

第4章 数据管理

1. 为什么要进行数据管理?

【参考答案】在 R 语言中，数据分析只是数据准备的第一步。一个数据集中可能含有无数个变量和无数个观测，但其中只有一部分才是我们需要的数据，为了简化问题，我们只需要创建这一部分数据即可。在根据不同的数据分析而采取策略前，我们还要完成对数据集变量的修改或者不同数据集的组合、排序等数据管理的任务。

2. 为什么要进行数据类型的转换？各数据类型之间怎样用函数进行转换？请举例说明。

【参考答案】因为计算机对不同的数据类型有不同的保存方式，所以 R 语言提供了一系列用来判断某个对象的数据类型和将其转换为另一种数据类型的函数，各种数据类型之间也可以相互转换。转换数据类型：

- `as.numeric()` 转化为数值型
- `as.character()` 转化为字符型
- `as.data.frame()` 转化为数据框
- `as.matrix()` 转化为矩阵
- `as.vector()` 将矩阵向量化
- `as.factor()` 转化为因子
- `as.logical()` 转化为逻辑型

3. 简述 R 语言中数据集取子集的方法。

【参考答案】从一个数据集中选择我们想要访问的元素，并对变量进行操作。可通过提供行索引和列索引来提取子集。常见的方法有：选择变量、删除变量、选择行、`subset()` 函数变量和行选择、`sample()` 函数随机抽样、用 `sqldf()` 函数、使用 SQL 中的 `select` 语句进行查询等的数据集取子集方法。