

课程大纲 COURSE SYLLABUS

1.	课程代码/名称 Course Code/Title	分类数据分析 Categorical Data Analysis
2.	课程性质 Compulsory/Elective	专业选修课 Major Elective Course
3.	课程学分/学时 Course Credit/Hours	3 学分/48 学时, 3 Credits/48 Hours
4.	授课语言 Teaching Language	英文 English
5.	授课教师 Instructor(s)	焦熙云 Xiyun Jiao
6.	是否面向本科生开放 Open to undergraduates or not	是 Open to undergraduate students
7.	先修要求 Pre-requisites	(如面向本科生开放, 请注明区分内容。 If the course is open to undergraduates, please indicate the difference.) 数理统计 (MA204) 统计线性模型 (MA329) Mathematical Statistics (MA204) Statistical Linear Models (MA329)
8.	教学目标 Course Objectives	(如面向本科生开放, 请注明区分内容。 If the course is open to undergraduates, please indicate the difference.) 本课程旨在系统性的介绍针对不同分类数据的模型及相应的推断方法, 使得学生在掌握理论的基础上具备使用软件进行实际数据分析的能力。具体授课内容包含以下几个方面: 分类数据介绍, 列联表的描述和推断, 广义线性模型, Logistic 回归模型, Logit 及 loglinear 模型, 针对特殊分类响应数据的模型和推断方法, 以及广义线性混合模型。每个主要章节均设置软件演示环节。 This course aims at systematically introducing the models and corresponding inference methods for various types of categorical data. After the course, the students are expected to not just have a good knowledge of the related theories, but also gain the ability to analyse real data with software. Specifically, the course will consist of the following sessions: introduction to categorical data, description and inference of contingency tables, generalized linear models, logistic regression models, logit and loglinear models, models and the related inference for special categorical response data, and generalized linear mixed models. Each part will be accompanied by a software demonstration session.
9.	教学方法 Teaching Methods	(如面向本科生开放, 请注明区分内容。 If the course is open to undergraduates, please indicate the difference.) 教师将采用理论与实践相结合的方式教授这门课程。具体而言, 在讲解理论与方法的基础上, 教师将向学生展示如何通过软件来解决实际中的分类数据分析问题, 并给学生提供充足的练习机会。 When teaching this course, the lecturer will combine theories with practice. Specifically, after introducing the theories and methods, the lecturer will demonstrate how to use software to implement categorical data analysis in practice and provide sufficient opportunities for the students to practice on their own.
10.	教学内容 Course Contents	

(如面向本科生开放, 请注明区分内容。 If the course is open to undergraduates, please indicate the difference.)

<p>Chapter 1: Introduction to Categorical Data 分类数据介绍</p>	<p>Description, modelling and inference of categorical data (2 hours). 分类数据的描述, 建模和推断 (2 学时)。</p>
<p>Chapter 2: Contingency Tables 列联表</p>	<p>Description and inference of contingency tables. Specifically, introducing the probability structure, parameter estimation and hypothesis testing of contingency tables by starting from two-way tables and extending to multiway ones; A software demonstration session included (8 hours). 列联表的描述和推断。 具体而言, 从二向列联表开始, 向多项列联表扩展, 介绍有关列联表的概率结构、参数估计和假设检验; 包含软件演示环节 (8 学时)。</p>
<p>Chapter 3: Generalized Linear Models 广义线性模型</p>	<p>Describing generalized linear models for binary and counts data, and introducing the likelihood and inference methods of generalized linear model (4 hours). 描述针对二项和记数数据的广义线性模型, 并介绍广义线性模型的似然函数和推断方法 (4 学时)。</p>
<p>Chapter 4: Logistic Regression Logistic 回归模型</p>	<p>Interpreting parameters in logistic regression models, describing the fitting methods, and introducing the building and selection of logistic regression models; A software demonstration session included (6 hours). 诠释 logistic 回归模型中的参数, 描述此类回归模型的拟合方法, 并介绍 logistic 回归的模型构建和选择; 包含软件演示环节 (6 学时)。</p>
<p>Chapter 5: Logit and Loglinear Models Logit 及 loglinear 模型</p>	<p>Introducing the logit models for multinomial responses and the corresponding inference methods, the loglinear models for contingency tables and the corresponding inference methods, and also the building and extension of logit/loglinear models; A software demonstration session included (10 hours). 介绍针对多项数据的 logit 模型及相应推断方法, 针对列联表的 loglinear 模型及相应推断方法, 以及 logit/loglinear 的模型构建和扩展; 包含软件演示环节 (10 学时)。</p>
<p>Chapter 6: Special Models and the Related Inference 特殊模型和相关的推断方法</p>	<p>Models and the related inference for special categorical response data, including matched pairs and repeated response data; A software demonstration session included (6 hours). 针对特殊分类响应数据的模型和推断方法, 包括配对数据及重复响应数据; 包含软件演示环节 (6 学时)。</p>
<p>Chapter 7: Random Effect Models 随机效应模型</p>	<p>Generalized linear mixed model (random effect) and other mixture models for categorical data: introducing the generalized linear mixed models for clustered, binary and multinomial data, and the corresponding fitting and inference methods; in addition, briefly introducing other mixture models for categorical data; A software demonstration session included (8 hours). 广义线性混合模型 (随机效应) 及其他针对分类数据的混合模型: 介绍针对聚类数据、二项数据和多项数据的广义线性混合模型及相应的拟合</p>

	推断方法；另外简单介绍其他针对分类数据的混合模型；包含软件演示环节（8学时）。
Chapter 8: Other Related Problems 其他相关问题	Other related problems and review (4 hours). 其他相关问题及回顾复习（4学时）。
11. 课程考核 Course Assessment	
	<p>(①考核形式 Form of examination; ②.分数构成 grading policy; ③如面向本科生开放, 请注明区分内容。 If the course is open to undergraduates, please indicate the difference.)</p> <p>考勤: 5% Attendance: 5% 平时作业: 40% Assignments: 40% 课程项目: 30% Course project: 30% 期末考试: 25% Final exam: 25%</p>
12. 教材及其它参考资料 Textbook and Supplementary Readings	
	<p>教材 (Textbook):</p> <p>[1] Alan Agresti (2013), Categorical Data Analysis, 3rd ed., Wiley and Sons</p> <p>参考资料 (References):</p> <p>[2] Maura Stokes, Charles Davis & Gary Koch (2012), Categorical Data Analysis Using SAS, 3rd ed., SAS Institute. [3] Alan Agresti (2019), An Introduction to Categorical Data Analysis, 3rd ed., Wiley and Sons. [4] Alan Agresti (2010), Analysis of Ordinal Categorical Data, 2nd ed., Wiley and Sons.</p>