BIOLOGY 4605/7220 Statistical Analysis in Biology and Environmental Science

FALL 2023
Version: 5 Sept

Lectures:	Mon Wed Fri 12 PM	ED 2003
Labs:	Tue 3-5 or 6-8 PM	C 2003
Instructors:	Amy Hurford	Office (CSF 4338) Tel 864-8301
	ahurford@mun.ca	Hours – M 2pm; W 2pm
	David Schneider	Office (CSF 3228) Tel 864-8393
	a84dcs@mun.ca	Hours - after class or by appointment (CSF4220)
Teaching Assistant:	Ranjan Waggle	rwagle@mun.ca

Course Summary. The goal of this course is for you to learn a model based approach to the statistical analysis of research data. Skill and confidence come with practice, so assignments and quizzes will be short and frequent. Lecture material will emphasize principles of good quantitative analysis, illustrated by complete examples. Laboratories will cover the computational aspects of problem solving, with a package of the student's choice.

Goals

1. Principles of good analysis.
2. Skill in application.
3. Capacity for self-instruction.
4. Confer with statistician.
5. Develop critical capacity.
6. Evaluate quantitative presentations.

Pre-Requisite: 1 course in statistics.

Evaluation	N	4605	7220	wt
Labs	9	25	25	2.8
Assignments	5	10	10	2.0
Quiz/SA	10	25	25	2.5
Exam 1	1	10	10	10
Exam 2	1	10	10	10
Final	1	20	20	20
Written Report			30	
	27	100	130	

Exams and quizzes are open book, emphasizing use of tools, rather than to memorizing formulas.

Graduate students (Biology 7220) will be required to prepare a written report on the analysis of a set of data of interest to the student. The topic will be decided during a conference early during the term, then discussed during tutorial sessions. The report will constitute 30% of the final mark.

If you are unable to complete evaluated work due to acceptable cause submit a written request stating your name, the date and name of work, and reason for non-completion.

All course material is at https://davidcschneider.github.io/StatisticalScience/

Required material: Lecture Notes in Statistical Science
Laboratories in Statistical Science

Additional material: Review Questions in Statistical Science

A calculator is required for quizzes, exams, and Lab 2. The calculator (which can be an app on a portable device) does not require statistical functions but does require y^x and e^x functions.

Labs and assignments are due in pdf format on the date stated in the syllabus. Work will be returned to students within a week (usually the next lecture after it is due}.

Late work will be penalized at 5% off per day (excluding weekends).

Lab 1 is a group project that requires attendance for successful completion. Labs 2 and 9a are group projects for which attendance is recommended.

About quizzes/short assignments. These cover lecture material since the previous quiz. Midterm and final exams will have the same format as quizzes. Examples of quizzes are posted on the course website. In class quizzes are easier to do on paper than electronically; please submit them electronically to BrightSpace. Short assessments (SA) are quizzes due by the end of the day.

About labs and assignments. Working together is encouraged in all labs and assignments. However, each person is responsible for preparing their *own* written report (don't share write-ups). Obvious duplicates will be considered misconduct (see below). Please submit all work on BrightSpace.

About statistical packages. Labs 3, 5, and 6 can be completed in a statistical package or in a spreadsheet using functions and data analysis tools. Labs 3-9 can be completed in any statistical package with a general linear model (GLM) routine. Lab 10 can be completed in any package with a logistic regression routine, or with a generalized linear model (GzLM) routine.

Packages that lack a GLM routine and a logistic regression routine are superficially attractive but lack any value in learning principles and best practice in statistical analysis.

Statistical packages consist of line code (you type the command) and a GUI– a graphics user interface (mouse clicks to set up the analytic model).

This course uses RStudio but it does not require any particular package. One of the instructors (DCS) can help you with any of the following.

SPSS. Easily learned GUI, line code not suitable for archiving, randomization not possible.
Minitab. Easily learned GUI. Line code adequate for GLM and simple calculations.
JMP. GUI for SAS code.
SAS. Relatively easy line code. Gold standard for statistical analysis. Expensive license.
RStudio. Open source freeware for editing and executing R code. https://www.rstudio.com/
R. Freeware for statistical computing and graphics. Line code more opaque than SAS or Minitab. R has overtaken SAS in academic settings because it has no licensing fee. The hidden cost is the learning curve. The course website has Rcode for all the labs.

Printing computer output. Please do not print the sometimes verbose files produced by statistical packages. Instead, copy and paste the appropriate sections of output into your lab report or assignment. Note that you will have to use a non-scalable font (`such as courier`) to print or display numerical output (ANOVA tables, *etc.*) without distortion.

Disruptions. *In the case of a class disruption or cancellation, and in the case of revisions to evaluation methods, the instructor or the Head of the department will notify all students registered in the course via the course shell in Brightspace. Any necessary revisions to the evaluation methods will be made in consultation with the students registered in this course. If a student demonstrates that they would be disadvantaged by the change, then, as per 6.7.4 of the University Calendar, accommodations will be made.*

TABS Teaching assessment by students.

Course evaluation questionnaires are of use to instructors in several aspects of teaching, especially delivery. Questionnaires at the end of a course are of little use to students, as it is too late for corrective action. No-name questionnaires and polls at the end of a lecture are short, a few minutes. Summaries of the results are usually delivered orally in class.

Use of Artificial Intelligence. Original work, completed wholly by you, is expected to be submitted in this course. The use of an artificial intelligence tool like ChatGPT prevents active learning. It is known to produce unsubstantiated statements with a high degree of plausibility. If you use artificial intelligence in graded work, state how you used it and how you checked its veracity. With this course on your transcript, you will be expected to be able complete statistical analysis. Your future professional competency will be compromised if you rely on artificial intelligence and do not master the course material.

Academic Conduct *Students are expected to adhere to those principles which constitute proper academic conduct. A student has the responsibility to know which actions, as described under Academic Offences in the University Regulations, could be construed as dishonest or improper. Students found guilty of an academic offence may be subject to a number of penalties*

commensurate with the offence including reprimand, reduction of grade, probation, suspension or expulsion from the University. For more information regarding this policy, students should refer to the University Regulations for Academic Misconduct (Section 6.12) in the University Calendar.

Inclusive Education. *Memorial University of Newfoundland is committed to supporting inclusive education based on the principles of equity, accessibility and collaboration. Accommodations are provided within the scope of the University Policies for the Accommodations for Students with Disabilities (www.mun.ca/policy/site/policy.php?id=239). Students who may need an academic accommodation are asked to initiate the request with the Glenn Roy Blundon Centre at the earliest opportunity (www.mun.ca/blundon).*

Tentative schedule

Day	Date	On Web	Topic	Labs	Due	Marking
Wed	6-Sep	Ch1	Intro to Course			
Fri	8-Sep	Ch2.1, 2.2	Quantities			
Mon	11-Sep	Ch2.5,2.6	Units, Dimensions		**A1 Quantities**	RW
Tues	12-Sep	Lab1	Inferential Cards	CSF2313		
Wed	13-Sep	Ch3	Rescaling		**Quiz 1**	AH
Fri	15-Sep	Ch4	Equations		**Lab1**	RW
Mon	18-Sep	Ch5	Data Eq		**A2**	RW
Tues	19-Sep	Lab2	RStudio Intro	CSF2313		
Wed	20-Sep	Ch6.1	Freq Dist I		**Quiz2**	AH
Fri	22-Sep	Ch6.2, 6.3	Freq Dist II		**Lab2**	RW
Mon	25-Sep	Ch7, 7.1,7.2	Inference, Randomization tests			
Tues	26-Sep	Lab3	Freq Dist [excel/StatPackage]	C2003		
Wed	27-Sep	Ch7.3, 7.5	Hypoth. Tests & Conf. Limits		**Quiz3**	AH
Fri	29-Sep	Ch8, 9.1	GLM Intro, Regression I		**Lab3/A3**	RW
Mon	2-Oct	No lecture - National Day				
Tues	3-Oct	Lab4	Randomization tests	C2003		
Wed (Mon)	4-Oct		**Unit I Exam**			AH
Fri	6-Oct	Ch9.2	Regression II		**Lab4/A4**	RW
Mon	9-Oct	Holiday				
Tues	10-Oct	Holiday				
Wed	11-Oct	Ch9.3, Ch10.2	Regression, t-test			
Thurs (Tues)	12-Oct	Lab 5a	Regression			
Fri	13-Oct	Ch10.3, 10.4	1-way ANOVA		**Quiz4**	DCS
Mon	16-Oct	Ch11	Rev: 1 Expl Var		**Lab5a**	
Tues	17-Oct	Lab6a	1-way ANOVA	C2003		
Wed	18-Oct	Ch12.1	Mult Regr		**Quiz5**	DCS
Fri	20-Oct	Ch13.1 (13.2)	2-way ANOVA		**Lab6a only**	RW
Mon	23-Oct	Ch13.3	Mixed Model - Paired t-test			
Tue	24-Oct	Lab7	Multifactor ANOVA	C2003		

Wed	25-Oct	Ch13.4	Mixed Model - Rand. Block		**Quiz6**	Peer/DCS
Fri	27-Oct	Ch13.6	Nested random effects ANOVA		**Lab 7**	RW
Mon	30-Oct	Ch14.1	ANCOVA			
Tue	31-Oct	Lab8	ANCOVA	C2003		
Wed	1-Nov	Ch15 review	Ch 15 Review			
Fri	3-Nov		**Unit II Exam**			DCS/AH/RW
Mon	6-Nov	Ch20.1, 20.4	Correlation & Multivariate Analysis		**Lab8**	RW
Tue	7-Nov	Lab9a	GLMM Problem setup	C2003		
Wed	8-Nov	Ch16, 16.2,3,4	Analysis of Deviance		**Quiz7**	AH
Fri	10-Nov	Ch18, 18.1	Logistic regression			
Mon	13-Nov		No class			
Tues	14-Nov	Lab9b	GLMM Problem execution	CSF2313		
Wed	15-Nov	Ch 18.2	Prospective analysis		**Quiz 8**	AH
Fri(Mon)	17-Nov	Ch 18.3	Retrospective analysis			
Mon	20-Nov	Ch 18.6	Logistic ANCOVA		**Lab 9a,b**	RW
Tue	21-Nov		Review - GLMM			
Wed	22-Nov	Ch 17.1	Poisson regression		**Quiz 9**	AH
Fri	24-Nov	Ch 17.4, 17.5	Contingency tests, Poiss ANCOVA			
Mon	27-Nov	Cn19.1-19.4	Model selection I - EDA and stepwise			
Tue	28-Nov		No Lab			
Wed	29-Nov	Ch19.5	MultiModel Inference (AIC, SIC)		**Worksheet**	AH
Fri	1-Dec		TBA			
Mon	4-Dec	Worksheet	Course review with Q&A		**Worksheets online**	AH
	7 Dec Exams begin		**Final Exam 2 hour synchronous**		**TBD**	DCS/AH/RW

Assignments. See schedule for due dates.

A1. Quantities

In the library or on line, find a journal reporting research results. Open the journal to an article, and list the first defined physical or biological quantity you encounter (if you must move to the next article, then so be it). State the journal name, volume, and page number. For this quantity, provide complete details for each of the 5 components of the quantity: name, symbol, typical value, units, and procedural statement. If a component is not present then state 'not present.'

Find 3 more quantities; complete the following 8 point check list for all 4 quantities.

Journal name, volume, and page number _____

name of quantity in words. Present? _____ If so, name is _____

symbol Present in article ? _____

number of values N = _____or cannot be determined _____

procedural statement Present ? _____

Reproducible by another investigator ? _____

type of measurement scale (nominal, ordinal, interval, ratio) _____

If ratio scale: units = _____

(4 due in all)

A2-A5 These require graphs and summary statistics from a text book or published literature. To find examples on a topic of interest to you try Google Scholar. Publications with suitable examples are listed on the course website.

A2. Data Equations.. In the published literature find a graph where a regression equation has been displayed. A list of such publications can be found on the course website: https://github.com/DavidCSchneider/StatisticalScience/tree/main/Data

State the source publication (request bonus point for an example not on the course website). Write the equation, write the name of each symbol or parameter value, and give its units. Immediately below the equation (symbolic form) display a data equation for each of 3 different values of the explanatory (X) variable.

A3. Hypothesis testing. Find, in the published literature, two mean values with associated standard deviations and sample size.

1. Report the 6 values with full citation of source of the published data.
2. Compute the t-statistic using the appropriate formula from Ch7.3. State which formula you used and why. Use the 10 step generic recipe for decision making with statistics (Ch7.3, Table 7.1) to declare a decision about the two means.

A4. Confidence intervals. For the same data used in A3, compute the confidence limits for each mean. Report all 6 values (means, sd, n), the source of the numbers, and both confidence limits. Use the generic recipe for confidence limits (Ch7.5, Table 7.5a).

To obtain critical *t*-values for confidence intervals, use commands you learned in Lab 3.

A5. Correlation. Find, in the published literature, a table of data that you consider appropriate for correlation. Enter the data into a spreadsheet or statistical package. Compute the mean and variance for each variable. Compute the correlation coefficient. State the source of the data (with full citation), why correlation is appropriate, then display the data (label each column), each mean and variance, and the correlation coefficient. Show calculation of the likelihood ratio from the correlation coefficient. State whether inference to a population is possible and defend your argument for or against calculating a p-value to make the inference.