BIOLOGY 4605/7220 Statistical Analysis in Biology and Environmental Science

FALL 2025
Version: 12 Sept

Lectures:	Mon Wed Fri 12 PM	ED 3048
Labs:	Tue 3-5 or 6-8 PM	C 2003
Instructor:	David Schneider	Office (~~CSF 3228~~) Tel ~~864-8393~~
	a84dcs@mun.ca	Hours - after class or by appointment (online)
Teaching Assistants:	Hallie Arno	HEArno@mun.ca
	Matthew Rideout	f99MWR
	Alex Arkilanian	AArkilanian@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	

Lectures in 2025 will be online delivery in a classroom because the instructor is restricted to limited use of feet by avulsion fractures (look it up) to tibia (left) and talus (right).

Exams and quizzes are open book, emphasizing use of tools, rather than 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. The course instructor 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.

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 AI tool like ChatGPT diplaces 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. AI is incapable of judgement and flunks the quizzes in this course, With this course on your transcript, you will be expected to be able to do better: use judgment and critical thinking to write and execute a generalized linear model.

Academic Conduct *Students are expected to adhere to principles of 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 refer to the University Regulations for Academic Misconduct (Section 6.12) in the University Calendar.*

Inclusive Education. *Memorial University is committed to inclusive education based--equity, accessibility and collaboration. Accommodations are provided within the scope of the University*

Policies (www.mun.ca/policy/site/policy.php?id=239). To initiate a request for academic accommodation contact the instructor or the Glenn Roy Blundon Centre (www.mun.ca/blundon).

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.'

For each of **3 more** quantities in the journal, complete the following checklist:

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)

Assignments A2-A5. Text books in statistics are good sources of data from the published literature. To find published data on a topic of interest to you try Google Scholar. Data often appear in print before 1960, less often in subsequent decades. Publications with data 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:
http://www.mun.ca/biology/schneider/b4605/Data/RefswithRegressionEquations.pdf

State the source publication (bonus point if you find 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 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

Day	Date	On Web	Topic (TA)	Lab	Due	Marked
				Location	(A = Assignment)	by
Wed	10-Sep	Ch1	Intro to Course			
Fri	12-Sep	Ch2.1, 2.2	Quantities HA		Worksheet for credit	DCS
Mon	15-Sep	Ch2.5,Ch4	Units, Equations MR		A1	AA
Tues	16-Sep	Lab2	Equations MR	CSF2325		
Wed	17-Sep	Ch2.5,Ch3	Dimensions Rescaling AA			
Fri	19-Sep	Ch5	Data Equations AA		Lab2, Quiz2	L:MR,Q:HA
Mon	22-Sep	Ch6,1	Freq Dist I AA			
Tues	23-Sep	Lab3	Probability distributions HA	CP2003		
Wed	24-Sep	Ch6.2 6.3	Freq Dist II HA		A2 Data Equations	A:MR
Fri	26-Sep	Ch7, 7.1,7.6	Statistical Inference MR		Lab3 , Quiz3	L:HA,Q:MR
Mon	29-Sep	Ch7.2 7.3	Hypothesis testing HA			
Tues	30-Sep			No Lab		
Wed	01-Oct	Ch 7.5	Confidence Limits MR		A3 (Ch 7.3), Quiz4	A:HA,
Fri	03-Oct	Ch8.1	Unit summary AA			Q:AA
Mon	06-Oct	Unit 1 Exam	MR		A4 Conf.Limits	Ex:All A:DCS
Tue	07-Oct	Lab 4	randomization p-values AA	CP2003		
Wed	08-Oct	Ch9.1 9.2	Regression HA			
Fri	10-Oct	Ch9.3, 9.5	Regression AA		Lab4	L:AA
Mon	13-Oct		Lecture Break			
Tue	14-Oct		Lecture Break			
Wed	15-Oct	Ch10.3, 10.4	1-way ANOVA HA			
Thurs	16-Oct	Lab5a	Regression MR			
Fri	17-Oct	Ch11	Rev: 1 Expl Var AA		Lab5a, Quiz5	L:MR,Q:AA
Mon	20-Oct	Ch12.1	Multiple Regression AA			
Tue	21-Oct	Lab6a	ANOVA HA	CP2003		
Wed	22-Oct	Ch13.1	2-way ANOVA HA			
Fri	24-Oct	Ch13.2	ANOVA w/interaction MR		Lab6a Quiz6	L:HA Q:MR
Mon	27-Oct	Ch13.3	Mixed Model - Paired t MR			
Tue	28-Oct	Lab7	2-factor ANOVA AA	CP2003		
Wed	29-Oct	Ch13.4	Mixed Model, Rand.Block AA			
Fri	31-Oct	Ch13.6	Nested ANOVA HA		Lab7, Quiz7	L:AA,Q:HA
Mon	03-Nov	Ch14.1	ANCOVA MR			
Tue	04-Nov	Lab8	ANCOVA MR	CP2003		
Wed	05-Nov	Ch15 review	Last day to drop course HA			
Fri	07-Nov	Unit II Exam	AA		Lab8	L:MR Ex:All
Mon	10-Nov	Ch20.1, 20.4	Correlation & Mvar HA			
Tue	11-Nov		No Lecture			
Wed	12-Nov	Lab9a	GLMM Problem setup HA A	CSF2325		
Fri	14-Nov	Ch16,16.2,16.4	GzLM Intro MR		Quiz 8	Q:MR
Mon	17-Nov	Ch18, 18.1	Logistic regression MR			
Tues	18-Nov	Lab9b	GLMM execution HA AA	CP2003		
Wed	19-Nov	Ch 18.2	Prospective analysis AA			
Fri	21-Nov	Ch 18.3	Retrospective analysis HA		Quiz 9	Q:HA
Mon	24-Nov	Ch 18.6	Logistic ANCOVA MR		Lab 9a,b	L:HA AA
Tue	25-Nov	Lab10 optional	Logistic regression.		Online (DCS)	
Wed	26-Nov	Ch 17.1	Poisson regression HA			
Fri	28-Nov	Ch 17.4,17.5	Contingency tests, Poisson AA		Quiz 10	Q:AA
Mon	01-Dec	Cn19.1-19.4	Model selection I - EDA and stepwise MR			
Tue	02-Dec		Review - GLMM	online		
Wed	03-Dec	Ch19.5	Multimodel Inf - AIC, SIC AA		Quiz11	Q:DCS
Fri	05-Dec	Worksheet	Course review with Q&A		Online(DCS)	
	TBA	Final Exam	2 hr Synchronous	TBA	TBA	All