《应用多元统计分析》教学大纲

课程编号：151393B

课程类型：□通识教育必修课 □通识教育选修课

□专业必修课 ☑专业选修课

□学科基础课

总学时：48讲课学时：32 实验（上机）学时：16

学　　分：3

适用对象：金融学（数据与计量分析）

先修课程：高等代数、概率论、数理统计

一、教学目标

**目标1：使学生掌握经典的多元统计分析理论与方法。**

**目标2：使学生能够将多变量的实际问题应用于社会经济领域的问题研究。**

多元统计分析是数理统计学的一个重要分支，它的研究对象是多个随机变量组成的随机向量。本课程讲授经典的多元统计分析方法，对于统计学专业来说要求掌握各个方法的基本原理与算法，并且学会使用统计软件实现计算。使得学生在掌握多元分析基础理论和方法的同时，能够对多变量的实际问题进行数据处理分析应用。本课程在传授学生知识的同时，还注重培养学生社会主义价值观、高尚的职业道德、和崇高的爱国情怀。本课程为双语教学。

二、教学内容及其与毕业要求的对应关系

本课程使学生掌握经典的多元统计分析理论与方法，并能将其应用于社会经济领域的问题研究。首先向学生介绍多元统计分析的概况，然后复习矩阵代数知识，之后再复习随机向量的知识，在此基础上拓展相关的内容，这是本课程的基础，因为考虑到学生基础不同，可能没有做好充分准备，需要多花些时间讲解。接下来讲授将一元统计推广到多元统计的理论内容，主要讲解多元分布的基本概念和多元正态总体的统计推断，该部分内容是课程本身的基础，需要精讲。之后为多元数据的图表示法。最后讲述多元统计分析的具体方法部分，包括：聚类分析[Cluster analysis]、判别分析[Discriminant analysis]、主成分分析[Principal Components Analysis]、因子分析[Factor Analysis]、对应分析[Correspondence Analysis]和典型相关分析[Canonical correlation analysis]等。具体重点讲授的内容应根据当前发展方向和任课教师本身特点而定。同时应结合当前我国建设和经济发展实际介绍统计方法在我国实际中的应用，增强学生学好这门课，掌握多元统计分析知识的紧迫感和使命感。

为实现教学目标所采取的教学方法：以教师讲授为主，以学生自主学习为辅。除了课堂三多元分析理论与方法精讲之外，还有实验课配合，使用SPSS等统计软件计算实现。并且简要介绍我国在统计软件方面的进展及成果，增强学生的自信心与自豪感。

本课程为考查课，考核方式可采用开卷形式，一类是上机分步骤做多元正态总体均值向量的检验，聚类分析和判别分析等；另一类是应用问题——关于多变量的数据使用统计软件操作一些计算复杂的方法如因子分析等，对输出结果的解释分析。

三、各教学环节学时分配

**教学课时分配**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **序号** | **章节内容** | **讲课** | **上机** | **其他** | **合计** |
| 1 | 第一章 绪论 | 4 |  |  | 4 |
| 2 | 第二章 多元正态分布及其它三个重要分布 | 2 | 2 |  | 4 |
| 3 | 第三章 多元正态总体均值向量和协差阵的假设检验 | 4 | 2 |  | 6 |
| 4 | 第四章 多元数据的图表示 |  | 2 |  | 2 |
| 5 | 第五章 聚类分析 | 5 | 2 |  | 7 |
| 6 | 第六章 判别分析 | 5 | 2 |  | 7 |
| 7 | 第七章 主成分分析 | 3 | 2 |  | 5 |
| 8 | 第八章 因子分析 | 5 | 2 |  | 7 |
| 9 | 第九章 对应分析 | 2 | 1 |  | 3 |
| 10 | 第十章 典型相关分析 | 2 | 1 |  | 3 |
| **合计** |  | 32 | 16 |  | 48 |

四、教学内容

**第一章绪论**

§1、什么是多元分析

§2、多元统计分析的应用

§3、了解本课程的主要安排

§4、复习并补充相关的知识

§5、统计软件介绍

本章介绍多元统计分析的研究对象及其研究方法，了解多元分析可以解决什么问题。复习矩阵代数与随机向量的知识，并拓展相关内容。介绍统计分析在我国社会主义建设中发挥的重要作用，取得的重要成果，培养学生对掌握多元统计分析能力的紧迫感和对我国多元统计分析取得的成就的自豪感。另外还要对将要使用的统计软件进行介绍。

**复习思考题**：

1. 多元统计分析的研究对象是什么？其研究方法是怎样的？
2. 多元统计分析的理论与方法各包括哪些内容？

**第二章 多元正态分布**

§1、基本概念

§2、多元正态分布的定义及性质

§3、多元正态分布的参数估计

§4、其它三个重要分布

类比一元的正态分布函数将其推广至多元，了解多元正态分布的性质。掌握多元正态分布的参数估计，使用统计软件计算常用的统计量：样本均值向量、样本离差阵、样本协差阵和相关系数矩阵。了解其它三个重要分布：Wishart分布、Hotelling T2和WilksΛ分布。

**复习思考题**：

1. 多元正态分布的数字特征都包括什么？如何估计它们？
2. 三个重要分布：Wishart分布、Hotelling T2和WilksΛ分布分别是一元的什么分布的推广？各自的作用如何？

**第三章多元正态总体均值向量和协差阵的假设检验**

§1、单个多元正态总体的均值向量的检验

§2、两个多元正态总体的均值向量的比较

§3、多个多元正态总体的均值向量的检验（MONOVA）

通过本章教学使学生重点掌握多元正态总体的均值向量的假设检验，可以使用MONOVA解决实际问题中多个总体均值向量是否有显著性差异的研究。并要求学生会用SPSS等统计软件对数据进行计算分析。

**复习思考题**：

1. 单个多元正态总体、两个多元正态总体和多个多元正态总体均值向量的假设检验所使用的统计量分别是什么？

**第四章多元数据的图表示**

通过本章教学使学生掌握如何使用不同软件实现多元数据的图表示。由于不同软件各具特色，如Excel可画出蛛网图，SPSS不能画出脸谱图因此可使用R软件。以此为契机引导学生理解研发我国独立自主可控的统计软件的重要性。

**复习思考题**：

1. 多元数据的图表示有哪些？
2. 怎样画出蛛网图和脸谱图？如何解读？

**第五章 聚类分析**

§1、什么是聚类分析

§2、距离和相似系数

§3、八种系统聚类方法

§4、快速聚类与两步聚类

通过本章教学，使学生知晓聚类分析分Q型聚类和R型聚类分析。熟知常用的测度距离和相似系数的定义。重点掌握系统聚类方法，对于最长、最短距离法会分步骤计算以理解原理。使用软件能够进行快速聚类的计算，另外了解两步聚类。并要求学生会用SPSS等统计软件将聚类分析应用于分类问题的研究中。

**复习思考题**：

1. 简述欧氏距离与马氏距离的区别和联系。
2. 系统聚类的基本思想是什么？

**第六章 判别分析**

§1、什么是判别分析

§2、距离判别法

§3、费歇判别法

§4、Bayes判别法

§5、逐步判别法

通过本章教学，使学生清楚判别分析的目的和意义，了解判别分析与聚类分析的区别与联系。本章重点讲解距离判别法，Fisher判别法, Bayes判别法以及逐步判别法的原理。要求学生掌握两个总体的距离判别法，Fisher判别法的分步骤计算，并会用SPSS等统计软件对实际问题数据进行计算做判别分析。

**复习思考题**：

1. 聚类分析与判别分析的异同与联系是怎样的？
2. 距离判别法，Fisher判别法，Bayes判别法的基本思想分别是什么？

**第七章主成分分析**

§1、什么是主成分分析

§2、主成分分析的数学模型及几何解释

§3、主成分的推导及性质

§4、计算步骤及实例

通过本章教学使学生了解主成分分析的目的和意义，主成分分析的数学模型及几何解释，主成分的推导及基本性质。掌握计算程序中有关主成分分析的算法基础，会分步骤进行计算。并要求学生掌握用SPSS的Factor过程导出主成分的计算，或者使用R软件直接计算。介绍主成分分析在关系国计民生的重要任务比如核酸检测分析中的重要性，是学生对统计分析在我国的应用和成就有进一步的了解。

**复习思考题**：

1. 主成分分析的基本思想是什么？
2. 主成分分析的作用体现在何处？

**第八章因子分析**

§1、什么是因子分析

§2、因子分析的数学模型

§3、因子载荷阵的估计方法

§4、因子旋转

§5、因子得分

§6、主成分分析与因子分析的比较

§7、应用案例

通过本章教学使学生了解因子分析的目的和基本思想，重点掌握R型因子分析的数学模型，因子载荷阵的估计方法，因子旋转，因子得分。能够用SPSS等统计软件对实际问题进行计算分析。

**复习思考题**：

1. 试述主成分分析与因子分析的异同与联系。
2. 简述因子模型中因子载荷矩阵**A**的统计意义。

**第九章 对应分析**

§1、对应分析的基本思想

§2、对应分析方法的原理

§3、计算步骤及实例

通过本章的学习，使学生了解对应分析的目的和基本思想、方法和基本原理。掌握不同数据类型使用SPSS软件实现计算的技巧。

**复习思考题**：

1. 试述对应分析与因子分析的关联。
2. 试述对应分析的基本思想及其计算步骤。

**第十章 典型相关分析**

§1、典型相关分析的基本思想及数学描述

§2、典型相关系数和典型变量

§3、典型相关系数的显著性检验

§4、计算步骤及实例

通过本章的学习，使学生了解典型相关分析的目的和基本思想。掌握典型相关分析的数学模型，总体和样本的典型相关系数以及典型变量，典型相关系数的假设检验。要求学生能够通过程序编写使用SPSS软件实现典型相关分析的计算。

**复习思考题**：

1. 试述典型相关分析的基本思想。
2. 什么是典型变量？它具有哪些性质？

五、考核方式、成绩评定

本课程考核方式为考查，以平时成绩和期末考试相结合，平时成绩占40%，考试（闭卷）成绩占60%。

六、主要参考书及其他内容

参考教材：

[1]任雪松、于秀林编著：《多元统计分析》（第二版），中国统计出版社，2011年

[2](美)约翰逊(Johnson,R.A.),(美)威克恩(Wichern,D.W.) 著《实用多元统计分析(第6版)》(清华管理学系列英文版教材), 清华大学出版社，2008年。

**主要参考书：**

[1] 吴喜之编著，《统计学：从数据到结论》，北京，中国统计出版社，2004年。

[2]【美】Richard A. Johnson，Dean W. Wichem著，陆璇译：《实用多元统计分析》（第4版），清华大学出版社，2005年。

[3] 高祥宝 董寒青编著，《数据分析与SPSS应用》，北京，清华大学出版社，2007年。

[4] 高惠璇 编著 《应用多元统计分析》，北京大学出版社出版 2005年。

[5] 薛毅 编著 《统计建模与R软件》清华大学出版社，2007年。

[6] 贾俊平编著，《统计学》，北京，清华大学出版社，2009年。

[7] 其他网络中英文资料

执笔人：侯杰 教研室主任：　　　　　系教学主任审核签名：