2022fall Biostatistics Final

191850112 Qiang Liu (刘强)

2022-12-13

目录

Question 1

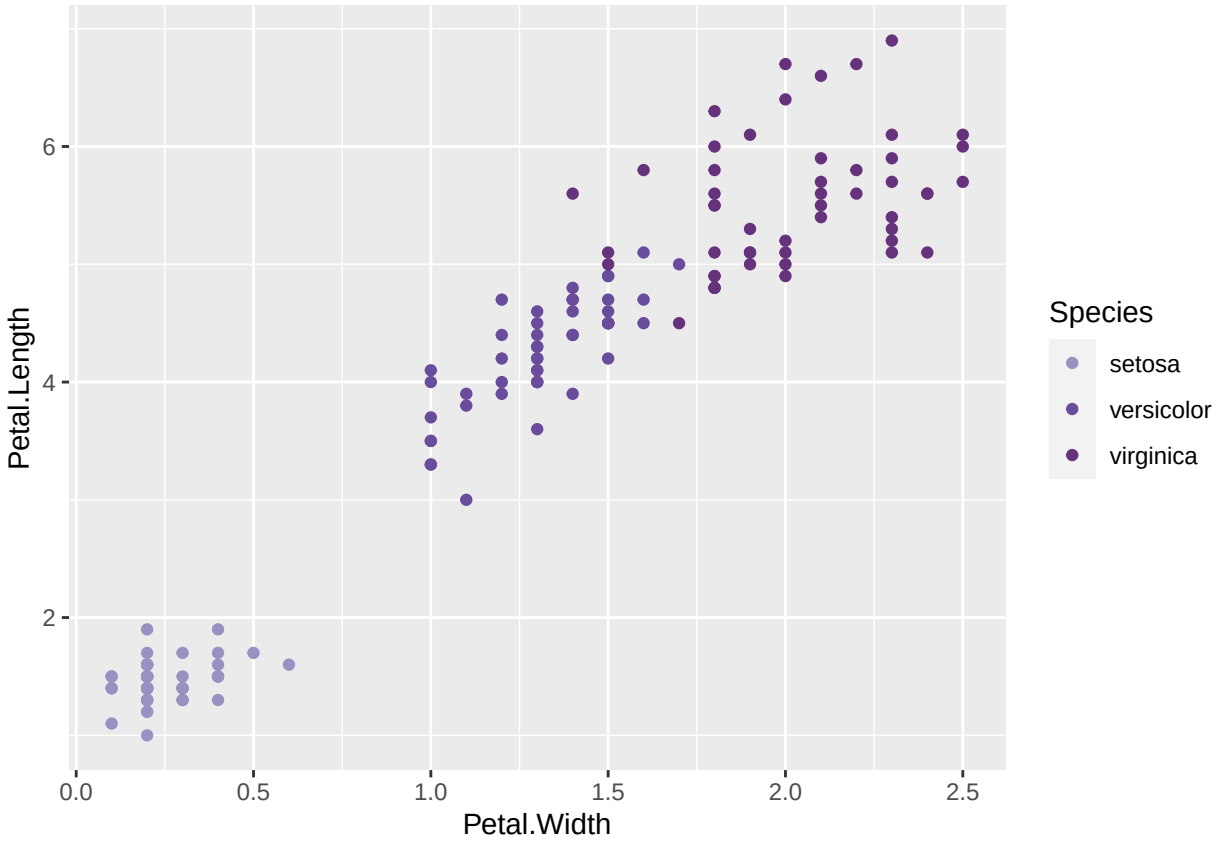
```
> library(ggplot2)
> library(dplyr)
> data("iris") # 内置数据集 iris
> head(iris) # 查看前六行数据
```

```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1          5.1         3.5          1.4         0.2  setosa
## 2          4.9         3.0          1.4         0.2  setosa
## 3          4.7         3.2          1.3         0.2  setosa
## 4          4.6         3.1          1.5         0.2  setosa
## 5          5.0         3.6          1.4         0.2  setosa
## 6          5.4         3.9          1.7         0.4  setosa
```

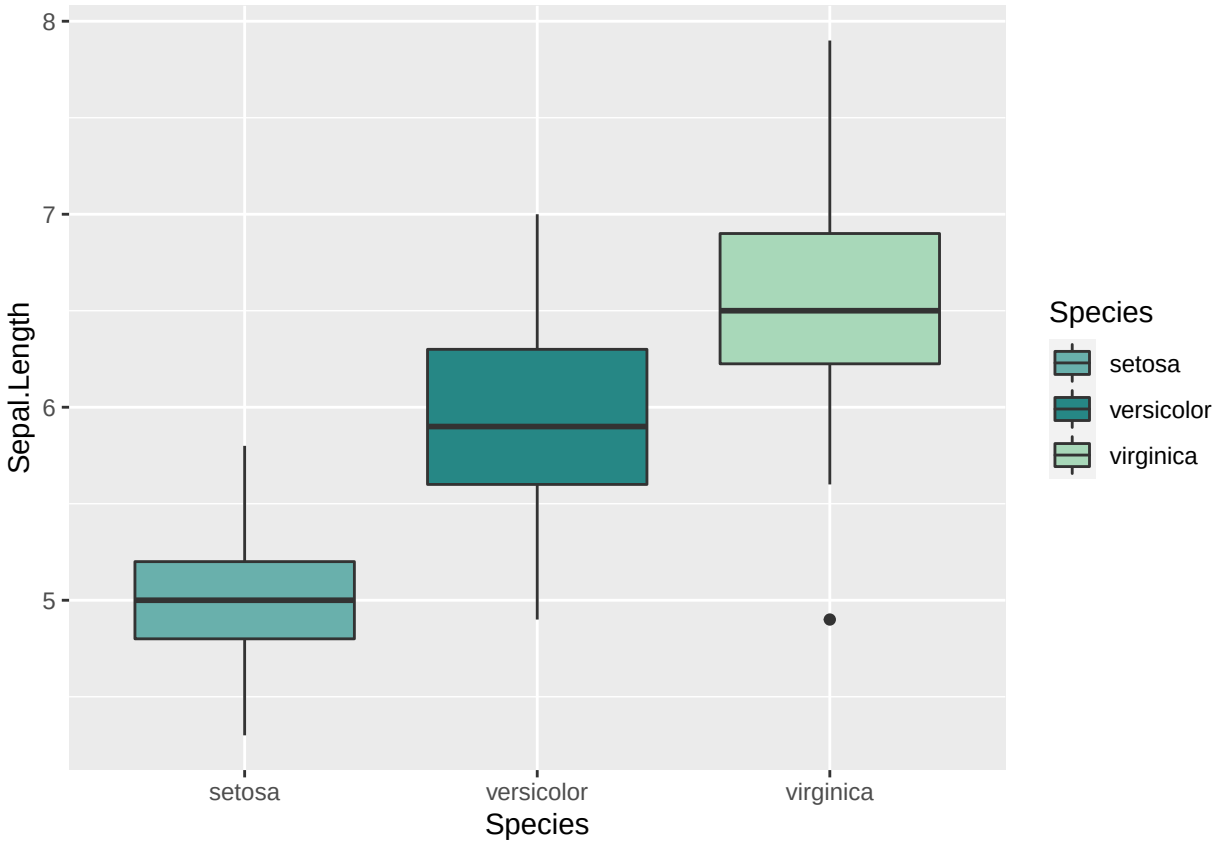
```
> linear_model <- lm(Petal.Length ~ Petal.Width,data = iris)
> linear_model$coefficients # 可知两种指标的相似度为 2.229940
```

```
## (Intercept) Petal.Width
##    1.083558    2.229940
```

```
> colorset1 <- c("#9B90C2","#6A4C9C","#66327C")
> colorset2 <- c("#69B0AC","#268785","#A8D8B9")
> iris %>% # 绘制散点图，数据点颜色指代不同的 Species
+   ggplot(aes(x = Petal.Width,y = Petal.Length,color = Species))+
+   geom_point()+
+   scale_color_manual(values = colorset1)
```




```
> iris %>% # 绘制箱线图，分析 Sepal.Length 特征
+   ggplot(aes(x = Species,y = Sepal.Length,fill = Species))+
+   geom_boxplot()+
+   scale_fill_manual(values = colorset2)
```



Question 1 answer description

(1)Petal.Length 和 Petal.Width 两种指标的相似度 (Coefficient) 约为 2.23, 散点图如上所示.

(2) 由箱线图可知, 三个品种之间的 Sepal.Length 特征存在明显差异, 其中 setosa 的总体分布最低,virginica 的总体分布最高.

Question 2

```
> n <- 20
> u1 <- 36.75
> s1 <- 2.77
> u2 <- 40.35
> s2 <- 1.56
> F12 <- s1^2/s2^2 # 用 F-test 进行方差齐性分析
> F12 #F1 的值为 3.152901, 大于 F(19,19,0.99)
```

```
## [1] 3.152901
```

```
> t <- (u1-u2)/((s1^2/n)+(s2^2/n))^0.5
> t
```

```
## [1] -5.064273
```

```
> k <- (s1^2/n)/((s1^2/n)+s2^2/n)
> df <- ((k^2/n)+((1-k)^2/n))^-1
> df # 对应 a=0.01 的 t 值为 2.45, 大于 t
```

```
## [1] 31.52715
```

Question 2 answer description

首先验证方差齐性，计算出 F 值为 3.15>F0.99(19,19)，因此两个品系的差异极显著

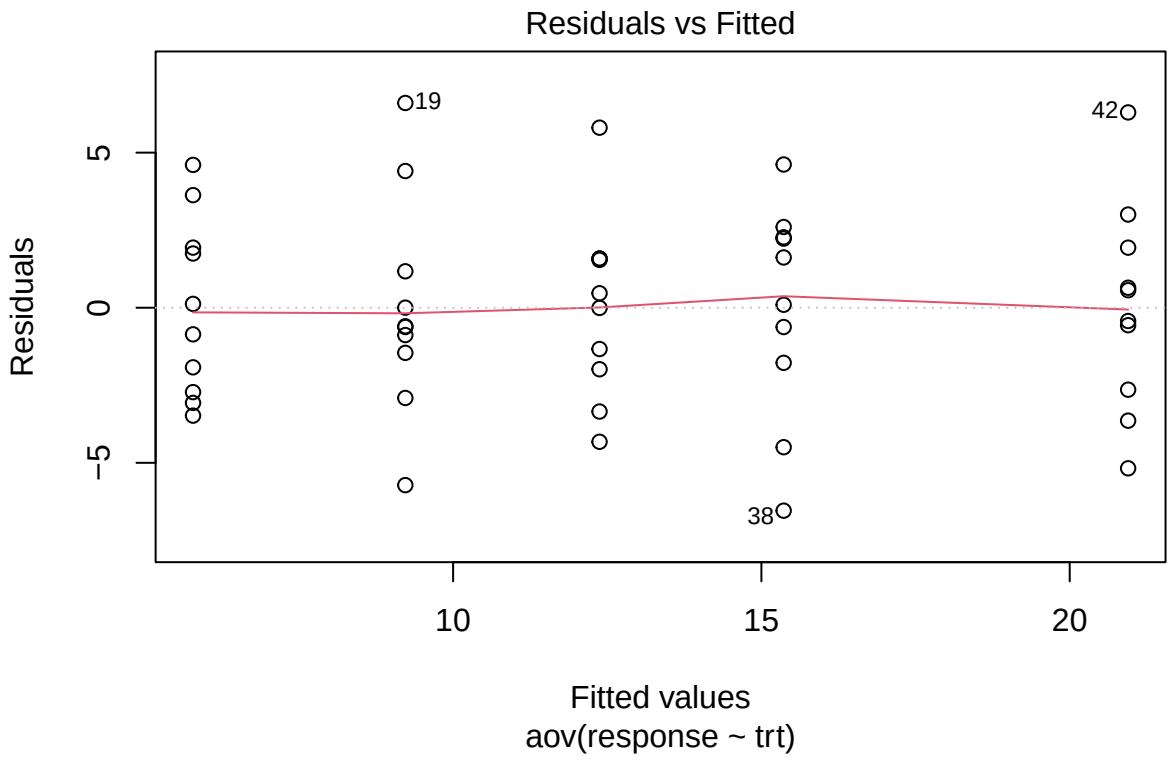
在进行均值检验. 由于方差不等，应该使用近似 t 检验，由于计算得到的 t 值为-5.06，而计算出的自由度对应 t 分布的值为 2.45,t<t(1-a,df)，拒绝原假设，因此可知新品系的均值大于原品系，值得推广.

Question 3

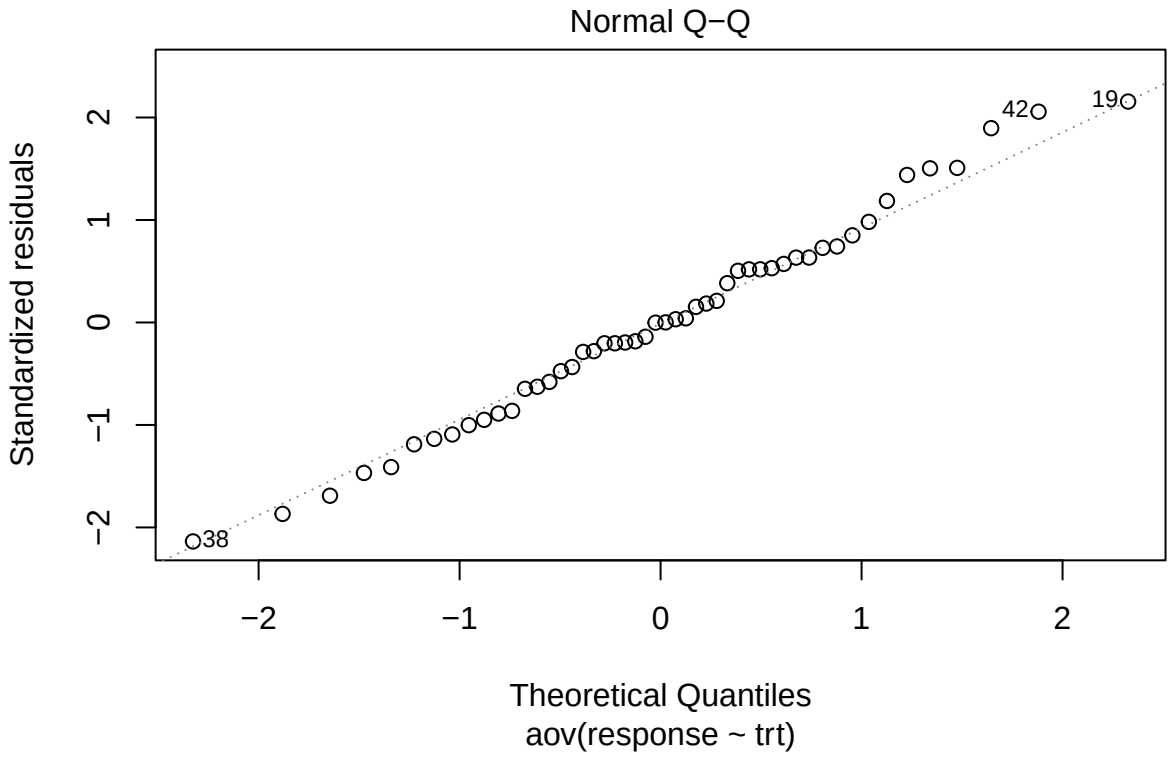
```
> library(multcomp)
> data(cholesterol) # 内置数据集
> head(cholesterol) # 查看前六行
```

```
##     trt response
## 1 1time   3.8612
## 2 1time  10.3868
## 3 1time   5.9059
## 4 1time   3.0609
## 5 1time   7.7204
## 6 1time   2.7139
```

```
> res.aov <- aov(response~trt,data = cholesterol) #ANOVA 分析
> plot(res.aov,1) # 方差齐性检验
```



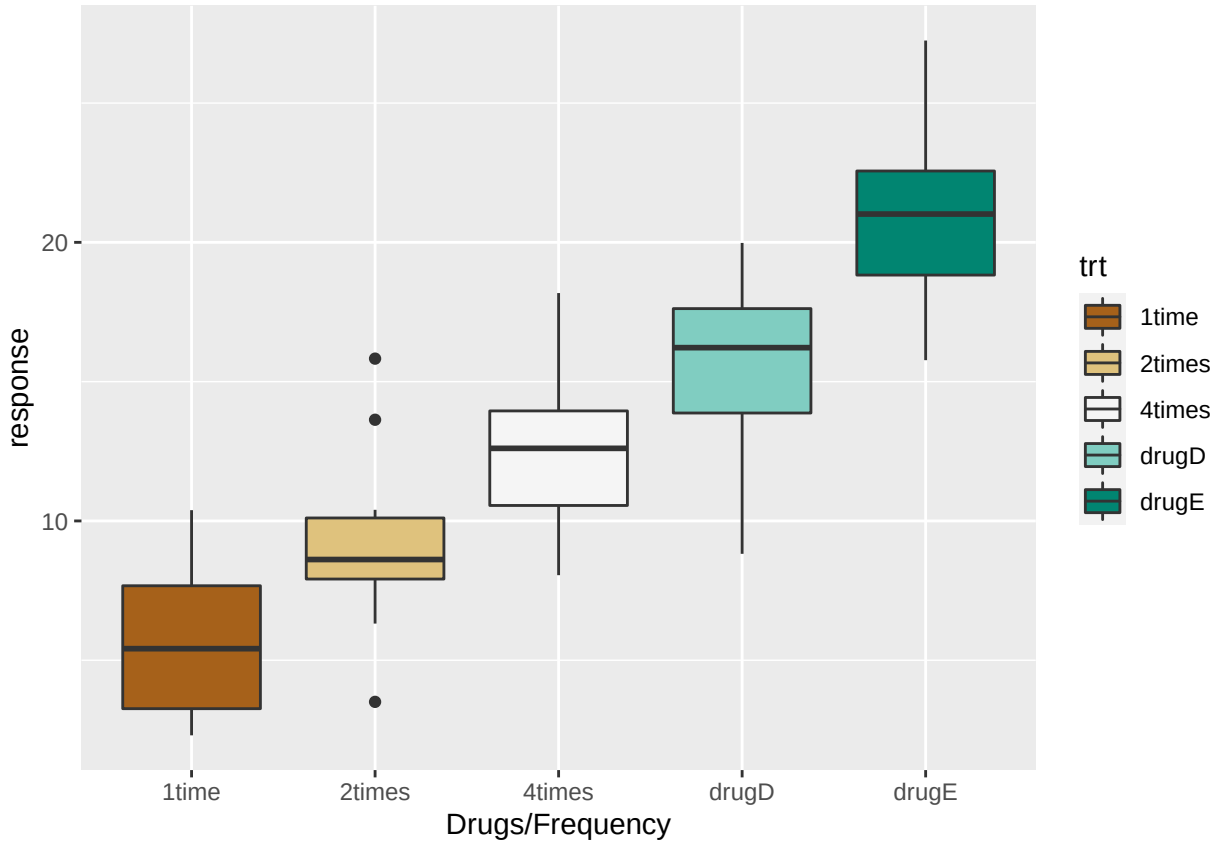

```
> plot(res.aov,2) # 正态性检验
```




```
> summary(res.aov) # 是否有显著性影响
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## trt          4 1351.4   337.8   32.43 9.82e-13 ***
## Residuals   45  468.8    10.4
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> cholesterol %>% # 箱线图绘制，对比不同药物/频次治疗方法效果
+   ggplot(aes(x = trt,y = response,fill = trt))+
+   geom_boxplot()+
+   labs(x = "Drugs/Frequency")+
+   scale_fill_brewer(palette = "BrBG")
```



Quesiton 3 answer description

由方差齐性检验图可知, 残差值 residual 与方差没有线性关系, 因此样本方差齐性较好.

由正态性检验图可知, 样本数据的正态性较好.

由 summary 的结果可知, 不同药物处理带来的效应明显, 因此通过绘制箱线图来对比不同药物治疗或不同频次治疗的效果差异.

由箱线图可知, 原药物 20mg 一天一次 (1time) 降低胆固醇最多.