第3章 数据集基本处理

教案

**课程名称：**R语言编程基础

**课程类别：**必修

**适用专业：**

**总学时：**48学时（其中理论18学时，实验30学时）

**总学分：x**学分

**本章学时**：10学时

# 材料清单

* 1. 《R语言编程基础》教材
  2. 配套PPT
  3. 引导性提问
  4. 探究性问题
  5. 拓展性问题及其资料
  6. R语言编程相关软件环境

# 教学目标与基本要求

### 教学目标

主要介绍到R语言中的新增数据属性列的方法，访问数据框变量，创建和重命名变量。然后阐述R语言中清洗数据的方法，包括处理缺失值，处理日期变量等。阐述选取变量，删除变量等的方法。介绍整合数据，其中使用sql语句操作数据等。最后阐述处理字符数据的方法，包括正则表达式和字符处理函数。

### 基本要求

* 1. 掌握新增数据属性列的方法。
  2. 掌握清洗数据（包括处理缺失值，处理日期变量等）的方法。
  3. 掌握选取变量，删除变量等的方法。
  4. 掌握整合数据的方法。
  5. 掌握正则表达式与字符处理的方法。

# 问题

### 引导性提问

引导性提问需要教师根据教材内容和学生实际水平，提出问题，启发引导学生去解决问题，提问，从而达到理解、掌握知识，发展各种能力和提高思想觉悟的目的。

* 1. R语言中怎么新增数据属性列？
  2. ‘脏’数据包括哪些？
  3. 怎么选取变量及数据？
  4. 数据可以怎么整合？
  5. 如何处理字符数据？

### 探究性问题

探究性问题需要在教师深入钻研教材的基础上精心设计，提问的角度或者在引导性提问上，从重点、难点问题切入，进行插入式提问。或者是对引导式提问中尚未涉及但在课文中又是重要的问题加以设问。

1. R语言中可以怎么处理‘脏’数据？
2. R语言中整合数据的方法？
3. R语言中如何使用正则表达式处理字符数据？

### 拓展性问题

拓展性问题需要教师深刻理解教材的意义，学生的学习动态后，根据学生学习层次，提出切实可行的关乎实际的可操作问题。亦可以提供拓展资料供学生研习探讨，完成拓展性问题。

1. 为什么需要处理数据？
2. R语言中选取数据有什么便捷的方法？
3. R语言中字符数据的处理还有什么方法？

# 主要知识点、重点与难点

### 主要知识点

1. R语言新增数据属性列的方法
2. R语言清洗数据的方法
3. R语言选取变量及数据的方法
4. R语言整合数据的方法
5. R语言处理字符数据的方法

### 重点

1. R语言新增数据属性列的方法
2. R语言清洗数据的方法
3. R语言选取变量及数据的方法
4. R语言整合数据的方法

### 难点

1. R语言处理字符数据的方法

# 教学过程设计

### 理论教学过程

1. 访问数据框变量
2. 创建及重命名变量
3. 处理缺失值数据
4. 处理日期变量
5. 数据排序及合并数据集
6. 选取变量及删除变量
7. 使用subset函数选取数据
8. 随机抽样
9. 使用SQL语句操作数据
10. 汇总统计数据及重塑数据
11. 正则表达式
12. 字符处理函数

### 实验教学过程

1. 新增数据属性列操作，包括数据框变量等
2. 清洗数据操作，包括缺失值等
3. 选取变量及数据操作
4. 整合数据操作
5. 使用正则表达式及字符处理函数处理字符数据

# 教材与参考书

### 教材

林智章 张良均．R语言编程基础 [M]．北京：人民邮电出版社．2017．

### 参考资料

[1] 张良均、谢佳标、杨坦、肖刚．R语言与数据挖掘．北京：机械工业出版社．2016．