

在 R 中进行数据正态性检验

李详

目录

1 为什么要进行正态性检验	1
2 shapiro.test 函数检验	1
3 直方图判断	2
4 Q-Q 图判断	5
5 小节	5

1 为什么要进行正态性检验

进行 t 检验、方差分析、相关性分析等数据分析时，都要求数据服从正态分布或者近似正态分布。但是这个条件往往被忽略。为了保证数据满足上述方法的使用条件，对数据进行正态性检验是十分重要的。本文就简单介绍如何在 R 语言中对数据进行正态性检验。本文的测试数据为 R 内嵌数据集 `iris`。

2 shapiro.test 函数检验

R 基础函数 `shapiro.test()` 可以对数据进行正态性检验。根据输出的 p -value 和 W 进行判断数据是否服从正态分布。 p -value <0.05 说明数据不符合正态分布； W 越接近 1 说明数据越接近正态分布。

```
suppressMessages(T)
```

```
## [1] TRUE
```

```
shapiro.test(iris$Sepal.Length)
```

```
##
```

```
## Shapiro-Wilk normality test
```

```
##
```

```
## data: iris$Sepal.Length
```

```
## W = 0.97609, p-value = 0.01018
```

根据输出的 $p\text{-value}=0.01018<0.06$, 说明数据不符合正态分布。

3 直方图判断

直接根据直方图判断数据是否服从正态分布。

1. 先看一个服从正态分布的例子

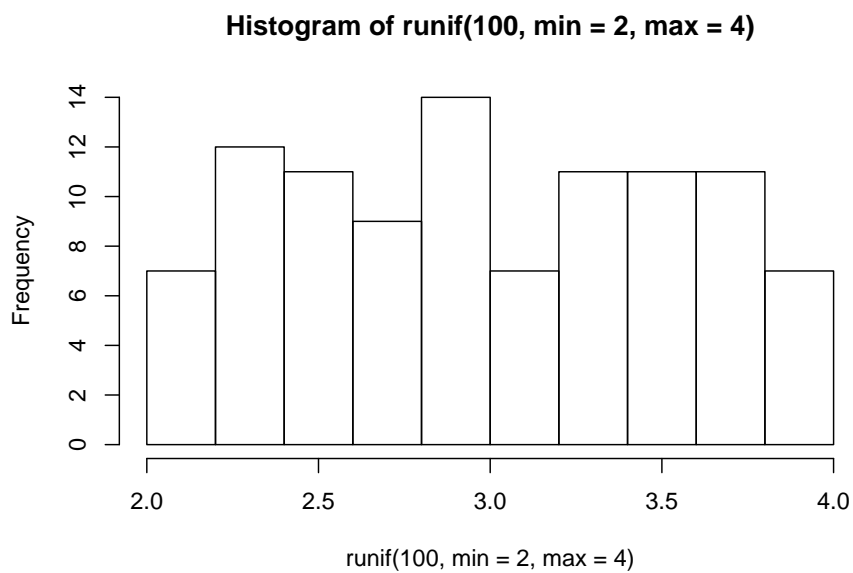
```
set.seed(123)
```

```
hist(rnorm(100,mean = 5, sd = 1))
```



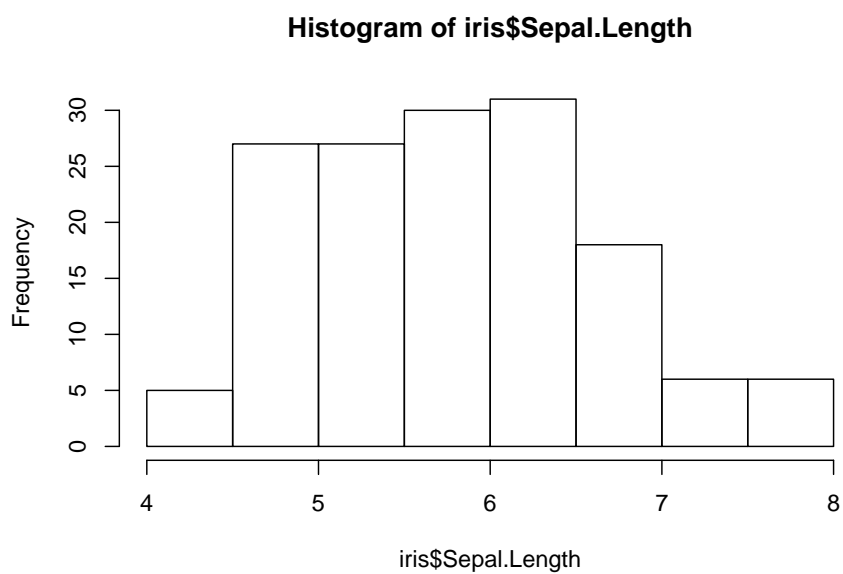
2. 再看一个不服从正态分布的例子

```
set.seed(123)
hist(runif(100, min = 2, max = 4))
```



3. 再看我们的测试数据

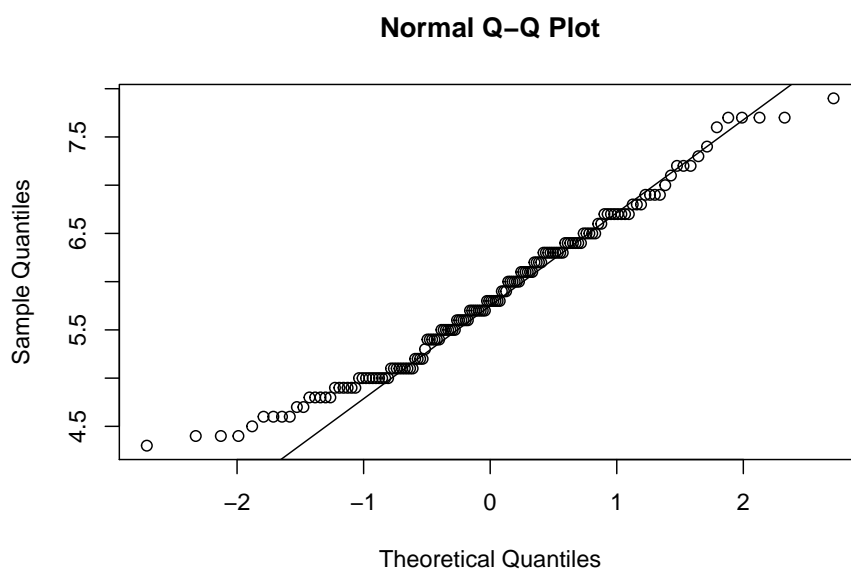
```
hist(iris$Sepal.Length)
```



4 Q-Q 图判断

Q-Q 图是根据数据的分位情况进行判断数据是否服从正态分布。使用 R 基础函数 `qqnorm()` 和 `qqline()` 即可绘制 Q-Q 图。x 轴是理论分位数，Y 轴是数据的分位数。如果数据符合正态分布，那数据点应该很好地拟合给出的直线。

```
qqnorm(iris$Sepal.Length)
qqline(iris$Sepal.Length)
```



5 小节

数据正态性检验很重要。数据正态性检验的方法很多，掌握一两种即可。