

## 《语言数据统计与分析》课程教学大纲

课程基本信息 (Course Information)					
课程代码 (Course Code)		*学时 (Credit Hours)	32	*学分 (Credits)	2
*课程名称 (Course Name)	(中文) 语言数据统计与分析 (英文) Statistical Analysis and Management of Linguistic Data				
课程类型 (Course Type)	微专业基础必修课				
授课对象 (Target Audience)	微专业学员				
授课语言 (Language of Instruction)	英汉双语, 英语为主				
*开课院系 (School)	上海交通大学外国语学院				
先修课程 (Prerequisite)		后续课程 (post)			
*课程负责人 (Instructor)	吴诗玉	课程网址 (Course Webpage)			
*课程简介 (中文) (Description)	<p>本课程为本科生专业选修课, 主要面向语言学、应用语言学及相关领域的学生。课程旨在培养学生使用统计方法进行语言数据分析的能力, 为后续的研究工作奠定基础。</p> <p>主要教学内容</p> <ol style="list-style-type: none"><li>R 语言基础<ul style="list-style-type: none"><li>R 语言环境配置与基本操作</li><li>R 语言的基本数据类型和数据结构</li><li>数据导入与导出</li></ul></li></ol>				

- 2. 数据管理
  - 数据清洗与预处理
  - 数据框架操作
  - 数据整理与转换
- 3. 数据可视化
  - 基础绘图：柱状图、散点图、线图等
  - 高级绘图：箱线图、热力图等
  - ggplot2 包的使用
- 4. 统计建模
  - 描述性统计与推断性统计
  - 回归分析（线性回归、逻辑回归）
  - 方差分析（ANOVA）
  - 假设检验（t 检验、卡方检验等）
- 5. 基础统计学知识
  - 概率与分布
  - 统计推断
  - 样本与总体
- 6. 语言数据分析案例
  - 文本数据分析
  - 实验数据分析
  - 问卷调查数据分析

#### 课程教学目标

1. 掌握 R 语言：通过学习 R 语言的基本操作和数据管理方法，学生能够熟练使用 R 进行数据处理和分析。
2. 提高数据分析能力：掌握数据可视化和统计建模的基本方法，能够针对具体的语言数据进行分析。
3. 理解统计学基础：通过学习基础统计学知识，理解统计方法的理论基础，并能在语言研究中正确应用这些方法。
4. 应用实际案例：通过具体案例的分析，学生能够将所学知识应用于实际语言数据的处理和分析，提升解决实际问题的能力。

#### 教学方法

1. 理论讲解：通过课堂讲授，帮助学生理解基本概念和理论知识。
2. 实践操作：通过计算机实验课和作业，强化学生对 R 语言和统计方法的实际操作能力。
3. 案例分析：通过真实语言数据的分析案例，培养学生将理论知识应用于实际问题的能力。
4. 小组讨论与报告：鼓励学生在小组讨论中交流学习心得，并通过报告形式展示学习成果。

#### 课程评估

	<ol style="list-style-type: none"> <li>1. 平时作业 (30%) : 通过布置平时作业, 考察学生对课堂知识的掌握情况和实践操作能力。</li> <li>2. 项目报告 (30%) : 学生选择一个实际语言数据分析项目, 进行全面的数据分析并提交报告。</li> <li>3. 期中考试 (20%) : 通过期中考试, 评估学生对基础统计学知识和 R 语言基本操作的理解。</li> <li>4. 期末考试 (20%) : 通过期末考试, 综合考察学生对整个课程内容的掌握程度。</li> </ol>
<p>*课程简介 (英文) (Description)</p>	<p>Course Nature</p> <p>This course is an elective for undergraduate students, primarily aimed at those studying linguistics, applied linguistics, and related fields. It is designed to equip students with the skills necessary to use statistical methods for analyzing linguistic data, laying a solid foundation for future research.</p> <p>Main Teaching Content</p> <ol style="list-style-type: none"> <li>1. Basics of R Programming <ul style="list-style-type: none"> <li>○ Setting up the R environment and basic operations</li> <li>○ Basic data types and structures in R</li> <li>○ Data import and export</li> </ul> </li> <li>2. Data Management <ul style="list-style-type: none"> <li>○ Data cleaning and preprocessing</li> <li>○ Data frame operations</li> <li>○ Data organization and transformation</li> </ul> </li> <li>3. Data Visualization <ul style="list-style-type: none"> <li>○ Basic plotting: bar charts, scatter plots, line graphs, etc.</li> <li>○ Advanced plotting: box plots, heat maps, etc.</li> <li>○ Using the ggplot2 package</li> </ul> </li> <li>4. Statistical Modeling <ul style="list-style-type: none"> <li>○ Descriptive and inferential statistics</li> <li>○ Regression analysis (linear regression, logistic regression)</li> <li>○ Analysis of variance (ANOVA)</li> <li>○ Hypothesis testing (t-tests, chi-square tests, etc.)</li> </ul> </li> <li>5. Basic Statistics Knowledge <ul style="list-style-type: none"> <li>○ Probability and distributions</li> <li>○ Statistical inference</li> <li>○ Samples and populations</li> </ul> </li> <li>6. Case Studies in Linguistic Data Analysis <ul style="list-style-type: none"> <li>○ Corpus data analysis</li> <li>○ Experimental data analysis</li> <li>○ Survey data analysis</li> </ul> </li> </ol> <p>Course Objectives</p>

	<ol style="list-style-type: none"> <li>1. Master R Programming: Enable students to proficiently use R for data processing and analysis through learning basic operations and data management methods.</li> <li>2. Enhance Data Analysis Skills: Equip students with the ability to use data visualization and statistical modeling techniques to analyze specific linguistic data.</li> <li>3. Understand Statistical Foundations: Provide students with a theoretical foundation in statistics, ensuring they can correctly apply these methods in linguistic research.</li> <li>4. Apply Real-World Cases: Allow students to apply their knowledge to real-world linguistic data, enhancing their ability to solve practical problems.</li> </ol> <p>Teaching Methods</p> <ol style="list-style-type: none"> <li>1. Lectures: Provide theoretical explanations to help students understand basic concepts and theoretical knowledge.</li> <li>2. Practical Exercises: Reinforce students' hands-on skills in R programming and statistical methods through computer lab sessions and assignments.</li> <li>3. Case Studies: Develop students' ability to apply theoretical knowledge to real-world problems through analysis of actual linguistic data cases.</li> <li>4. Group Discussions and Presentations: Encourage students to share their learning experiences in group discussions and present their findings through reports.</li> </ol> <p>Course Evaluation</p> <ol style="list-style-type: none"> <li>1. Assignments (30%): Assess students' understanding and practical skills in class knowledge through regular assignments.</li> <li>2. Project Report (30%): Students choose a real linguistic data analysis project, conduct comprehensive data analysis, and submit a report.</li> <li>3. Midterm Exam (20%): Evaluate students' understanding of basic statistical knowledge and R programming through a midterm exam.</li> <li>4. Final Exam (20%): Assess students' overall grasp of the course content through a final exam.</li> </ol>
--	--

**课程目标与内容 (Course objectives and contents)**

<p>*课程目标 (Course Object)</p>	<p>本课程“语言数据统计与分析”旨在实现以下三项主要目标：</p> <ol style="list-style-type: none"> <li>1. 知识掌握： <ul style="list-style-type: none"> <li>○ 统计与编程基础：学生应掌握 R 语言的基本操作和数据管理方法，能够进行数据导入、清洗、处理及可视化。</li> <li>○ 统计建模与分析：学生应具备使用 R 进行描述性统计、推断性统计、回归分析及方差分析的能力，并理解这些统计方法在语言数据分析中的应用。</li> </ul> </li> <li>2. 能力提升： <ul style="list-style-type: none"> <li>○ 数据分析能力：学生应具备对语言数据进行科学分析的能力，包括设计实验、收集数据、数据建模和结果解读，能够独立完成语言数据分析项</li> </ul> </li> </ol>
----------------------------------	---

	<p>目。</p> <ul style="list-style-type: none"> <li>问题解决能力：通过案例分析和实践操作，培养学生在面对复杂语言数据问题时的解决能力，提升其在实际应用中的灵活性和创新性。</li> </ul> <p>3. 素质与价值：</p> <ul style="list-style-type: none"> <li>科研素养：培养学生严谨的科研态度和批判性思维，鼓励其在语言学研究中保持开放的思维方式和探索精神。</li> <li>职业素养：提升学生在未来职业生涯中的竞争力，培养其在多学科团队中的合作能力和沟通技巧，适应现代社会对语言数据分析人才的需求。</li> </ul>						
	章节	教学内容（要点）	学时	教学形式	作业及考核要求	课程思政融入点	对应课程目标
*教学内容进度安排及对应课程目标 (Class Schedule & Requirements & Course Objectives)	示例：						
	第一章	数据管理	8	讲解及使用电脑进行实践操作	每节课之后完成相应的练习，共有4次练习	通过数据管理实践，培养学生一丝不苟、认真严谨的工作作风	1, 2, 3
	第二章	数据可视化	6	讲解及使用电脑进行实践操作	每节课之后完成相应的练习，共有3次练习	通过数据可视化实践，培养学生严谨、踏实的工作作风	1, 2, 3
	第三章	统计建模	18	通过大量语言研究案例展示如何对不同类型的语言数据进行统计建模	每节课之后完成相应的练习，共有6次练习	通过统计模型实践，培养学生严谨、科学的工作作风	1, 2, 3
注1：建议按照教学周周学时编排。							
注2：相应章节的课程思政融入点根据实际情况填写。							

*考核方式 (Grading)	<ol style="list-style-type: none"> <li>1. 平时作业 (30%) : 通过布置平时作业, 考察学生对课堂知识的掌握情况和实践操作能力。</li> <li>2. 项目报告 (30%) : 学生选择一个实际语言数据分析项目, 进行全面的数据分析并提交报告。</li> <li>3. 期中考试 (20%) : 通过期中考试, 评估学生对基础统计学知识和 R 语言基本操作的理解。</li> <li>4. 期末考试 (20%) : 通过期末考试, 综合考察学生对整个课程内容的掌握程度。</li> </ol>
*教材或参考资料 (Textbooks & Other Materials)	<ul style="list-style-type: none"> <li>○ 教材名称: R for Data Science, 作者: Hadley Wickham &amp; Garrett Golemund, 出版社: O'Reilly Media, 出版年份: 2016, 版次: 第 1 版, 书号: ISBN 978-1491910399</li> <li>○ 教材名称: An Introduction to Statistical Learning, 作者: Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani, 出版社: Springer, 出版年份: 2013, 版次: 第 1 版, 书号: ISBN 978-1461471370</li> <li>○ 教材名称: Statistics for Linguistics with R, 作者: Stefan Th. Gries, 出版社: De Gruyter Mouton, 出版年份: 2009, 版次: 第 1 版, 书号: ISBN 978-3110205644</li> <li>○ 教材名称: 基于 R 的语言研究多变量分析, 作者: 吴诗玉, 出版社: 上海交通大学出版社, 出版时间: 2024 年 03 月, 书号: ISBN 978-7313298676</li> <li>○ 教材名称: R 在语言科学研究中的应用, 作者: 吴诗玉, 出版社: 科学出版社, 出版时间: 2021 年 10 月, 装帧: 平装, 书号: ISBN 978-7030694119</li> <li>○ 教材名称: 第二语言加工及 R 语言应用(中青年学者外国语言文学学术前沿研究丛书), 作者: 吴诗玉, 出版社: 外语教学与研究出版社, 出版时间: 2019 年, 页数: 315, 书号: ISBN 978-7521311761</li> </ul>
其它 (More)	
备注 (Notes)	

备注说明:

1. 带\*内容为必填项。
2. 课程简介字数为 300-500 字; 课程大纲以表述清楚教学安排为宜, 字数不限。