

# BIOLOGY 4605/7220 Statistical Analysis in Biology and Environmental Science

FALL 2024

Version: 6 Sept

<b>Lectures:</b>	Mon Wed Fri 12 PM	ED 3048
<b>Labs:</b>	Tue 3-5 or 6-8 PM	C 2003
<b>Instructor:</b>	David Schneider a84dcs@mun.ca	Office (CSF 3228) Tel 864-8393 Hours - after class or by appointment (CSF4220)
Teaching Assistant:	Lindsey Butt Emmerson Wilson	LMButt@mun.ca ERWilson@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.

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	

**Pre-Requisite:** 1 course in statistics.

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. 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 prevents active learning. It is known to produce unsubstantiated statements with a high degree of plausibility. It fails the open book quizzes. 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](http://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](http://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

### Tentative schedule

Day	Date	On Web	Topic	Location of Lab	Due (A = Assignment)	Marked by
Wed	4-Sep	Ch1	Intro to Course			
Fri	6-Sep	Ch2.1, 2.2	Quantities		Worksheet For Cred	DCS
Mon	9-Sep	Ch2.5,2.6	Units, Dimensions		A1	
Tues	10-Sep	Lab1	Inferential Cards	CSF2325	DCS	
Wed	11-Sep	Ch3	Rescaling			
Fri	13-Sep	Ch4	Equations		Lab1, Quiz2	L:LB,Q:LB
Mon	16-Sep	Ch5	Data Eq			
Tues	17-Sep	Lab2	Equations	CSF2325	DCS	
Wed	18-Sep	Ch6.1	Freq Dist I		A2 Data Equations	A:LB
Fri	20-Sep	Ch6.2, 6.3	Freq Dist II		Lab2 , Quiz3	L:EW,Q:EW
Mon	23-Sep	Ch7, 7.1,7.3	Hypothesis testing			
Tues	24-Sep	Lab3	Freq Dist	CP2003	LB, EW	
Wed	25-Sep	Ch7.2 7.5	Confidence Limits		A3 (Ch 7.3), Quiz4	A:EW,Q:LB
Fri	27-Sep	Ch8.1, 9.1	Randomization tests		Lab3	L:EW
Mon	30-Sep	No lecture	Diagram Ch 7.2		A4 Conf Limits	A:LB
Tue	1-Oct	Lab 4	Computing p-values	CP2003	LB, EW	
Wed	2-Oct	Unit 1 Exam	Monday schedule			All
Fri	4-Oct	Ch9.2	Regression		Lab4	L:LB
Mon	7-Oct	Ch9.3, 9.5	Regression			
Tue	8-Oct	Lab5a	Regression	CP2003		
Wed	9-Oct	Ch10.3, 10.4	1-way ANOVA			
Fri	11-Oct	Ch11	Rev: 1 Expl Var		Lab5a, Quiz5	L:EW,Q:LB
Mon	14-Oct	Holiday				
Tue	15-Oct	Holiday				
Wed	16-Oct	Ch12.1	Multiple Regression			
Thurs	17-Oct	Lab6a	1-way ANOVA	CP2003		
Fri	18-Oct	Ch13.1 (13.2)	2-way crossed ANOVA		Quiz6	Q:EW
Mon	21-Oct	Ch13.3	Mixed Model - Paired t-test		Lab6a only	L:LB
Tue	22-Oct	Lab7	2-factor ANOVA	CP2003	Lab7	
Wed	23-Oct	Ch13.4	Mixed Model - Rand. Block			
Fri	25-Oct	Ch13.6	Nested effects ANOVA		Lab7, Quiz7	L:EW,Q:LB
Mon	28-Oct	Ch14.1	ANCOVA			
Tue	29-Oct	Lab8	ANCOVA	CP2003	Lab8	
Wed	30-Oct	Ch15 review	Last day to drop course			
Fri	1-Nov	Unit II Exam				All
Mon	4-Nov	Ch20.1, 20.4	Correlation & Multivariate Analysis		Lab8	L:LB
Tue	5-Nov	Lab9a	GLMM Problem setup	CSF2325	DCS	
Wed	6-Nov	Ch16, 16.2,16.4	Analysis of Deviance		A5 correlation	A:DCS
Fri	8-Nov	Ch18, 18.1	Logistic regression		Quiz 8	Q:EW
Mon	11-Nov	Holiday				
Tues	12-Nov	Lab9b	GLMM execution	CP2003	DCS+(LB,EW ?)	
Wed	13-Nov	Ch 18.2	Prospective analysis			
Fri	15-Nov	Ch 18.3	Retrospective analysis (Mon)		Quiz 9	Q:LB
Mon	18-Nov	Ch 18.6	Logistic ANCOVA		Lab 9a,b	L:EW&LB
Tue	19-Nov	Lab10 optional	Logistic regression. Optional		DCS	
Wed	20-Nov	Ch 17.1	Poisson regression			
Fri	22-Nov	Ch 17.4, 17.5	Contingency tests, Poiss ANCOVA		Quiz 10	Q:EW
Mon	25-Nov	Cn19.1-19.4	Model selection I - EDA and stepwise			
Tue	26-Nov		Review - GLMM	CSF2325		
Wed	27-Nov	Ch19.5	Multi Model Inference -AIC, SIC			
Fri	29-Nov		TBA		Quiz11	Q:EW
Mon	2-Dec	Worksheet	Course review with Q&A			
	TBA	Final Exam	2 hr Synchronous	TBA	TBA	All