BIOLOGY 4605/7220 Statistical Analysis in Biology and Environmental Science   FALL 2022

Lectures:	Mon Wed Fri 12 PM	ED 3034B
Labs:	Tue 2-5 or 6-9 PM	Online + drop in help
Instructors:	David Schneider	Office (202 Elizabeth Ave) Tel 864-8393
	84dcs@mun.ca	Hours - after class or by appointment (CSF- TBA)
	Amy Hurford	Office (CSF 4338) Tel 864-8301
	ahurford@mun.ca	Hours -

Teaching Assistant: Brendan Carswell   bmcarswell@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. 2/3 of students in 2017 considered this unnecessary.

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.
Contact the instructor if you are unable to complete evaluated work due to acceptable cause.

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 Quantitative Biology

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.

COVID disruption. In the event of a rapid rise in COVID19 cases, this course will shift to online for labs and lectures, as delivered in the Fall 2020 term.
For latest information and updates see: https://www.mun.ca/covid19/

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.

Assignments. See schedule (page 4) 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 8-point 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 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 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.

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 the course are of little use to students, as it is too late for corrective action. No-name questionnaires and polls during the course have been a regular part of these two courses since 1994. They are typically short, a few minutes before the end of a lecture period. Short summaries of the results are usually delivered orally in class.

*   *   *

Day	Date	On Web	Topic	Labs	Due	Assessed by
Wed	7-Sep	Ch1	Intro to Course			
Fri	9-Sep	Ch2.1, 2.2	Quantities			
Mon	12-Sep	Ch2.5,2.6	Units, Dimensions		A1 Quantities	A:BC
Tues	13-Sep	Lab1	Inferential Cards	CSF2313	DCS	
Wed	14-Sep	Ch3	Rescaling		Quiz 1	Q:DCS
Fri	16-Sep	Ch4	Equations		Lab1, Quiz 1	L:BC
Mon	19-Sep	Ch5	Data Eq			
Tues	20-Sep	Lab2	Equations @2 [R@6?,C2003]	CSF2313	BC/AH/DCS	
Wed	21-Sep	Ch6.1	Freq Dist I		A2,Quiz2	A:BC,Q:AH
Fri	23-Sep	Ch6.2, 6.3	Freq Dist II		Lab2	L:BC
Mon	26-Sep	Ch7, 7.1,7.2	Inference, Randomization			
Tues	27-Sep	Lab3	Freq Dist	C2003	AH/DCS	
Wed	28-Sep	Ch7.3. 7.5	Hypoth.Tests & Conf. Limits		SA/Quiz3	AH
Thur	29-Sep				Lab3	BC
Fri	30-Sep	No lecture - National Day			A3	DCS
Mon	3-Oct	Unit 1 Exam				DCS/AH/BC
Tues	4-Oct	Lab4	Randomization tests	C2003	AH/DCS	
Wed	5-Oct	Ch8, 9.1	GLM Intro, Regression I		A4, Q4	A:DCS, Q:AH
Fri	7-Oct	Ch9.2	Regression II		Lab5a	BC
Mon	10-Oct	Holiday				
Tues	11-Oct	Holiday				
Wed(M	12-Oct	Ch9.3, Ch10.2	Regression, t-test		SA/Q5	SA/Q: TBA
Fri	14-Oct	Ch10.3, 10.4	1-way ANOVA			
Mon	17-Oct	Ch11	Rev. 1 Expl Var			
Tues	18-Oct	Lab5a	Regression	C2003	AH/DCS	
Wed	19-Oct	Ch12.1	Mult Regr		SA/Quiz6	SA/Q: TBA
Fri	21-Oct	Ch13.1 (13.2)	2-way ANOVA		Lab5a only	BC
Mon	24-Oct	Ch13.3	Mixed Model - Paired t-test			
Tue	25-Oct	Lab6a	1-way ANOVA	C2003	AH/DCS	
Wed	26-Oct	Ch13.4	Mixed Model - Rand. Block		SA/Quiz7	Peer/DCS
Fri	28-Oct	Ch13.6	Nested random effects		Lab6a only	BC
Mon	31-Oct	Ch14.1	ANCOVA			
Tue	1-Nov	Lab7	Multifactor ANOVA	C2003	AH/DCS	
Wed	2-Nov	Ch15 review				
Fri	4-Nov	Unit II Exam or alternative (Wsheet for credit?)				DCS/AH/BC
Mon	7-Nov	Ch20.1, 20.4	Correlation & Multivariate Analysis		Lab7	BC
Tue	8-Nov	Lab8	ANCOVA	C2003	DCS/AH	
Wed(Fri)	10-Nov	Ch16, 16.2,3,4	Analysis of Deviance		SA/Quiz8	DCS
Fri	11-Nov			No class		
Mon	14-Nov	Ch18, 18.1	Logistic regression		Lab8	BC
Tues	15-Nov	Lab9a	GLMM Problem setup	CSF2313	AH/DCS	
Wed	16-Nov	Ch 18.2	Prospective analysis		SA/Q9	DCS
Fri	18-Nov	Ch 18.3	Retrospective analysis		Quiz	DCS
Mon	21-Nov	Ch 18.6	Logistic ANCOVA		Lab 9a,b	BC
Tue	22-Nov	Lab9b	GLMM execution			
Wed	23-Nov	Ch 17.1	Poisson regression			
Fri	25-Nov	Ch 17.4, 17.5	Contingency tests, Poiss ANCOVA		SA/Q10	DCS
Mon	28-Nov	Cn19.1-19.4	Model selection I - EDA and stepwise			
Tue	29-Nov	Ch19.5	MultiModel Inference (AIC, SIC	TBA		
Wed	30-Nov		Review - GLMM		Worksheet (credit?	DCS
Fri	2-Dec	Worksheet	Course review with Q&A		Worksheets online	
	7 Dec Exams begin		Final Exam 2 hour synchronous		Date: TBA	DCS/AH/BC
	16 Dec Exams end					

First Day Survey (No names please)   Biology 4605/7220   Fall 2022

The purpose of this survey is to gather information about student background, including experience and familiarity with online submission of activities (assignments, labs, quizzes, exams) listed on the first page. The information will be used for planning course delivery.

1. To be completed in class. If not, please print it from the website, fill it out in pen, take a photo, and mail it to David.Schneider@mun.ca.

2. Which course are you taking? Circle or highlight one.   B4605   B7220

   Grad students: list department or program   _____

   Honors students: List program _____ and topic _____

3. Using a calculator or spreadsheet can you calculate 5^4 Y / N ? $e^{0.75}$ Y / N

4. Have you ever used the R-Studio statistical package ?   Y / N
   List other statistical packages you have used.   _____ _____ _____ _____

5. Have you ever used one of the functions in a spreadsheet ?   Y / N

6. Do you have a laptop or touchscreen device (larger than a phone) that can be used in a classroom setting?   Y / N

7. How long has it been since you used logarithms ? _____

8. The variable Y has three values, Y = [9 6 3]   What is the mean value of Y ? _____

9. How many online courses have you taken? Circle or highlight one: <u>None</u> <u>Several</u> <u>Too many</u>

   B7220 only: Which platforms have you used? <u>WebEx Y/N</u> <u>Zoom Y/N</u> Other _____   ____

   B7220 only: How many courses have you taken with the
   Brightspace / D2L Learning Management System (LMS) at MUN? <u>None</u> <u>Several</u> <u>Many</u>

10a. How many university courses in mathematics? _____In statistics? _____

10b. Have you already learned how to do an ANCOVA in a previous course? Y / N

11. How many university courses where you have collected data? _____

12. How many courses where you have analyzed data?   _____

13. List (opposite side) any non-classroom experience with quantitative techniques, including any data sets you have collected.

14. Describe what you hope to learn during this course and how it relates to your academic program.