实验四 列表与因子

**实验目的**

（一）掌握列表的创建与操作。

（二）掌握因子变量的定义。

（三）掌握与因子相关的R函数。

**实验内容**

实验一：R语言列表操作练习。创建脚本文件test0401.R，完成下面操作。

⑴ 使用list()函数创建一个包含以下元素的列表my.list1：

● 一个向量（vector）：包含5个整数，从1到5；

● 一个标量（scalar）：赋值为6.78

● 一个矩阵（matrix）：3行4列，填充数据由rpois(12, 20)产生；

● 一个数据框（data.frame）：含有2列，列名分别为Name和Age，Name的数据为"Alice", "Bob", "Charlie", "David", "Eve"；Age的数据为23, 27, 22, 31, 29。

⑵ 用[ ]运算符切片列表，获取my.list1中的第1、第3、第4个子对象，并将其赋值给一个新的变量vector\_slice。

⑶ 用$和[[ ]]运算符分别引用my.list1中的列表的标量和矩阵部分。

⑷ 提取my.list1中的data.frame子对象中Name为Bob的行。

**实验二**：R语言列表操作进阶练习。创建脚本文件test0402.R，完成下面操作。

⑴ 用list()函数创建如实验一所示的列表，并赋值给my.list2。

⑵ 给列表my.list2添加一个子对象char，值为"New Element"。

⑶ 删除列表my.list2中的子对象matrix。

⑷ 修改列表my.list中vector向量的最后一个元素，将其改为6。

⑸ 给列表my.list2中vector向量的第一个元素重命名，改为"First "。

⑹ 使用lapply()函数计算列表my.list2中的所有子对象的长度。

⑺ 使用sapply()函数计算列表my.list2中的所有子对象的长度。

**实验三**：因子变量的定义与操作。创建脚本文件test0403.R，完成下面操作。

⑴ 用data.frame函数定义一个数据框df，数据如下所示，并用参数stringsAsFactors将diabetes和status这两个变量设为因子变量，然后用str查看df的结构。

|  |  |  |  |
| --- | --- | --- | --- |
| patientID | age | diabetes | status |
| 1 | 25 | Type1 | Poor |
| 2 | 34 | Type2 | Improved |
| 3 | 28 | Type1 | Excellent |
| 4 | 52 | Type1 | Poor |

⑵ 用factor函数将数据框df中status列转换为有序因子变量，用str函数观察status的结构。

⑶ 用factor函数创建一个有序因子变量ord\_factor，变量的值为"low", "medium", "high", "medium", "low", "high"；因子的序为"low"、"medium"、"high"。

**实验四**：split、lapply和tapply函数的使用。创建脚本文件test0404.R，完成下面操作。下面操作使用鸢尾花（iris）数据集，该数据集包含了150个鸢尾花的四个属性（Sepal.Length, Sepal.Width, Petal.Length, Petal.Width）和一个分类变量Species。

⑴ 用split函数以Species为分组标准对数据集进行分组。

⑵ 计算每组数据的数据量。

⑶ 计算每组的花萼（sepal）和花瓣（petal）的平均长度。

**实验五**：tapply函数、by函数和aggregate函数的应用。创建脚本文件test0405.R，完成下面操作。

⑴ 用by函数以Species为分组标准对iris数据集分组计算花萼和花瓣的平均长度。

⑵ 用tapply函数对iris数据集按Species进行分组汇总，计算花萼和花瓣宽度的中位数。

⑶ 用aggregate函数对iris数据集进行分组汇总，计算花萼和花瓣的平均长度与平均宽度。

**实验六**：cut函数和table函数的应用。打开脚本文件test0406.R，完成下面操作。

⑴ 用cut函数将age变量分为三个年龄段（"Young", "Middle", "Old"），对应的年龄区间为(18, 35], (35, 55], (55, 70]，并赋值给变量age.group。

⑵ 用table函数统计每个age.group年龄段在数据集data出现的频数。

⑶ 统计每个年龄段不同收入水平的频数。