Physical & Environmental Sciences
UNIVERSITY OF TORONTO
SCARBOROUGH

EES1119 – Quantitative Environmental Analysis

Dates

Tuesdays 1-4pm. From September 14th to November 30th, 2021

Lecturer

Carlos Alberto Arnillas
Email: carlos.arnillasmerino@mail.utoronto.ca
Office hours: TBD

Teaching assistant

Coreen Daley
Email: c.daley@mail.utoronto.ca
Office hours: TBD (either Mondays 5-6, or Tuesdays 5-6)

Description

This course provides an introduction to the field of ecological statistics using R. Students will become familiar with several methods of statistical analysis of categorical and multivariate environmental data, and with the critical aspects needed to understand them, such as assumptions, underlying distribution, and decision criterion if relevant. The course will provide a comprehensive presentation of some of the most commonly used methods: analysis of variance and regression analysis, and basic concepts of ordination (principal component & non-metric multidimensional scaling) methods and Monte Carlo and Bayesian approaches. Emphasis will be placed on how these methods can be used to identify significant cause-effect relationships, detect spatiotemporal trends, and assist environment management by elucidating ecological patterns (e.g., assessment of atmospheric conditions based on pollutants, experimental identification of the role of fertilizers on natural settings).

Along the course, the students will manipulate and analyze the data using R. R is a programming language often used in environmental sciences and other academic and non-academic settings to perform statistical analysis. It is a free software with the active support of a huge community in the environmental sciences.

Learning outcomes

1. Students should be able explain the critical aspects of a statistical inference method (assumptions, core underlying distribution, statistic being tested and how it related to real data)
2. Students will be provided with the information to explain linear regressions as representation of data generation models and communicate their validity in specific conditions.
3. Students will be provided with tools to implement data validation and quality control strategies in their common practice.
4. Students should be able to integrate, summarize and share critical results of their analyses using R.
5. Students will have the opportunity to contextualize and share their expected learning process in terms of foundational knowledge, applications, integration, human dimension, caring, learning how to learn.

Course Grades

Three assignments (each 30 %)	90%
Participation	10%

The evaluation will be carried out in accordance with the Graduate Grading and Evaluation Practices Policy (and how that policy is interpreted and applied in this Dept.) http://www.governingcouncil.utoronto.ca/Assets/Governing+Council+Digital+Assets/Policies/PDF/grading.pdf

Meetings

Zoom address:

https://utoronto.zoom.us/j/85656669249 (meeting ID: 856 5666 9249)

Tuesdays Sep 14th to Nov 30th. 1-4pm

Prerequisites

No prior knowledge of environmental science or statistics is required.

Readings

The required textbook for this course is:

N.J. Gotelli and A.M. Ellison (2018). A Primer of Ecological Statistics. Oxford University Press. New York, US.

Concepts and theory	Implementing in R	Activities & assignments†
A. Orientation (Before Week 1) • Course Outline. Lecture Schedule • What is "statistics" about? • What to expect from this course?	What are R and RStudio? Installing the software	Weekly plan: *Asynchronous activities* 48h before the class: Muddiest point & *ah-ha!* moment *Synchronous activities*: Review of muddiest points R coding: peer-review, group discussion, other activities

Concepts and theory	Implementing in R	Activities & assignments†

B. Random variables and probability distributions (Weeks 1-2)
• Dealing with real data:
  ○ Metadata
  ○ Quality control
  ○ Precision and accuracy
• Types of variables: values, functions and operations
• Probabilities as a type of variable: "or", "and", "not", "conditional" operations
• Important distributions:
  ○ Discrete Random Variables. Key distributions: Bernoulli, binomial, Poisson
  ○ Continuous Random Variables. Key distributions: Uniform, normal, lognormal.
  ○ Transforming distributions: From binomial and Poisson to normal, normal to normal, normal to chi-square, chi-square to chi-square, chi-square to F.
• Describing distributions: Summary Statistics
  ○ Histograms, density plots, and cumulative density plots
  ○ Measures of Location and Spread
    ▪ Central tendency: arithmetic mean, median and mode
    ▪ Absolute distance, variance and the standard deviation
  ○ Skewness, kurtosis and central moments
  ○ Quantiles
	Notation: <-, NA, #, ?, ??
	Some good coding practices (comments and indentation)
	Variables and functions
	Data in long and wide format
	Types of variables:
	- character, factor, integer, real, Boolean
	- Vectors and matrices
	- Data frames and data.table
	Importing data:
	- read.csv(), fread(), read.xls()
	Functions
	- c(), is.na()
	- runif(), dunif(), qunif(), punif()
	- +, -, *, /, ^, …
	- &, |, <, >, !=, <=, >=, ()
	- […]
	- abs()
	- Simple plots: hist(), curve(), abline()
	Describing the distribution:
	- mean(), sd(), median()
	- mode, skewness, kurtosis, quantiles
	Libraries
	GGplot:
	- Basic logic: data + aesthetics
	- ggplot() + geom_point() + stat_function()
		Read chapters 2 and 3 (week 2)
		
		First assignment: Exploring the data

<table>
<tr><th>Concepts and theory</th><th>Implementing in R</th><th>Activities & assignments†</th></tr>
<tr><td>C. Framing and testing hypotheses-frameworks for statistical inference (Week 3-4)
• Samples and populations
• The data generation process: the underlying mechanisms and its representation
• Testing statistical hypotheses
  ○ The inductive method: Bayesian inference
  ○ The hypothetico-deductive method
  ○ Where is happiness coming from? Decision theory:
    ▪ Statistical experiment and null hypotheses
    ▪ Error types I and II. Power
    ▪ Monte Carlo analysis
    ▪ Parametric analysis (statistical vs. “biological” significance, p-values, effect size)
    ▪ Bayesian analysis
  ○ Dealing with confounding mechanisms (indices, statistical methods, experiments)
• Guided examples:
  ○ Sampling distribution, Central Limit Theorem and building a simple test
  ○ Goodness-of-fit tests</td><td>Loops: for()

Some statistical functions:
chisq.test()
ks.test()
shapiro.test()
ad.test()
lillie.test()
pearson.test()</td><td>Read chapters 4 and 5

In-class group discussion</td></tr>
</table>

Concepts and theory	Implementing in R	Activities & assignments†
D. Simple linear regression: Overview and parameter estimates (Week 5) • Distinction between regression and correlation models ○ Pearson r2 • Model and data generation process • Overview of steps in regression analysis • Estimation of regression function and error terms: Least squares	cor() cor.test() lm() coef() plot() shapiro.test()	*First assignment due date* *Second assignment: Implementing regression models*

E. Simple linear regression: assumptions, inference, and predictions (Week 6)
• Assumptions and inference in regression analysis
• Diagnostics and remedial measures:
  ○ Effects of measurement errors on X and Y
  ○ Normality: Residuals & data transformations
  ○ Heteroscedasticity: Residuals & data transformations, extra random terms
  ○ Non-independence: Correlations among the residuals
  ○ Endogeneity: Correlations among error-term and predictor
• Prediction intervals for new observations
• Inference and interpretation: When does correlation imply causation?
	anova()
	summary()
	predict()
	glm()
	logit()
	plot.lm()
	par()
	boxcox()
		Chapter 9 (before Monte Carlo and Bayesian Analysis)

Concepts and theory	Implementing in R	Activities & assignments†

F. Multiple linear regression (Week 7)
• Representing non-linear trends:
  ○ Polynomial terms
  ○ Other transformations
• Adjusting the linear model:
  ○ Multiple predictors
  ○ Interaction terms
  ○ Collinearity: Correlations among the predictors
  ○ Coefficients of partial determination
• Standardized multiple regression model
• Parsimony: Stepwise regression and other automatic search procedures for variables reduction
• Equifinality
	poly()
	step()
	AIC()
	BIC()
	anova() # between two lm models
	visreg()
	vif()
		Chapter 9 (Other Kinds of regression analyses – Multiple regression, Model selection criteria)

G. Generalized Linear Models: Logistic regression & Poisson regression (Week 8-9)
• Generalized linear models
  ○ Simple logistic regression models
  ○ Poisson regression models
• Inference and interpretation
• Estimation of regression function and error terms: Maximum likelihood
	glm()	*Chapter 9 (Other Kinds of regression analyses – Logistic regression)*

Concepts and theory	Implementing in R	Activities & assignments[†]

H. Analysis of Variance (ANOVA) (Weeks 10-11)
• Relation between regression and analysis of variance (dummy variables)
• ANOVA Model I-Fixed factor levels
• Analysis of factor level effects (pre-defined contrasts, post-hoc tests: Tukey, Scheffé, Bonferroni)
• Generalizing the linear model
  ○ Multifactorial design
  ○ Random terms
  ○ Random and mixed effects models
  ○ ANOVA Models II and III
	anova()
	# factor and controlling the reference level
	pairwise.t.test()
	TukeyHSD()
	lmer()
	Anova()
		Chapter 10
		
		Second assignment due date
		
		Third assignment: Implementing a mixed ANOVA with covariates

I. Design lab (Week 12)
• Explain, predict, describe
• Designing and implementing a model
• Common problems and how to fix them
  ○ Missing data
  ○ Outliers: Error in the data or "eureka" moment?
  ○ Missing variables and confounding mechanisms
  ○ Bidirectional causal links
• Open questions:
  ○ How to find the right approach for my problem?
  ○ Conceptual figures
  ○ Identifying assumptions, underlying distribution (if present), and interpreting the results
  ○ Testing the compliance to underlying assumptions in a relevant document
  ○ What to report in a paper?
		Third assignment due date two weeks after.
		
		Discuss:
		Shmueli 2010 To explain or to predict. Statistical Science 25(3) 289-310
		
		In class discussion:
		Use a peer-reviewed paper, find a test, discuss if all the assumptions were covered, and explain the implications if one of them was not.

Concepts and theory	Implementing in R	Activities & assignments†
J. Ordination analysis (Time dependent) • Principal component analysis: ○ Transformation (eigenvalues and eigenvectors) and interpretation • Non-metric multidimensional scaling • Advantages and disadvantages of ordination	pca() nmds()	*Chapter 12 (Ordination)*

Note: †Reading material and in-class activities may change as we move throughout the course to adapt the course pace.

Accessibility

Students with diverse learning styles and needs are welcome in this course. In particular, if you have a disability/health consideration that may require accommodations, please feel free to approach me and/or the AccessAbility Services Office as soon as possible. If you require accommodations for a disability or have any accessibility concerns about the course, or course materials, please contact the UTSC Accessibility Services as soon as possible:

https://www.utsc.utoronto.ca/ability/welcome-accessability-services

We also suggest you also refer to the following University of Toronto Scarborough Library link

http://utsc.library.utoronto.ca/services-persons-disabilities

The sooner you let us know your needs the quicker we can assist you in achieving your learning goals in this course.

Verification of illness

A ***Verification of Illness*** **(also known as a "doctor's note") is temporarily not required.** Students who are absent from academic participation for **any reason** (e.g., COVID, cold, flu and other illness or injury, family situation) and who require consideration for missed academic work should report their absence through the online absence declaration. The declaration is available on **ACORN under the Profile and Settings menu.** Students should also advise their instructor of their absence. Visit COVID-19 Information for University of Toronto Students page on the Vice-Provost, Students website for information on this and other frequently asked questions.

Emergency planning

Students are advised to consult the university's preparedness site for information and regular updates regarding procedures relating to emergency planning.

Plagiarism

University of Toronto Code of Behaviour on Academic Matters states that "it shall be an offence for a student knowingly: to represent as one's own any idea or expression of an idea or work of another in any academic examination or term test or in connection with any other form of academic work, i.e., to commit plagiarism."

For accepted methods of standard documentation formats, including electronic citation of internet sources please see the UofT writing website at http://advice.writing.utoronto.ca/using-sources/documentation.

The full Code of Behaviour regulations could be found from consulting https://www.sgs.utoronto.ca/policies-guidelines/academic-integrity-resources/

Writing and English language

You can find support at English Language and writing support at University of Toronto or the Centre for Teaching and Learning at UTSC (see also https://uoft.me/AcademicLearningSupport).