辅导课程N° 3

描述性统计

练习 0. 更改工作目录并创建一个名为 TP3.R 的脚本,您将使用该脚本保存此会话的所有结果。

1. 数量变量

在本节中,您将对文件 cats.txt 中包含的变量执行描述性分析。该文件包含有关猫的性别、体重（以千克为单位)和心脏重量（以克为单位)的数据。

下载文件 cats.txt 并导入数据集。要了解数据结构,您可以使用以下命令显示变量（列)的名称

> 名字（猫）

和观察数（行数）

> nrow(猫)

当表非常大时,最好只显示数据帧的某些行。例如,使用命令head(nom.data.frame,n)仅显示前 n 行。

练习 1. 熟悉数据集。

(1) 变量名是什么?

(2) 数据集包含多少个变量和观测值?

(3) 显示前 10 个观测值。猫的心脏的性别和重量是多少
6 号?

Attach ()函数允许使用表中变量的名称,而无需回忆数据框的名称,即只需输入 Bwt 即可代替 cats$Bwt。第一次尝试：

> 顺便说一句

接下来,执行命令：

> 附加（猫）

最后,重试：

> 顺便说一句

练习 2. 描述性统计

(1) 下表显示了R中计算集中1趋势的描述性统计以及离散度统计的函数。

---

1. 警告:使用 R 中的var()函数计算的样本方差返回 s 给出的无偏方差
$^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - x)^2$ （见讲座）。不要与有偏方差 ^σ 混淆 $^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - x)^2$。

统计	函数平均值
样本平均值	（变量名)分位数（变量名,0.5）
样本中位数	var（变量名)sd（变量名)分位数（变量名，
样本方差	p)diff（范围（变量名））
样本标准差	
p 阶分位数	
范围	
四分位数范围	IQR（变量名）

计算变量 Bwt 的样本均值;中位数、1、2 和 3 阶四分位数;样本方差和标准差、极差和四分位距。对结果发表评论。

(2) 命令summary()返回数据帧的描述性统计信息。执行
命令

> 摘要(cats) 并将获得的

结果与上一个问题的结果进行比较。

练习 3.箱线图

(1) 箱线图是描述数据集主要特征（中位数、四分位数、最小值、最大值和离群值)的图形表示。绘制变量箱线图的 R 函数是
boxplot(nom_variable)。

绘制代表猫体重和心脏重量的变量的箱线图。对获得的结果进行评论并将其与使用summary() 获得的结果进行比较。

练习 4. 频率直方图

hist ()函数显示数据集中变量的直方图。默认情况下， hist()表示直方图中的频率,即每个类的观测值数量。使用选项freq=FALSE绘制相对频率（以便直方图的总面积为 1）。函数hist()的附加选项允许：

定义类的数量,使用选项Breaks=n我们获得 n+1 的直方图
类。

定义构建直方图的间隔。使用Breaks=vec on 可以获得一个直方图,其中间隔（类别)的限制由向量 vec 的值给出， - 更改颜色:例如 col= blue

(1) 绘制与猫的体重变量相关的直方图。尝试不同的类数值:2、20、200 和 2000。您观察到什么?最好的班级数量是多少?

(2) 命令hist()返回一个列表类型的对象,允许您查找与直方图关联的频率表的元素。例如,如果hist()返回的对象是 histo,则命令
histo$breaks允许查找类的间隔， histo$counts返回每个间隔的频率。绘制与 4 个类别的直方图相关的频率表。

2. 定性变量

定性变量（也称为分类变量或因子)是其取值为类别的变量,可以是序数变量或名义变量。这些类别也称为级别。例如,变量 Sex：

> 班级（性别）

[1]“因素”

函数levels()用于找出定性变量所采用的值集。

> 级别（性别）

[1]“F”“M”

为了以图形方式分析定性变量,我们可以使用函数bar-plot()绘制条形图。该函数的参数是一个带有条形高度的向量。后者可以从函数table()中获得,该函数返回定性变量的频率表。

例如,输入

> 表（性别）

练习 5.

(1)使用函数table()计算各类别的相对频率
可变性别。

(2) 使用barplot()函数显示变量 Sex 的条形图。

(3) 还可以使用函数pie()绘制饼图。绘制饼图。哪个
您更喜欢代表吗?

定性变量对于按组绘制箱线图很有用。更准确地说,可以使用以下命令根据变量 var.factor 的值按组构建变量 var.num 的箱线图：

> 箱线图(var.num ~ var.factor)

请注意,命令plot(var.num ~ var.facteur) 与上一个命令等效。

练习 6. * 绘制公猫和母猫体重的箱线图并解释结果。

3.两个定量变量之间关系的研究

为了研究两个定量变量之间的关系,我们可以显示散点图。

练习 7.

(1) 使用plot()函数显示变量(Bwt,Hwt)的散点图。关于这两个变量之间的关系你能说些什么?

(2) 使用以下公式计算这些变量的样本协方差和相关系数
函数cov()和cor()。对结果发表评论。

(3) 为了可视化性别对变量 Bwt 和 Hwt 的影响,请使用不同颜色为男性和女性绘制点。评论一下剧情。