linear regression I
4	11	WQD	Multiple linear regression II
5	16	WQD	Logistic regression I
6	18	WQD	Logistic regression II
	23		Victoria Day – no class
7	25	JE	Poisson and count data regression
8	30	SIB	Additional topics: explanatory vs predictive models; handling missing data; causal models; model building with correlated independent variables; how to report regression analyses in applied publications
9	June 01	SIB	Survival analysis
10	06	SIB	Cox proportional hazards regression

11	08	SIB	Multilevel Analysis I - clustering
12	13	SIB	Multilevel Analysis II – repeated measures
13	15	SIB	Summary lecture – tying it all together - “Modelling as an Art”

Expectations for Tutorials

All students: Should submit their answers prior to class via Avenue2Learn for all questions.

Leaders: Tutorial Assignments will consist of at least 3 questions. Each leader for that week is expected to present on at least 1 question.

Tutorial Questions: Aim to answer tutorial questions as if presenting to a non-statistician. Provide output, answer the research question (clinical and statistical interpretation of the data) and discuss limitations (e.g. model fit, residuals, sampling methods).