R期末测试

一：（5/20分）

1. 产生一个1维15元素数组，分别取其5,3-5，（1,3,5,7）元素。

代码：

array1 <- *array(*1:15*)  
print(*array1*[*5*])  
print(*array1*[*3:5*])  
print(*array1*[c(*1, 3, 5, 7*)])*

结果：

[1] 5

[1] 3 4 5

[1] 1 3 5 7

1. 产生一个2维5 x 6元素数组，取其第四行元素。

代码：

array2 <- *array(*1:30, dim = *c(*5, 6*))  
print(*array2*[*4,*])*

结果：

[1] 4 9 14 19 24 29

> array2

[,1] [,2] [,3] [,4] [,5] [,6]

[1,] 1 6 11 16 21 26

[2,] 2 7 12 17 22 27

[3,] 3 8 13 18 23 28

[4,] 4 9 14 19 24 29

[5,] 5 10 15 20 25 30

1. 产生20个平均数为5，标准差为4的正态分布数据。求其平均值、最大值、最小值并进行大—小，小—大的排序输出。

程序：

data1 <- *rnorm(*20, 5, 4*)  
print(sort(c(mean(*data1*)*, *max(*data1*)*, *min(*data1*))*, decreasing = TRUE*))  
print(sort(c(mean(*data1*)*, *max(*data1*)*, *min(*data1*))))*

结果：

[1] 8.298098 3.957786 -3.978508

[1] -3.978508 3.957786 8.298098

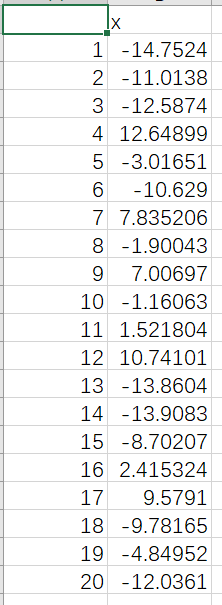
1. 生成20个[-15,15]均匀分布随机数保存至csv文件中。

程序：

data1 <- *runif(*20, -15, 15*)  
write.csv(*data1, "runif.csv"*)*

结果：

见runif.csv文件



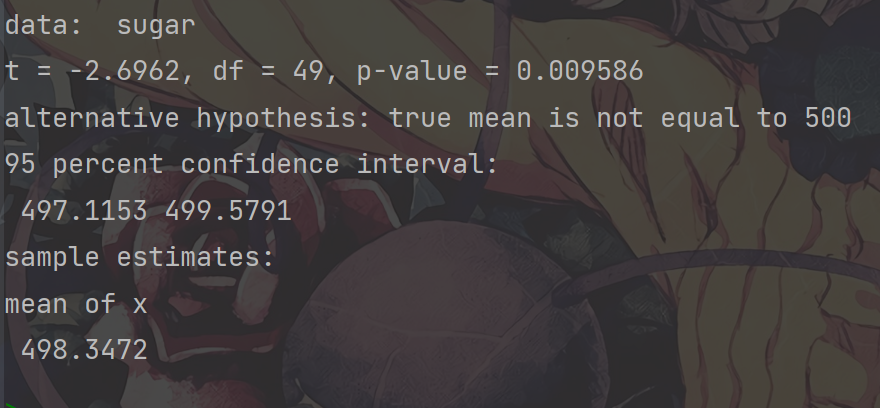
二、T检验（15分）

一个顾客买了一包标有500g重的一包红糖，觉得份量不足，于是找到监督部门；于是监督部门就去商店称了50包红糖（数据：sugar.txt）；其中均值（平均重量）是498.35g；这的确比500g少，问是否能够说明厂家生产的这批红糖平均起来不够份量呢？阐述分析结果。

程序：

sugar <- *read.csv(*"R\_learning/exam/data/sugar.txt", header = FALSE*)*result1 <- *t.test(*sugar, mu = 500*)  
print(*result1*)*

结果：



H0为mean=500，根据t测验的P值来看，需要否定H0：mean=500，即厂家生产的这批红糖平均起来是不够分量的

三、方差分析（15分）

用四种安眠药，两种不用的饮食方式，特选24只健康的兔子进行试验，睡眠时间如下所示：

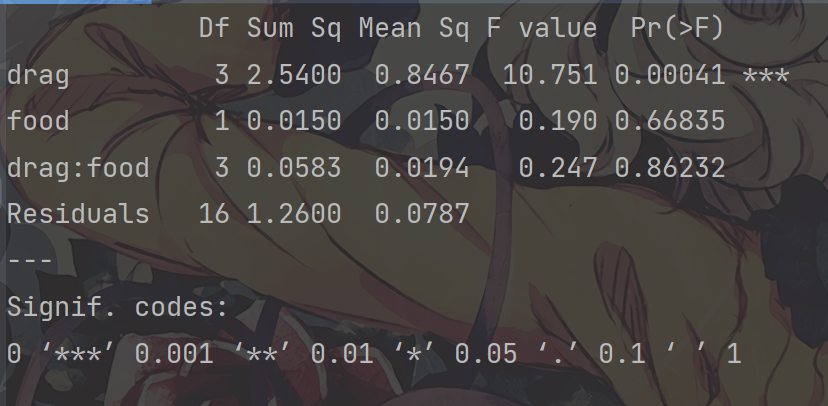
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 安眠药 | 睡眠时间/h | | | | | |
| 饮食B1 | | | 饮食B2 | | |
| A1 | 6.2 | 6.1 | 6.0 | 6.3 | 6.1 | 5.9 |
| A2 | 6.3 | 6.5 | 6.7 | 6.6 | 7.1 | 6.4 |
| A3 | 6.8 | 7.1 | 6.6 | 6.8 | 6.9 | 6.6 |
| A4 | 5.4 | 6.4 | 6.2 | 6.3 | 6.0 | 5.9 |

在显著水平α=0.05下对其进行双因素方差分析，分析4种安眠药、2种饮食的作用是否相同，阐述数据分析结果。

程序：

rabbit <- *data.frame(* sleep\_time = *c(*6.2, 6.1, 6.0, 6.3, 6.1, 5.9, 6.3, 6.5, 6.7, 6.6, 7.1, 6.4,  
 6.8, 7.1, 6.6, 6.8, 6.9, 6.6, 5.4, 6.4, 6.2, 6.3, 6.0, 5.9*)*,  
 drag = *gl(*4, 6, 24, labels = *c(*"A1", "A2", "A3", "A4"*))*, # 四组，一组六个  
 food = *gl(*2, 3, 24, labels = *c(*"B1", "B2"*)))* # 两组，一组三个  
rabbit\_aov <- *aov(*sleep\_time ~ drag + food + drag:food, data = rabbit*)  
summary(*rabbit\_aov*)*

结果：



4种安眠药、2种饮食的作用不相同，其中安眠药对兔子的睡眠时间有显著影响，而饮食以及安眠药和饮食的交互作用则对兔子的睡眠时间没有影响

五、数据整理（15分）

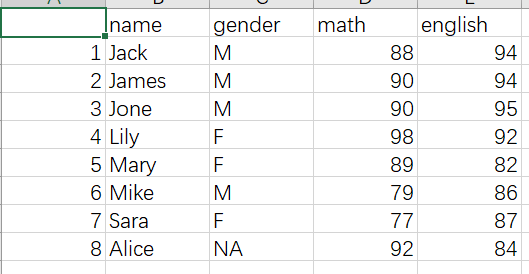
现在有某一班级的成绩表格grade.csv,（表格内有三列，分别为“ID”“english”和“math”）， 现在有本班级的信息表格 name.csv（表格内有两列，分别为 “name”和“gender”)。现在需要通过匹配两个表格获得该班级的新的成绩单，成绩单中能在name.csv中查找到的学生信息应该匹配，查找不到的用NA表示，并保存为grade2.csv。

程序：

grade <- *read.csv(*"R\_learning/exam/data/grade.csv"*)*name <- *read.csv(*"R\_learning/exam/data/name.csv"*)*new\_file <- *merge(*name, grade, by.x = "name", by.y = "ID", all.y = TRUE*)  
write.csv(*new\_file, "grade2.csv"*)*

结果：

见grade2.csv文件



四、画图（20分）

cars.txt数据是周一到周五car、truck、suv的销售总量，读入数据，要求画出car、truck以及suv周一到周五总的销售量的柱形图，细节要求如下：

主标题“Autos”（2分）；

纵坐标为“Total”（2分）；

横坐标的内容分别为“car”，“truck”和“suv”，颜色分别为1,2,3；（3分）

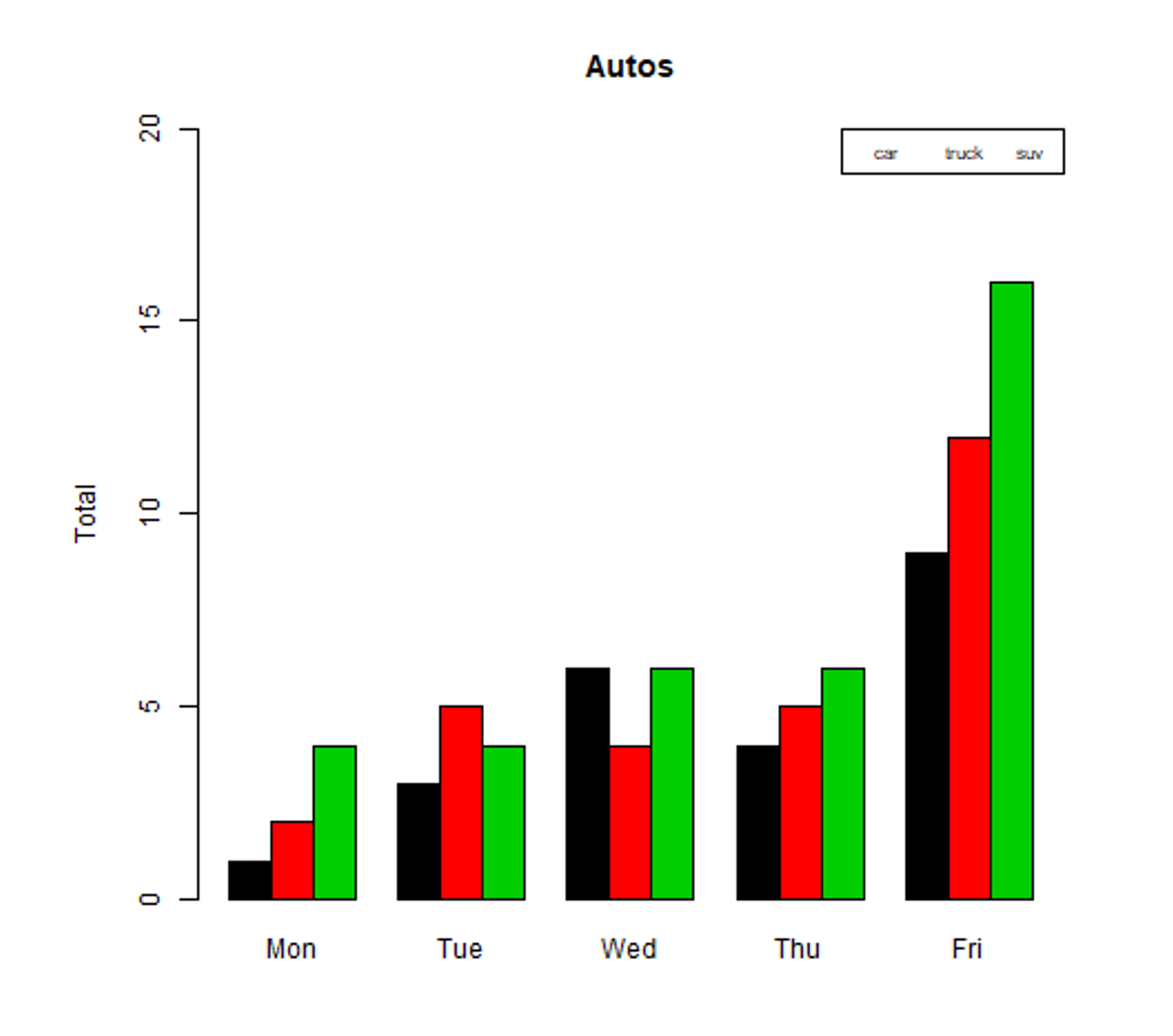
图例为"Mon","Tue","Wed","Thu","Fri"（4分），颜色分别为1,2,3,4,5，（3分），图列大小为0.7（2分）；

保存为“Auto.png”(4分)。

程序：

cars <- *read.table(*"R\_learning/exam/data/cars.txt", header = TRUE*)  
png(*"Auto.png"*)*cars\_data <- *t(as.matrix(*cars*))  
rownames(*cars\_data*)* <- NULL  
  
*barplot(*cars\_data,  
 names.arg = *c(*"Mon", "Tue", "Wed", "Thu", "Fri"*)*,  
 col = *c(*1, 2, 3*)*,  
 beside = TRUE,  
 ylim = *c(*0, 20*)*,  
 ylab = "Total",  
 main = "Autos"*)  
legend(*"topright",  
 *c(*"car", "truck", "suv"*)*,  
 col = *c(*1, 2, 3*)*,  
 cex = 0.7,  
 horiz = TRUE*)  
dev.off()*

结果：



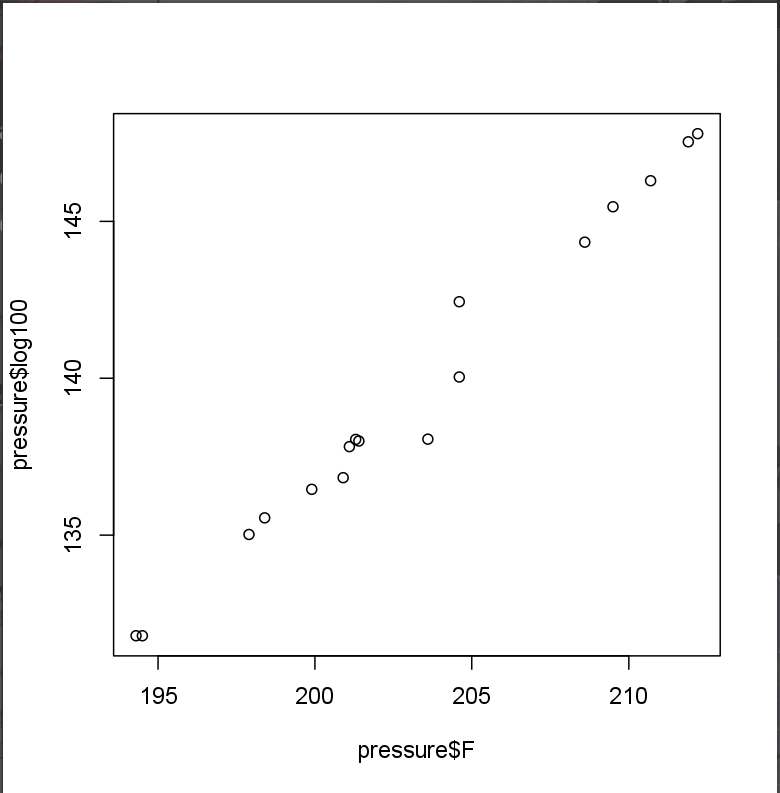
六、线性回归（15分）

物理学家James．D．Forbes试图通过水的沸点来估计海拔高度，他知道通过气压计测得的大气压可用于得到海拔高度，气压越低，高度越高，他测量了17个地方水的沸点（F）及大气压数据，并且对数据作了简单的处理，得到了较为明确的数学关系，所得数据见（大气压.csv）：

1. 画出表中以F为横坐标与log100为纵坐标的散点图，查看它们之间的关系（5分）。

程序：

pressure <- *read.csv(*"R\_learning/exam/data/pressure.csv"*)  
plot(*pressure$F, pressure$log100*)*

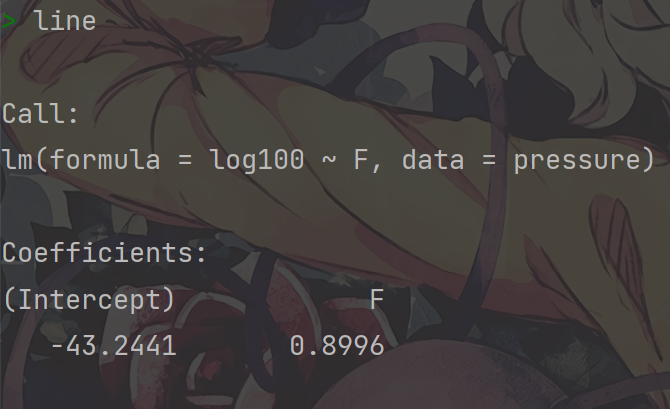


1. 以y=log100为因变量，x=F为自变量，计算x与y的相关系数，建立一元线性回归方程。（5分）。

程序：

line <- *lm(*log100 ~ F, data = pressure*)*

结果：



回归方程为：y=-43.2441+0.8996x

1. 在散点图上加上回归模型红线（2分），保存为“大气压.pdf”（3分）。

程序：

pressure <- *read.csv(*"R\_learning/exam/data/pressure.csv"*)  
pdf(*"pressure.pdf"*)  
plot(*pressure$F, pressure$log100*)*line <- *lm(*log100 ~ F, data = pressure*)  
abline(*line$coefficients,col = "red"*)  
dev.off()*

结果：

