

应用多元统计分析作业 (1)

钟瑜 222018314210044

2020 年 9 月 21 日

1. 利用 R 中自带的 iris 数据分别计算三种鸢尾花的 sepal 长度和宽度的均值和标准差, petal 的长度和宽度的均值和标准差;

1.1 不分花种

```
1 > print(head(iris))
2       Sepal.Length Sepal.Width Petal.Length Petal.Width Species
3 1           5.1         3.5         1.4         0.2   setosa
4 2           4.9         3.0         1.4         0.2   setosa
5 3           4.7         3.2         1.3         0.2   setosa
6 4           4.6         3.1         1.5         0.2   setosa
7 5           5.0         3.6         1.4         0.2   setosa
8 6           5.4         3.9         1.7         0.4   setosa
9
10 > library(psych)
11 > describe(iris)
12 vars          n mean sd median trimmed mad min max range skew kurtosis se
13 Sepal.Length  1 150 5.84 0.83   5.80   5.81 1.04 4.3 7.9   3.6   0.31  -0.61 0.07
14 Sepal.Width   2 150 3.06 0.44   3.00   3.04 0.44 2.0 4.4   2.4   0.31   0.14 0.04
15 Petal.Length  3 150 3.76 1.77   4.35   3.76 1.85 1.0 6.9   5.9  -0.27  -1.42 0.14
16 Petal.Width   4 150 1.20 0.76   1.30   1.18 1.04 0.1 2.5   2.4  -0.10  -1.36 0.06
17 Species*      5 150 2.00 0.82   2.00   2.00 1.48 1.0 3.0   2.0   0.00  -1.52 0.07
```

1.2 分花种

```
1 > vars<- c("Sepal.Length", "Sepal.Width", "Petal.Length", "Petal.Width")
2 > aggregate(iris[vars], by=list(Species=iris$Species), mean)
3 Species      Sepal.Length Sepal.Width Petal.Length Petal.Width
4 1      setosa      5.006      3.428      1.462      0.246
5 2 versicolor      5.936      2.770      4.260      1.326
6 3  virginica      6.588      2.974      5.552      2.026
7
8 > aggregate(iris[vars], by=list(Species=iris$Species), sd)
9 Species      Sepal.Length Sepal.Width Petal.Length Petal.Width
10 1      setosa      0.3524897  0.3790644  0.1736640  0.1053856
```

```

11 2 versicolor    0.5161711    0.3137983    0.4699110    0.1977527
12 3 virginica     0.6358796    0.3224966    0.5518947    0.2746501

```

2. 检验三种鸢尾花的 sepal 长度和宽度, petal 的长度和宽度是否存在差异;

2.1 散点图的矩阵绘制 (1)

```

1 library(GGally)
2 ggscatmat(data=iris, columns=1:4, color="Species")

```

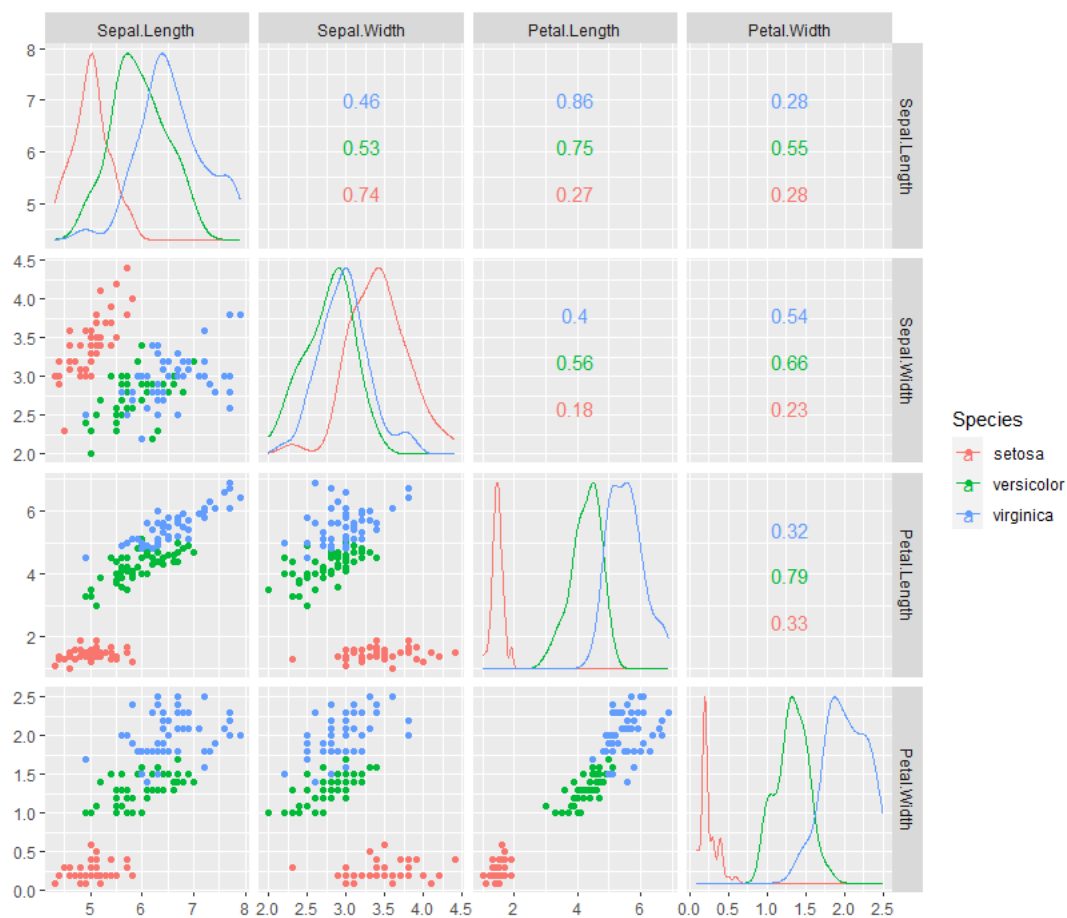


图 1: 散点图矩阵

2.1 散点图的矩阵绘制 (2)

```

1 mycolors<-c("#FF95CA", "#FFAF60", "#95CACA")
2 ggpairs(data = iris, aes(color=Species),
3 columns = c("Sepal.Length", "Sepal.Width", "Petal.Length",
4             "Petal.Width", "Species"))+
5 scale_fill_manual(values=alpha(mycolors, 0.5))

```

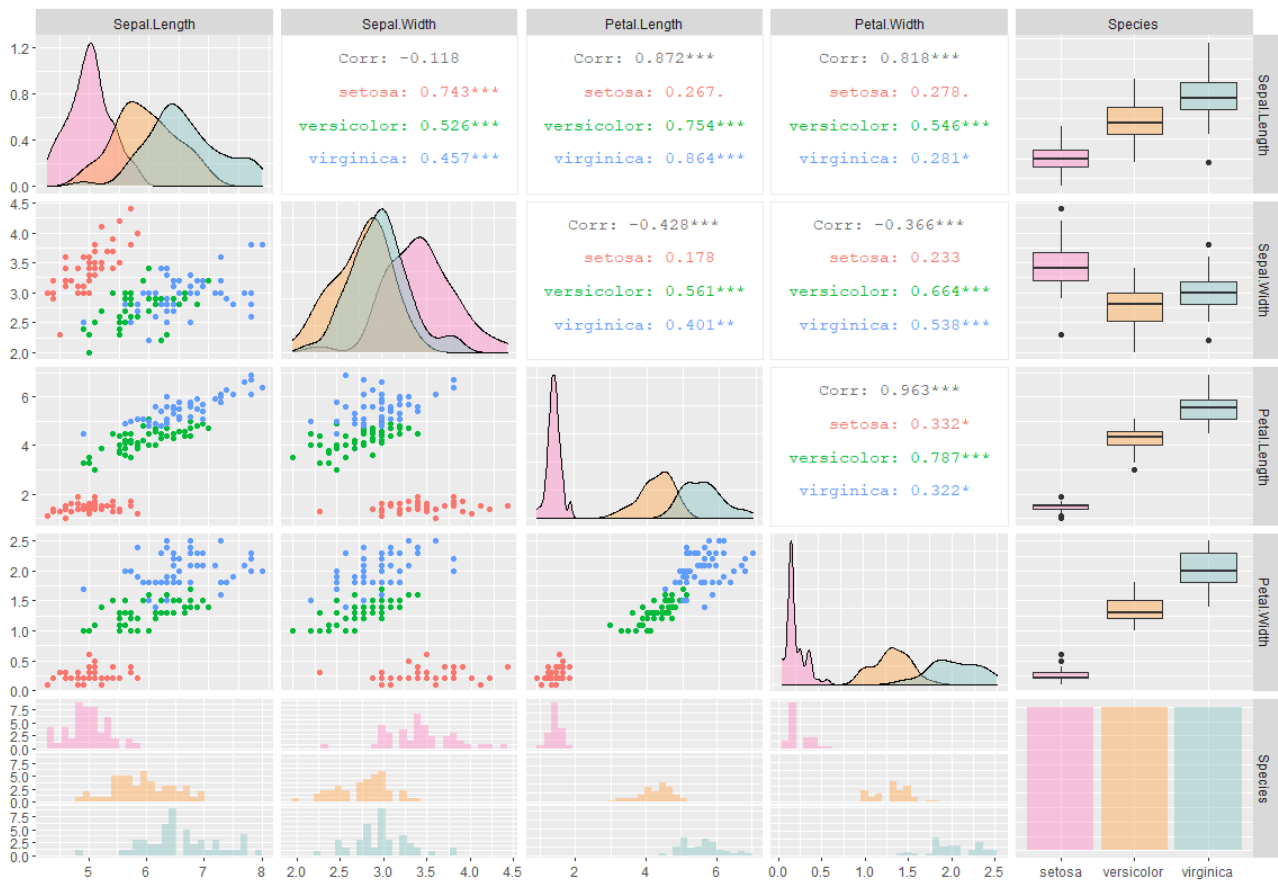


图 2: 散点图矩阵

3. 计算每朵花的 sepal 长度和宽度的差, petal 的长度和宽度的差, 并且将计算结果保存到原始数据中, 并分别命名为 dif_sepal, dif_petal;

```

1 dif_Sepal<-iris$Sepal.Length-iris$Sepal.Width
2 as.matrix(dif_Sepal)
3 as.data.frame(dif_Sepal)
4 myiris1<-cbind(iris ,dif_Sepal)
5
6 dif_Petal<-iris$Petal.Length-iris$Petal.Width
7 as.matrix(dif_Petal)
8 as.data.frame(dif_Petal)
9 myiris<-cbind(myiris1 ,dif_Petal)

```

4. 将第三步中修改后的数据另外保存到 G 盘, 并命名为 myiris.csv;

```

1 write.table(myiris, "G:/myiris.csv",
2             row.names=FALSE, col.names=TRUE, sep=", ")

```

5. 用 R 重新读取 myiris.csv。

```
1 > library(readr)
2 > cancer <- read_csv("G:/myiris.csv")
3 Parsed with column specification:
4 cols(
5   Sepal.Length = col_double(),
6   Sepal.Width = col_double(),
7   Petal.Length = col_double(),
8   Petal.Width = col_double(),
9   Species = col_character(),
10  dif_Sepal = col_double(),
11  dif_Petal = col_double()
12 )
13 > View(cancer)
```