

课程详述

COURSE SPECIFICATION

以下课程信息可能根据实际授课需要或在课程检讨之后产生变动。如对课程有任何疑问，请联系授课教师。

The course information as follows may be subject to change, either during the session because of unforeseen circumstances, or following review of the course at the end of the session. Queries about the course should be directed to the course instructor.

1.	课程名称 Course Title	Environmental Data Analysis
2.	授课院系 Originating Department	School of Environmental Science and Engineering
3.	课程编号 Course Code	ESE335
4.	课程学分 Credit Value	3.0
5.	课程类别 Course Type	Major Elective Courses
6.	授课学期 Semester	Spring
7.	授课语言 Teaching Language	English
8.	授课教师、所属学系、联系方式（如属团队授课，请列明其他授课教师） Instructor(s), Affiliation & Contact (For team teaching, please list all instructors)	Lei ZHU School of Environmental Science and Engineering Taizhou Hall 429 zhul3@sustech.edu.cn 13342959636
9.	实验员/助教、所属学系、联系方式 Tutor/TA(s), Contact	To be announced
10.	选课人数限额(可不填) Maximum Enrolment (Optional)	

11. 授课方式 Delivery Method	讲授 Lectures	习题/辅导/讨论 Tutorials	实验/实习 Lab/Practical	其它(请具体注明) Other (Please specify)	总学时 Total
学时数 Credit Hours	48		0		48
12. 先修课程、其它学习要求 Pre-requisites or Other Academic Requirements	None				
13. 后续课程、其它学习规划 Courses for which this course is a pre-requisite	None				
14. 其它要求修读本课程的学系 Cross-listing Dept.	None				

教学大纲及教学日历 SYLLABUS

15. 教学目标 Course Objectives

As an interdisciplinary, Environmental Science gains insights from various data sets of field studies, lab experiments, remote sensing, and model simulations. Analyzing and visualizing data sets has become one of the most critical skills for carrying out Environmental Science studies. However, senior ESE undergraduate students often find their opportunities to access such courses confined, and specific training toward developing such skills limited.

This course will teach students how to apply suitable statistical methods and visualization tools to analyze environmental data sets. Topics include environmental data sets characteristics, checking data sets, comparisons between two groups, comparisons among several groups, correlation, simple linear regression, multiple linear regression, logistic regression, time series analysis, and spatial data analysis. Students will also learn how to conduct data analysis properly with the R language.

16. 预达学习成果 Learning Outcomes

This course facilitates student learning through lectures, exercises, labs, assignments, final project, and one-on-one interactions. Students should be able to analyze and visualize environmental data sets using suitable statistical methods and R tools. This course would also boost students' programming skills, which are broadly applicable in their later study and research.

17. 课程内容及教学日历 (如授课语言以英文为主, 则课程内容介绍可以用英文; 如团队教学或模块教学, 教学日历须注明主讲人)

Course Contents (in Parts/Chapters/Sections/Weeks. Please notify name of instructor for course section(s), if this is a team teaching or module course.)

Section 1 Introduction (2 credit hours).

Course logistics, topics to be covered, grading policy, R tutorial

Section 2 Review of statistic basics (4 credit hours)

Probability density function, distributions, statistical hypothesis, random draw, confidence interval, p-value, R programming exercises

Section 3 Characteristics of environmental data sets (3 credit hours)

Types of environmental data sets, format of environmental data sets, normal distribution, log normal distribution, log transformation, detection limit, missing values, R programming exercises

Section 4 Checking data sets: Quick summaries (2 credit hours)

Mean, median, quantile, standard deviation, variance, outlier, R programming exercises

Section 5 Checking data sets: Quick plots (4 credit hours)

Histogram, barplot, boxplot, scatterplot, time series plot, image plot, surface maps, R programming exercises

Section 6 Comparisons between two groups: t-tools (3 credit hours)

t distribution, assumptions of t-test, comparing means of two groups, R programming exercises

Section 7 Comparisons between two groups: Alternatives to t-tools (3 credit hours)

Rank-Sum test, permutation test, Welch t-test, sign test, signed-rank test, R programming exercises

Section 8 Comparisons among several groups (3 credit hours)

One-way ANOVA, F-test, two-way ANOVA, R programming exercises

Section 9 Linear combinations and multiple comparisons of means (3 credit hours)

Linear combinations of group means, multiple comparison procedures, R programming exercises

Section 10 Correlation and simple linear regression (3 credit hours)

Pearson's test, Spearman's test, Kendall's test, simple linear regression, least squares regression estimation, R programming exercises

Section 11 Assumptions for simple linear regression (3 credit hours)

Robustness of least squares inferences, model assessment, fit assessment, R programming exercises

Section 12 Multiple linear regression (3 credit hours)

Least squares estimates, model assessment, fit assessment, R programming exercises

Section 13 Over-fitting and variable selection (3 credit hours)

Over-fitting, AIC, BIC, backward selection, forward selection, step-wise selection, R programming exercises

Section 14 Logistic regression (3 credit hours)

Binary responses, binomial responses, Poisson responses, building logistic regression model, R programming exercises

Section 15 Time series analysis (3 credit hours)

MA, AR, seasonal decomposition, ARIMA, forecasting, R programming

Section 16 Spatial data analysis (3 credit hours)

Raster and shape files, interpolation, spatial regression, R programming

Total credit hours: 48

18. 教材及其它参考资料 Textbook and Supplementary Readings

Agresti A., C. Franklin, and B. Klingenberg, **Statistics: The Art and Science of Learning From Data**, Pearson, 4th Edition (January 7, 2016), 816 pages, ISBN-10: 0133860825, ISBN13: 978-0133860825.

Freedman D., R. Pisani, and R. Purves, **Statistics**, W. W. Norton & Company, 4th Edition (February 13, 2007), 720 pages, ISBN-10: 0393929728, ISBN-13: 978-0393929720.

Hastie, T., R. Tibshirani, and J. Friedman, **The Elements of Statistical Learning: Data Mining, Inference, and Prediction**, Springer, 2nd Edition (January 1, 2016), 767 pages, ISBN-10: 0387848576, ISBN-13: 978-0387848570.

James G., D. Witten, T. Hastie, and R. Tibshirani, **An Introduction to Statistical Learning: with Applications in R**, Springer, 1st ed. 2013, Corr. 7th printing 2017 Edition (June 25, 2013), 440 pages, ISBN-10: 1461471370, ISBN-13: 978-1461471370.

Ramsey F. and D. Schafer, **The Statistical Sleuth: A Course in Methods of Data Analysis**, Cengage Learning, 3rd Edition (May 2, 2012), 784 pages, ISBN-10: 1133490670, ISBN-13: 978-1133490678.

Shumway R. and D. Stoffer, **Time Series Analysis and Its Applications With R Examples**, Springer, 4th ed. 2017 Edition (April 19, 2017), 575 pages, ISBN-10: 3319524518, ISBN-13: 978-3319524511.

课程评估 ASSESSMENT

19. 评估形式 Type of Assessment	评估时间 Time	占考试总成绩百分比 % of final score	违纪处罚 Penalty	备注 Notes
出勤 Attendance				
课堂表现 Class Performance		10		
小测验 Quiz				
课程项目 Projects				
平时作业 Assignments		30		
期中考试 Mid-Term Test				
期末考试 Final Exam		30		
期末报告 Final Presentation		30		
其它（可根据需要 改写以上评估方 式） Others (The above may be modified as necessary)				

20. 记分方式 **GRADING SYSTEM**

- A. 十三级等级制 **Letter Grading**
 B. 二级记分制（通过/不通过） **Pass/Fail Grading**

课程审批 **REVIEW AND APPROVAL**

21. 本课程设置已经过以下责任人/委员会审议通过
This Course has been approved by the following person or committee of authority