**《R语言》考试大纲**

一、课程简介

《R语言》是现今最受欢迎的数据分析和可视化平台之一。它是统计领域广泛使用的S语言的一个分支，可以认为R是S语言的一种实现。而S语言是由AT&T贝尔实验室开发的一种用来进行数据探索、统计分析和作图的解释型语言。R语言是自由的开源软件，并同时提供 Windows、Mac OS X 和 Linux 系统的版本。本课程中，在介绍R语言的基本操作和基本数据处理方法之后，我们着重讲授通过R语言解决回归、方差分析、功效分析、重抽样与自助法、广义线性模型、主成分分析和因子分析、时间序列、聚类分析、以及分类的方法。这些内容都是统计学的重要组成部分。通过本课程的学习，能使学生掌握以R语言为工具处理各种统计工作中的实际问题的能力，并为一些后续课程的学习及统计学各领域的实践工作提供必要的保证，同时对于培养学生的计算机操作能力、编程能力有着重要的作用。

二、考查目标

考查学生对《R语言》理论基础知识掌握的情况及分析解决某些实际问题的能力。通过考试，选拔出具有较好统计学功底的学生来攻读我校统计学专业的硕士研究生，以使录取的研究生具有较扎实与系统的进一步学习统计学专业知识及从事有关统计科研工作所需的《R语言》能力。

三、考试内容及要求

第一章　回归

（一）考核知识点

1、拟合线性模型

2、评价模型适用性

3、解释模型的意义

（二）考核要求

1、深刻理解各项内容，熟练掌握线性模型的拟合及其评价方法。

2、能够应用本章的知识，通过R语言进行与回归问题相关的计算并解决相关的实际问题。

第二章　方差分析

（一）考核知识点

1、R中基本的实验设计建模

2、拟合并解释方差分析模型

3、检验模型假设

（二）考核要求

1、深刻理解各项内容，熟练掌握方差分析模型的拟合、解释及其检验的方法。

2、能够应用本章的知识，通过R语言进行与方差分析问题相关的计算并解决相关的实际问题。

第三章　功效分析

（一）考核知识点

1、判断所需样本量

2、计算效应值

3、评价统计功效

（二）考核要求

1、深刻理解各项内容，熟练掌握效应值的计算方法以及统计功效的评价方法。

2、能够应用本章的知识，通过R语言进行与功效分析问题相关的计算并解决相关的实际问题。

第四章　重抽样与自助法

（一）考核知识点

1、理解置换检验的逻辑

2、在线性模型中应用置换检验

3、利用自助法获得置信区间

（二）考核要求

1、深刻理解各项内容，熟练掌握置换检验以及自助法的使用方法。

2、能够应用本章的知识，通过R语言进行与重抽样及自助法问题相关的计算并解决相关的实际问题。

第五章　广义线性模型

（一）考核知识点

1、建立广义线性模型

2、预测类别型变量

3、计数型数据建模

（二）考核要求

1、深刻理解各项内容，熟练掌握广义线性模型的使用方法。

2、能够应用本章的知识，通过R语言进行与广义线性模型问题相关的计算并解决相关的实际问题。

第六章　主成分分析和因子分析

（一）考核知识点

1、主成分分析

2、探索性因子分析

3、理解其他潜变量模型

（二）考核要求

1、深刻理解各项内容，熟练掌握主成分分析和因子分析的使用方法。

2、能够应用本章的知识，通过R语言进行与主成分分析及因子分析问题相关的计算并解决相关的实际问题。

第七章　时间序列

（一）考核知识点

1、生成时间序列

2、分解时间序列

3、建立预测模型

4、预测未来值

（二）考核要求

1、深刻理解各项内容，熟练掌握时间序列的使用及以此预测未来值的方法。

2、能够应用本章的知识，通过R语言进行与时间序列问题相关的计算并解决相关的实际问题。

第八章　聚类分析

（一）考核知识点

1、找出可能的类

2、确定类的个数

3、获得类的嵌套层级

4、获得离散的类

（二）考核要求

1、深刻理解各项内容，熟练掌握聚类分析的使用方法。

2、能够应用本章的知识，通过R语言进行与聚类分析问题相关的计算并解决相关的实际问题。

第九章　分类

（一）考核知识点

1、用决策树进行分类

2、用随机森林进行集成分类

3、生成支持向量机模型

4、评价分类准确性

（二）考核要求

1、深刻理解各项内容，熟练掌握决策树和随机森林的使用以及分类的评价方法。

2、能够应用本章的知识，通过R语言进行与分类问题相关的计算并解决相关的实际问题。

第十章　处理缺失数据的高级方法

（一）考核知识点

1、识别缺失数据

2、缺失数据模式的可视化

3、完整案例分析

4、缺失数据的多重插补法

（二）考核要求

1、深刻理解各项内容，熟练掌握缺失数据的识别和处理的方法。

2、能够应用本章的知识，通过R语言进行与缺失数据问题相关的计算并解决相关的实际问题。

四、考试形式和试卷结构

1、试卷满分及考试时间

本试卷满分为100分，考试时间为180分钟

2、答题方式

答题方式为闭卷，笔试。

3、试卷内容结构

基本概念20分；应用80分。

4、试卷题型结构

题型包括单选题、填空题、计算题、证明题。

1. 参考书目

《统计建模与R软件》 薛毅 陈立萍著，清华大学出版社，2007年。