

《多元统计分析与软件》课程教学大纲

一、课程基本信息

课程编码: MATH3009

课程名称: 多元统计分析与软件

英文名称: Multivariate Statistical Analysis and Software

课程学时: 总学时: 32 讲课学时: 24 实验学时: 0 上机学时: 8

课程学分: 2 学分

课程类别: 专业必修课

开课学期: 3 春

适用专业: 数据科学与大数据技术专业

先修课程: 概率论; 统计学; 高等代数与几何; 数学分析

负责人: 冯峥晖

二、课程目标

本课程是为数据科学与大数据技术专业本科生开设的专业课程,旨在教授分析多元数据的统计方法,特别是这些分析方法的数学原理和发展,为学生将来从事数据分析工作打下坚实的基础。课程主要内容包含高维均值与协方差阵及其统计推断、多元回归分析、主成分分析、因子分析、典型相关分析、分类分析、判别分析、聚类分析等。在本门课程中推荐 R 作为编程语言。

通过课堂讲授让学生具有坚实的理论基础,通过上机课训练学生运用 R 语言软件进行统计计算。并在课后的习题练习中掌握理论和实践技能。逐渐形成观察、思考、分析和解决实际问题的综合能力。

通过本课程的教学,应使学生熟练掌握求解各种常见的多元统计分析方法及其实现手段,熟练掌握相关实际应用软件,如 R 语言,将理论与实践相结合。

课程目标对学生的能力要求如下:

课程目标 1. 了解和掌握多元统计分析的基本概念;

课程目标 2. 掌握多元统计分析的基本方法,理论知识点和实现手段;

课程目标 3. 掌握使用 R 语言等统计软件实现多元统计分析在计算机上的实现,并能够用其解决实际问题。

三、课程目标与毕业要求指标点的对应关系

| 毕业要求 | 毕业要求指标点 | 课程目标 |
|-----------------|---|--------|
| 2 问题分析 4. 研究 | 2-1 了解和掌握多元统计中的基本概念 4-1 了解和掌握多元统计的统计相关背景及其应用 | 课程目标 1 |

| | | |
|-----------------|--|--------|
| 2 问题分析 4. 研究 | 2-2 掌握多元统计相关的基本方法 4-2 掌握多元统计的理论知识点 | 课程目标 2 |
| 2 问题分析 4. 研究 | 2-3 了解使用 R 语言软件实现多元统计算法。 4-3 了解将多元统计分析方法和统计软件编程应用于解决实际问题。 | 课程目标 3 |

四、课程目标与课程内容的对应关系

| 序号 | 教学内容 | 教学要求 | 学时 | 教学方式 | 对应 课程目标 |
|----|--------------|--|----|------|----------------------------|
| 1 | 矩阵运算 | 1. 引言 2. 准备知识介绍 | 2 | 授课 | 课程目标 1 课程目标 2 |
| 2 | 数据可视化与 R 语言 | 1. 数据可视化概述 2. R 语言简介 3. 多元统计数据的可视化 | 2 | 授课 | 课程目标 1 课程目标 2 |
| 3 | R 语言介绍 | 1. R 语言简介 2. R 语言绘图基础 | 2 | 上机 | 课程目标 1 课程目标 2 课程目标 3 |
| 4 | 多元正态分布 | 1. 随机向量 2. 多元正态分布 3. 矩阵多元正态分布 | 2 | 授课 | 课程目标 1 课程目标 2 |
| 5 | 多元正态总体抽样分布 | 1. Wishart 分布 2. Hotelling T^2 分布 | 2 | 授课 | 课程目标 1 课程目标 2 |
| 6 | 多元正态分布的参数估计 | 1. 多元正态分布的参数估计 2. 多元正态分布参数估计的性质 | 2 | 授课 | 课程目标 2 课程目标 3 |
| 7 | 方差分析 | 1. 单因素方差分析 2. 多因素方差分析 | 2 | 授课 | 课程目标 1 课程目标 2 |
| 8 | R 语言分析多元正态分布 | 1. R 语言分析多元正态分布 2. 方差分析 | 2 | 上机 | 课程目标 2 课程目标 3 |
| 9 | 主成分分析 | 1. 主成分分析的定义 2. 主成分的计算和性质 3. 样本主成分分析 4. 大样本性质和应用 | 4 | 授课 | 课程目标 1 课程目标 2 |
| 10 | 因子分析 | 1. 因子分析模型 2. 载荷矩阵的估计方法 3. 因子分析与主成分分析的 | 2 | 授课 | 课程目标 1 课程目标 2 |

| | | | | | |
|----|-------------------|---|---|----|------------------|
| | | 关系 | | | |
| 11 | 使用 R 语言实现多元统计分析 1 | 1. 使用 R 语言进行主成分分析 2. 使用 R 语言进行因子分析 | 2 | 上机 | 课程目标 2 课程目标 3 |
| 12 | 判别分析 | 3. 判别分析简介 4. 两个已知多元正态总体的判别 5. 多个总体的判别 | 2 | 授课 | 课程目标 1 课程目标 2 |
| 13 | 聚类分析 | 1. 聚类分析概述 2. K 均值聚类 3. 系统聚类法 | 4 | 授课 | 课程目标 1 课程目标 2 |
| 14 | 使用 R 语言实现多元统计分析 2 | 1. 使用 R 语言判别分析 2. 使用 R 语言聚类分析 | 2 | 上机 | 课程目标 2 课程目标 3 |

五、课程教学方法

1. 课堂授课

1.1 采用启发式教学，通过提问和互动激发学生主动学习的兴趣，培养学生独立思考、分析问题和解决问题的能力。以理论讲授为主，通过例题和实例的介绍加深学生对问题的理解。

1.2 采用 PPT 课件、网络视频等多种辅助方式增强教学的直观性，提高课堂教学信息量，同时帮助同学提高对数学发展史的理解认知。

2. 上机课

2.1 上机课讲授 R 语言的使用，学生通过上机课进行实践动手编程练习，学以致用，理论练习实际。

2.2 通过上机课习题加深学生 R 语言编程技能。

3. 课程 Project

通过课程 Project 让学生寻找实际问题，将所学理论和方法应用于实际的数值模拟问题，解决实际问题，锻炼学生综合运用知识的能力。

六、课程考核方法

课程考核以检验课程目标的达成度为手段，进而评价学生学习成果的达成度。考核的环节包括课后平时作业和期末报告，总成绩以百分计，满分 100 分，各考核环节所占分值比例及考核细则如下表。

| 考核环节 | 所占分值 | 考核与评价细则 | 对应课程目标 |
|------------|------|-----------------------------|-------------|
| 平时作业 | 20 | 每次按 100 分计算，取平均分按比例折算成实际得分 | 课程目标 1,2 |
| 出勤 | 10 | 不定时抽查出勤，出勤次数/总抽查次数按比例折算计入总分 | 课程目标 1,2 |
| 上机作业 | 20 | 每次按 100 分计算，取平均分按比例折算成实际得分 | 课程目标 3 |
| 课程 Project | 10 | 报告，满分 100 分，按比例折算成实际得分 | 课程目标 3 |
| 期末考试 | 40 | 期末闭卷考试，满分 100 分，按比例折算成实际得分 | 课程目标 1,2, 3 |

七、课程目标达成度评价方法

课程目标达成度评价包括课程分目标达成度评价和课程总目标达成度评价，具体计算方法如下表。

| 对应课程 | 考核环节 | 所占分 | 学生平均得分 | 达成度计算结果 |
|--------|------------|-----|-------------------|---------------------------------|
| 课程目标 1 | 平时作业 | 10 | A | $\frac{A+B+C}{25}$ |
| | 出勤 | 5 | B | |
| | 期末考试 | 10 | C | |
| 课程目标 2 | 平时作业 | 10 | D | $\frac{D+E+F}{35}$ |
| | 出勤 | 5 | E | |
| | 期末考试 | 20 | F | |
| 课程目标 3 | 上机作业 | 20 | G | $\frac{G+H+I}{40}$ |
| | 课程 Project | 10 | H | |
| | 期末考试 | 10 | I | |
| 课程总体 | 总评成绩 | 100 | A+B+C+D+E+F+G+H+I | $\frac{A+B+C+D+E+F+G+H+I}{100}$ |

主要教材(Main teaching materials):

李高荣, 吴密霞 (2021). 多元统计分析. 北京: 科学出版社.

参考文献(Reference):

1. 《多元统计分析》, 王静龙, 科学出版社, 2008
2. Johnson, Richard A., and Dean W. Wichern. Applied multivariate statistical analysis. (2020).
2. Gareth, James, et al. An introduction to statistical learning: with applications in R. Springer, 2013.