ESM 566/CE 566 Environmental Data Analysis

CRN: 11077/10468

M&W 10:00 am-11:50 am; VSC B1-08

Instructor: Yangdong Pan (email: pany@pdx.edu)

Teaching assistant: Wendy Sangucho Loachamin (email: wen27@pdx.edu)

Student hours: Monday 1:00-2:00 pm in SRTC 238F or by appointment (either in-person or via Zoom)

Syllabus Content

Health Check, Illness, Exposure or Positive Test for COVID-19

- If you are feeling sick or have been exposed to COVID-19, do not come to campus. Call The Center for Student Health and Counseling (SHAC) to discuss your symptoms and situation (503.725.2800). They will advise you on testing, quarantine, and when you can return to campus.
- If you test positive for COVID-19, do not come to campus. SHAC will advise you on quarantine, notification of close contacts and when you can return to campus.
- Please notify me, (i.e. your instructor), should you need to miss a class period for any of these reasons so that we can discuss strategies to support your learning during this time.
- If I become ill or need to quarantine during the term, either I or the department chair will notify you via PSU email about my absence and how course instruction will continue.

Guidance May Change

Please note that the University rules, policies, and guidance may change at any time at the direction of the CDC, State, or County requirements. Please review the

University's main COVID-19 Response webpage and look for emails from the University on these topics.

Course Learning Goals and Objectives:

The main purpose of this course is to help students ***think statistically*** and use statistics effectively in *framing, conducting, and reporting* their researches. At the end of the term, students are expected to understand and be able to use statistical analyses within an overall environmental research framework (i.e., conceptual model, problem formulation, sampling/study design, model parameter estimates, hypothesis testing, statistical inference, and environmental conclusions). Specifically, students are expected to

- Design statistically sound field studies and lab experiments
- Graphically and numerically summarize the data
- Select appropriate statistical methods, and be able to assess the assumptions and find possible remedies if the assumptions are violated
- Draw statistical conclusions and understand the uncertainty associated with each statistical decision (e.g., Type I and II error)
- Communicate scientific results effectively
- Critically interpret statistical analyses in the environmental research literature
- Use statistical computing programs, specifically, R, effectively to analyze environmental data

Textbook and Reading Materials:

- Gotelli, N. J. and A. M. Ellison. 2013. *A Primer of Ecological Statistics*. Sinauer Associates, Inc. Publishers, Sunderland, MA. (2nd edition) (required)
- Additional journal articles will be distributed on Canvas.

Software/APP:

- *R* (free downloadable from <https://cran.r-project.org/>. This is the 'engine' for all data analysis and programming that we will do in this class. Please install the software to your computer before the term starts.
- *RStudio*: a text editor for R and others (free downloadable from <http://rstudio.org/download/desktop>). This is a very user-friendly platform for us to write R script, execute them, save all work, and much more. We will primarily use *RStudio* for most our work in the class. Please install the software to your computer before the term starts. Please watch this short video on how to get start.
- For the basic R tutorial to get a start with R, please go to <http://www.cyclismo.org/tutorial/R/>. For more helpful documents and

webinars on *RStudio*, please go to <http://rstudio.org/docs/> and <RStudio Webinars - RStudio>.

- *Canvas*: All lecture PowerPoint slides, worksheets, data sets, homework assignments and additional reading materials will all be posted on Canvas.

Useful R Resources:

The number of materials on R available for users is overwhelming. This class will start with very basic introduction of R with weekly worksheets with the assumption that every student doesn't know to use R. If you would like to explore more on R on your own, you may start with the following's sites:

- The Comprehensive R Archive Network for free R reference manuals/books and other information <http://cran.cnr.berkeley.edu/>
- UCLA's IDRE Institute for Digital Research and Education: Rich resources for helping you learn and use R
- Stack overflow: many R-related questions and answers.
- R-Bloggers: R news and tutorials contributed by >500 R bloggers
- The R Journal: an open access, refereed journal of the R project for statistical computing
- The R Graph Gallery: a collection of 50 types of graphs and associated R scripts.

Approach:

This course emphasizes tremendously student-led and group-based learning. Each class period will start with a brief introduction, followed by an in-class hands-on group-based exercise with worksheets, and then frequent class discussion. The class will be divided into groups, each with 2-3 members. Each group will work together on the assignments during the class, and present/critique the results and interpretation among the group members.

1. **PowerPoint slides with audio**: Most of course contents will be delivered via PowerPoint slides coupled with audiotaped explanation by the instructor. The slides with audio will be loaded to Canvas weekly and students are encouraged to go through the slides prior to attending the class.
2. **Worksheets and in-class exercises**: A worksheet with several in-class exercises will be distributed weekly on Canvas. A large amount of each class period will be allocated for group-based hands-on in-class exercises. In other words, during each class period, students are encouraged to work as a group on each of the in-class exercises and discuss any issues with the instructor or GTA.
3. **Miniature lectures**: The instructor will limit the lecture to 10-15 minutes each time. Each class will start with a short introduction and then follow by in-class exercises using either students' own laptops or desktops in B1-82 (we do have multiple desktops available). When students run into similar issues during the exercise, the instructor will start to lecture to address the issues to the entire class.
4. **Research project**: ***Please see the details at the end of the syllabus***.

5. **Peer-evaluation**: Since the class emphasizes tremendously on teamwork and student-based learning, each member will have a chance to evaluate their peers' performance at the end of the term. The outcome of the peer evaluation will affect a student's final grade.

Grades:

- *NO EXAMS*
- "*Homework*" (3 homework): 60%, late homework with no acceptable excuse will still be accepted but with penalty. If you need an extension for homework, please discuss it with your TA prior to the submission deadline.

Homework #1 Hypothesis testing if watershed geology affects water quality impairment (Turbidity) in Oregon coastal streams (5 points). *Available on Oct.2 and due on Oct. 9.*
Homework #2 Simple linear regression model on the effects of urbanization on stream biota (10 points). *Available on Oct.9 and due on Oct.16.*

Homework #3 Multiple linear regression model on the relationships between the SO_2 concentrations and environmental predictors in 41 US cities (15 points). *Available on Oct.16 and due on Oct.30.*

- *Class and group participation*: 5%

Final project (35%): You are required to formulate a conceptual model, study question, collect/"borrow" data, analyze the data and interpret the results with relation to the study question, and write a professional research paper. Please see more details in **the Group-based Research Project** on Page 6.

Grading Scale (percent scores and grade break points for letter grades):
A: "excellent", comprehensive knowledge and understanding of subject matter;
B: "good", moderately broad knowledge and understanding of subject matter;
C: "satisfactory", reasonable knowledge and understanding of subject matter;
D: "inferior", minimum knowledge and understanding of subject matter
A: 94–100; A^-: 90–93; B^+: 87–89; B: 84–86; B^-: 80–83; C^+: 77–79; C: 74–76; C^-: 70–73

Pass: C^- or above
Incomplete: Departmental and university policies dictate that incompletes can be given only for verified medical reasons (through the Office of the Dean of Student Life).

Statement on Academic Honesty

Plagiarism of any form will not be tolerated in this class and will result in failing grades for the assignment and course participation, and a referral to the Office of the Dean of Student Life. For more information on academic integrity and misconduct, please see the Portland State University's Bulletin and the Office of the Dean of Student Life.

Tentative Course Outline

Both lecture and workshop topics and their orders will be subject to changes depending on students' interests and their data sets.

Week	Topics	Reading
1	Course introduction; Exploratory data analysis: Summarize data numerically Display data graphically	 Ch. 8 (p.212-236)
2	Hypothesis testing Analysis of Variance (ANOVA) Experimental design, Statistical power;	Ch. 4 (p.79) Ch. 10 (p.288-300) Ch. 10 (p.300-308)
3	Simple ordinary least-square linear models: Estimation and inference Residual analysis and diagnostic checks	Chapter 9 (p239-264)
4	Multiple linear models: Variance inflation factor and multicollinearity Step-wise regression hypothesis-testing	Chapter 9 (p275) Chapter 9 (p282)
5	Generalized Least Squares (GLS) models for Dependent data Generalized linear models (GLM): Logistic regression models	Zuur et al. (2009) Guisan et al. (2002) *Ecol. Modeling 157*
6	Polynomial regression and LOESS Generalized additive models (GAM)	
7	Multi-level regression models for nested data Random intercept models Random slope models Random intercept and slope models	Zuur et al. (2009)
8	Tree-based models: Regress tree models; Classification tree models; Random Forests and Boosted models	De'ath & Fabricius (2000) *Ecology 81*
9	Time-series: Autoregression (AR) Moving-averaging (MA) ARIMA models	

10	Frequentist vs. Bayesian data analysis
11	**Final paper due** (Dec. 9, 2024, Monday by Midnight)

Group-based Research Project

Group-based research projects allow students to work together and use what they learn from the class to solve a real-life environmental problem. Several research papers from past students using different statistical methods are posted on Canvas for your reference.

- *Research group*: By week three, 2-3 students will form a research group on their own.
- *Research*: Each group will then formulate a research question based on their shared research interest, construct a conceptual model based on the relevant literature, identify a dataset which is suitable for addressing their research question (it is preferred that the students use their own research datasets), select appropriate data analyses, perform the analyses on the datasets, interpret the results both statistically and scientifically. Each group is required to write a paper based on their research results following a scientific format. **This process is typically more time-consuming than we usually expected and thus act early.**

A suggested time frame for the group-based research will be:

- Week 3: form a research group and formulate a research question
- Week 4: build a conceptual model based on relevant literature
- Week 5: secure a dataset and start exploratory data analysis to get familiar with the dataset
- Week 6-7: data analysis and interpretation
- Week 8: writing a draft paper

- *Peer-review*: A complete draft paper will be due in Week 9 for peer review. Both instructor and a selected peer group will critically review the draft paper and provide constructive comments and suggestions. Each group will review the assigned draft paper together and submit one critique as well as any edits/suggestions marked on the original manuscript using "Track Changes". Both critique and marked manuscript will be on Canvas.
- *Final revision*: Each group should take all peer review comments and suggestions into consideration during the revision process. Each group will submit a table of "Point-to-point" responses to review comment. If you strongly disagree with the comments or suggestions, you are required to

provide a detailed rationale on why a particular major comment/suggestion is rejected in the revision.

- *Final paper due*: The final revised paper will be due on **Week 11 (Dec. 9, 2024)**

Grading Guidelines on the ESM/CE 566/666 Final Project Paper

Title & Abstract (2 pts)
Does the abstract concisely summarize the study with (1) purpose (2) research objective/question/hypothesis (3) study design (4) major findings and (5) conclusion(s)?

Introduction (5 pts)
Is the purpose of the paper clear in the introduction?
Does the author summarize the current knowledge on the subject using a conceptual model?
Does the author effectively use the conceptual model to provide enough background information that it is very clear to the reader on why this study is necessary?
Is the question well formulated and unambiguous?

Methods (10 pts)
Does the author clearly explain the study design (e.g., experiment/study unit, data collection, sample size, site selection criteria, etc. Hint: the best way is to use a diagram to illustrate the study design)?
Are the study design and methods appropriate for the research questions/hypotheses?
How does the author prepare the data for the analysis (e.g., missing data, outliers, test for assumptions and remediation if required)? Is each decision on the data preparation defensible with enough rationale?
Does the author clearly explain how the data are analyzed (e.g., using Generalized least square regression) and the rationale for selecting a particular method over the others (e.g., Generalized least square regression vs. ordinary least square regression)? In other words, how this method is suitable for addressing the research question? Does the author make an explicit linkage between each analytic analysis and the study objective/question?
Is the analytical method appropriate for the study design? Does the author describe ALL methods used in this study?
Can the reader repeat the same analysis according to this section?

Results (10 pts, the number of figures<4, the number of tables<3)
Does the author adequately characterize the patterns and trends in the data prior to any statistical analysis? In theory, if there is any pattern or trend, we should characterize them without using statistics?
Does the author provide too many results which are not well related to the study objectives?
Does the author provide key findings with relation to the study question?
Does the author report key results from each key data analysis (e.g., P-value and R2 for regression)?
Does the author present the key finding effectively (using appropriate illustrations such as figures or tables)?
Is every figure or table included in the paper referred to in the text?

Does each figure/table have a complete caption, axis labels, so that each figure/table can stand alone?
Does the author provide any results from the analysis which is not described in the methods section?
Does the author present the results only in this section (with no discussion/explanations and methods)?

Discussion (8 pts)
Does the author discuss the major findings with relation to the study objective?
Is the conclusion fully supported by the results?
Does author provide any logical and meaningful interpretation of the results?
Has the author been objective in the discussion of the topic?
Is all the discussion relevant to the study questions?
If the results are negative (contrast to the author's original expectations), does the author adequately discuss major possible reasons and provide any leads for further studies?
Does the author simply repeat all results in this section?

PSU Student Resources

- Title IX reporting
- PSU Prohibited Discrimination & Harassment Policy
- Disability accommodations and the Disability Resource Center
- Dean of student life
- Religious accommodations policy
- Library
- Writing Center
- Food assistance
- General PSU Policies (e.g., Student Conduct and Responsibility Policy)
- Student Resources and Centers (e.g., campus public safety, veterans resource center, etc.)
- Sanctuary campus information and resources
- DACA resources