

统计数据分析（SAS）课程大纲

- 1、2020 年春季学期--2021 年春季学期
- 2、2022 年春季学期起

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	统计数据分折(SAS) Statistical Data Analysis with SAS				
2.	授课院系 Originating Department	数学系 Department of Mathematics				
3.	课程编号 Course Code	MA409				
4.	课程学分 Credit Value	3				
5.	课程类别 Course Type	专业选修课 Major Elective Courses				
6.	授课学期 Semester	春季 Spring 【2020 年春季学期--2021 年春季学期】				
7.	授课语言 Teaching Language	英文 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)	田国梁教授 Professor Guoliang TIAN 数学系 Department of Mathematics				
9.	实验员/助教、所属学系、联系方式 Tutor/TA(s), Contact	无 NA / 待公布 To be announced / 已确定的实验员/助教联系方式 Please list all Tutor/TA(s) (请保留相应选项 Please only keep the relevant information)				
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				48

12. 先修课程、其它学习要求 Pre-requisites or Other Academic Requirements	统计线性模型 (MA329) Statistical Linear Models(MA329)
13. 后续课程、其它学习规划 Courses for which this course is a pre-requisite	
14. 其它要求修读本课程的学系 Cross-listing Dept.	

教学大纲及教学日历 SYLLABUS

15. 教学目标 **Course Objectives**

在已经学过的统计方法和统计建模的基础上，本课程将会使这些学生更深入的理解数据分析的整个过程。它旨在发展学生的模型选择技术，使得手中的实际问题能够被合适地转化为假设检验问题。最重要的是当用第一个模型拟合数据发现不合适时，怎样选择出适合的模型。学生将会学习怎样探索数据，如何建立可靠的模型以及如何清楚阐释统计分析的结果。

Building on prior coursework in statistical methods and modeling, students will obtain a deeper understanding of the entire process of data analysis. The course aims to develop skills of model selection so that practical questions at hand can be properly formulated as statistical null and alternative hypotheses. An important step is how to select a reasonable model, when one's first attempt does not adequately fit the data. Students will learn how to explore the data, to build reliable models, and to communicate the results of data analysis to a variety of audiences.

16. 预达学习成果 **Learning Outcomes**

On successful completion of the course, students should be able to:

- make good sense of the problem and identify what to measure for the question of interest;
- summarize and describe the quantitative and qualitative data using some simple appropriate statistical measures;
- identify the association among several continuous or discrete variables;
- carry out appropriate and comprehensive statistical analyses based on real life data using SAS including model selection, perform model diagnostics, formulate testable hypotheses, make appropriate statistical inferences, make interpretations on the findings and report writing.

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.)

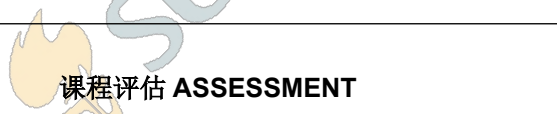
课程内容：描述性统计量，数据陈述与可视化，用参数统计方法进行单样本和双样本情况下的简单统计分析，用非参数统计方法进行单样本和双样本情况下的简单统计分析；回归分析，模型拟合，变量选择和模型诊断；单因子、双因子、多因子方差分析；协方差分析 logistic 回归和 Poisson 回归。用 SAS 软件进行实例数据建模与分析使学生能够获得第一手经验。

This course covers: descriptive statistics, presentation and visualization of data; Simple statistical analyses for the one-sample and two-sample case using parametric and nonparametric methods; Regression analyses: model fitting; variable selection and model diagnostic checking; Analysis of Variance (ANOVA): 1-way, two-way and higher-way ANOVA; Covariance analysis; Categorical and count data: binary logistic regression, Poisson regression. Real data sets will be presented for modelling and analysis using statistical software for gaining hands-on experience.

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

In this course, no single textbook can cover all the topics. Relevant references are as follows:

- [1] Ramsey, F. and Schafer, D. (2012). The Statistical Sleuth: A Course in Methods of Data Analysis, 3rd edition. Cengage Learning.
- [2] Cody, R. (2011). SAS Statistics by Example. SAS Institute.
- [3] Cody, R.P. and Smith, J.K. (2005). Applied Statistics and the SAS Programming Language, 5th edition. Pearson.
- [4] Elliott, R.J. (2009). Learning SAS in the Computer Lab, 3rd edition. Cengage Learning.
- [5] Kleinbaum, D.G., Kupper, L.L., Nizam, A. and Muller, K.E. (2007). Applied Regression Analysis and Other Multivariable Methods, 4th edition. Cengage Learning.



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				
小测验 Quiz				
课程项目 Projects				
平时作业 Assignments				
期中考试 Mid-Term Test				

期末考试
Final Exam
期末报告
Final Presentation
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



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	统计数据分折(SAS) Statistical Data Analysis with SAS
2.	授课院系 Originating Department	统计与数据科学系 Department of Statistics and Data Science
3.	课程编号 Course Code	MA409
4.	课程学分 Credit Value	3
5.	课程类别 Course Type	专业必修课 Major Compulsory Courses
6.	授课学期 Semester	春季 Spring 【2022 春季学期起】
7.	授课语言 Teaching Language	双语 Chinese and 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)	徐匆 Cong Xu 统计与数据科学系 Department of Statistics and Data Science Email: xuc6@sustech.edu.cn
9.	实验员/助教、所属学系、联系方式 Tutor/TA(s), Contact	待公布 To be announced
10.	选课人数限额(可不填) Maximum Enrolment (Optional)	50

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			48
12. 先修课程、其它学习要求 Pre-requisites or Other Academic Requirements	统计线性模型 (MA329) Statistical Linear Models (MA329)				
13. 后续课程、其它学习规划 Courses for which this course is a pre-requisite					
14. 其它要求修读本课程的学系 Cross-listing Dept.					

教学大纲及教学日历 SYLLABUS

15. 教学目标 Course Objectives

在已经学过的统计方法和统计建模的基础上，本课程将会使学生更深入的理解数据分析的整个过程。它旨在培养学生的综合数据分析能力，能够将手中的实际问题转化为假设检验问题并使用合适的统计模型进行分析。最重要的是当通过一个模型的拟合数据发现模型不合适时，怎样选择出更为适合的模型。学生将会学习怎样探索数据，如何建立可靠的模型以及如何清楚阐释统计分析的结果。

Building on prior coursework in statistical methods and modeling, students will obtain a deeper understanding of the entire process of data analysis. The course aims to develop skills of comprehensive data analysis so that practical problems at hand can be properly formulated as statistical hypotheses and analyzed with statistical models. An important step is to select a reasonable model when one's first attempt does not adequately fit the data. Students will learn how to explore the data, to build reliable models, and to communicate the results of data analysis to a variety of audiences.

16. 预达学习成果 Learning Outcomes

通过本课程的学习，学生预期可以学会以下技能：

- 将实际问题转化为统计问题，并使用合适的方法进行数据分析；
- 深入理解回归分析、方差分析、广义线性模型等常用统计方法与其应用场景；
- 能够对模型进行比较与选择，并针对模型的假设进行诊断；
- 使用 SAS 软件进行系统的、合理的统计数据分析，并清楚阐述数据分析的结果；

On successful completion of the course, students should be able to:

- Formulate practical problems as statistical problems and perform data analyses properly;
- Have a deep understanding of commonly used statistical methods and their application scenarios, e.g. regression analysis, analysis of variance, generalized linear model;
- Perform model selection and model diagnosis;
- Use SAS to carry out appropriate and comprehensive statistical data analyses and provide clear interpretations on the findings.

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.)

第一章：简介（4学时）

- 1.1 数据分析的一般步骤
- 1.2 数据类型与常用分布
- 1.3 描述性统计量与探索性分析

第二章：SAS 基础（6学时）

- 2.1 数据导入与处理
- 2.2 基础 SAS 程序与函数
- 2.3 数据可视化

第三章：基础假设检验（8学时）

- 3.1 两组连续数据的比较
- 3.2 两组分类数据的比较
- 3.3 多重比较

第四章：回归分析 I – 模型拟合（8学时）

- 4.1 多元线性回归模型
- 4.2 最小二乘法
- 4.3 极大似然估计
- 4.4 模型统计推断

第五章：回归分析 II – 变量选择与模型诊断（8学时）

- 5.1 变量选择
- 5.2 模型诊断
- 5.3 异常观测值识别
- 5.4 复共线性检测

第六章：方差分析（6学时）

- 6.1 单因子方差分析
- 6.2 两因子方差分析
- 6.3 协方差分析

第七章：广义线性模型（8学时）

- 7.1 逻辑回归
- 7.2 泊松回归

Chapter 1: Introduction (4 hours)

- 1.1 General data analysis process
- 1.2 Types of Data and common distributions
- 1.3 Descriptive Statistics and exploratory analysis

Chapter 2: A Brief Introduction to SAS (6 hours)

2.1	Data import and manipulation
2.2	Basic SAS procedures and functions
2.3	Data visualization
Chapter 3: Basic Hypothesis Testing (8 hours)	
3.1	Comparing two groups for continuous data
3.2	Comparing two groups for categorical data
3.3	Multiple comparison
Chapter 4: Regression Analysis I – Model Fitting (8 hours)	
4.1	The multiple linear regression model
4.2	The least squares method
4.3	Maximum likelihood estimation
4.4	Model inference
Chapter 5: Regression Analysis II – Variable Selection and Model Diagnosis (8 hours)	
5.1	Variable selection
5.2	Model diagnosis
5.3	Unusual observation identification
5.4	Multicollinearity detection
Chapter 6: Analysis of Variance (6 hours)	
6.1	One-way analysis of variance
6.2	Two-way analysis of variance
6.3	Analysis of covariance
Chapter 7: Generalized Linear Models (8 hours)	
7.1	Logistic regression
7.2	Poisson regression

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

Textbook:
Ramsey, F. and Schafer, D. (2012). The Statistical Sleuth: A Course in Methods of Data Analysis, 3rd edition. Cengage Learning.
Supplementary Readings:
[1] Cody, R. (2011). SAS Statistics by Example. SAS Institute.
[2] Cody, R.P. and Smith, J.K. (2005). Applied Statistics and the SAS Programming Language, 5th edition. Pearson.
[3] Elliott, R.J. (2009). Learning SAS in the Computer Lab, 3rd edition. Cengage Learning.
[4] Kleinbaum, D.G., Kupper, L.L., Nizam, A. and Muller, K.E. (2007). Applied Regression Analysis and Other Multivariable Methods, 4th edition. Cengage Learning.

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				
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		35		
期末报告 Final Presentation		35		
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