**成都信息工程大学计算机学院**

**数据统计与分析实验报告**

**“基于R语言的数据统计综合”实验**

|  |  |
| --- | --- |
| **姓 名：** |  |
| **学 号：** |  |
| **班 级：** |  |
| **完成日期：** |  |
| **任课教师：** |  |
| **总 分：** | **分** |

**实验四 基于R语言的数据统计综合实验**

**一、实验目的**

1. 综合运用R语言，实现数据的收集、处理和分析。

2. 掌握R语言绘图方法。

1. **实验内容**
2. 已知15位学生的体重（kg）：76.0, 65.0, 48.4, 67.9, 63.2, 63.2, 59.7, 64.5, 67.6, 65, 58, 70, 57.9, 51, 73。 （共33分，得 分）
3. 求学生的平均体重。

答：

|  |  |
| --- | --- |
| **题目要求** | **求学生的平均体重。** |
| 代码  （2分） | *data<-c(76.0, 65.0, 48.4, 67.9, 63.2, 63.2, 59.7, 64.5, 67.6, 65, 58, 70, 57.9, 51, 73)*  *mean(data)* |
| 运行结果  （2分） |  |

1. 求学生体重的方差和标准差。

答：

|  |  |
| --- | --- |
| **题目要求** | **求学生体重的方差和标准差。** |
| 代码  （4分） | *var(data)*  *sd(data)* |
| 运行结果  （4分） |  |

1. 将学生的体重从小到大排序。

答：

|  |  |
| --- | --- |
| **题目要求** | **学生的体重从小到大排序。** |
| 代码  （2分） | *data[order(data)]* |
| 运行结果  （2分） |  |

1. 求学生体重的中位数。

答：

|  |  |
| --- | --- |
| **题目要求** | **求学生体重的中位数。** |
| 代码  （2分） | *median(data)* |
| 运行结果  （2分） |  |

1. 求学生体重的分位数，分位点分别是0, 0.25, 0.5, 0.75和1。

答：

|  |  |
| --- | --- |
| **题目要求** | **求学生体重的分位数，分位点分别是0, 0.25, 0.5, 0.75和1。** |
| 代码  （2分） | *quantile(data)* |
| 运行结果  （2分） |  |

1. 画出茎叶图。

答：

|  |  |
| --- | --- |
| **题目要求** | **画出茎叶图。** |
| 代码  （2分） | *stem(data)* |
| 运行结果  （2分） |  |

1. 判断学生体重是否服从正态分布。

答：

|  |  |
| --- | --- |
| **题目要求** | **判断学生体重是否服从正态分布。** |
| 代码  （2分） | *qqnorm(data)* |
| 运行结果  （2分） |  |
| 结论  （1分） | 基本呈正态分布 |

1. 某种元件的寿命X(小时)，服从正态分布，，其中和均未知，16只元件的寿命如下： （共18分，得 分）159,280,101,212,224,379,179,264,222,362,168,250,149,260,485,170。

问是否有理由认为元件的平均寿命大于190小时（显著性水平α=0.05）。

答：在0.05显著性水平下，判断元件的平均寿命是否大于190小时。

|  |  |
| --- | --- |
| **题目要求** | **在0.05显著性水平下，判断元件的平均寿命是否大于190小时。** |
| 统计量  （3分） | 由于是对总体均值进行假设检验，且总体方差未知，故选择为统计量。 |
| 原假设  （2分） | H0: > |
| 备择假设  （2分） |  |
| 置信度  （2分） | 是 右 (填左或右)侧检验，置信度为0.05。 |
| 代码  （4分） | *x = c(159,280,101,212,224,379,179,264,222,362,168,250,149,260,485,170)*  *t.test(x, mu=190,alternative = 'greater')* |
| 运行结果  （3分） |  |
| 结论  （2分） | 拒绝H0 |

1. 随机抽取某大学16名在校大学生，了解到他们每月的生活费平均为800元，标准差S为300元，假定该大学学生的每月平均生活费服从正态分布，试以95%的置信度估计该大学学生的月平均生活费及其标准差的置信区间。（共34分，得 分）

答：在置信水平0.95下，月平均生活费的置信区间。

|  |  |
| --- | --- |
| 统计量  （3分） | 对均值进行区间估计，但总体方差未知，故选择 作为统计量。 |
| 代码  （6分） | datamean=800  datasd=300  y=qt(0.975,df=15)\*datasd/sqrt(16)  c(datamean-y, datamean+y) |
| 运行结果  （4分） |  |
| 置信区间  （4分） | [640.1413,959.8587] |

答：在置信水平0.95下，月生活费标准差 的置信区间。

|  |  |
| --- | --- |
| 统计量  （3分） | 对总体标准差进行假设检验，选择 作为统计量 |
| 代码  （6分） | sqrt(15)\*datasd/sqrt(qchisq(0.975,15))  sqrt(15)\*datasd/sqrt(qchisq(0.025,15)) |
| 运行结果  （4分） |  |
| 置信区间  （4分） | [221.6115, 464.3074] |

1. 假设某国的货币供给量y与国民收入x的历史数据如表1所示。（共15分，得 分）

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 年份 | y | x | 年份 | y | x |
| 1985 | 2 | 5 | 1991 | 4.2 | 8.4 |
| 1986 | 2.5 | 5.5 | 1992 | 4.6 | 9 |
| 1987 | 3.2 | 6 | 1993 | 4.8 | 9.7 |
| 1988 | 3.6 | 7 | 1994 | 5 | 10 |
| 1989 | 3.3 | 7.2 | 1995 | 5.2 | 11.2 |
| 1990 | 4 | 7.7 | 1996 | 5.8 | 12.4 |

1. 作出散点图，判断线性关系。

答：

|  |  |
| --- | --- |
| **题目要求** | **作x与y的散点图。并判断线性关系。** |
| 代码  （3分） | y<-c(2,2.5,3.2,3.6,3.3,4,4.2,4.6,4.8,5,5.2,5.8)  x<-c(5,5.5,6,7,7.2,7.7,8.4,9,9.7,10,11.2,12.4)  plot(x,y) |
| 散点图  （2分） |  |
| 结论  （2分） | 线性相关 |

1. 求货币供给量y对国民收入x的回归直线方程。

答：

|  |  |
| --- | --- |
| **题目要求** | **求y对x的经验回归直线方程。** |
| 代码  （2分） | lm(y~x) |
| 运行结果  （2分） |  |
| Y对X的回归直线方程  （2分） | y = 0.485193x + 0.009777 |
| 叠加散点图与回归直线  （2分） |  |

**四、实验总结** （不评分）

**总结内容要求：**不能少于300字，成果的总结或做题目心得和体会。

R语言是一种强大的统计分析语言，它可以帮助我们快速、准确地进行数据分析和统计分析,从而更好地了解数据的分布情况。它语法简单，第三方库极大地增强了它的功能。