

第一次作业

1 第一题

此作业源代码丢失，为重写代码/作业，为简略只做第一问，第二问类似可做。

以下分析time变量

1.1 统计描述

最小值：5.0

25%分位数：166.0

中位数：266.0

75%分位数：413.0

最大值：1022.0

1.2 集中趋势

算数均数：307.8815

几何均数：227.7131

1.3 离散趋势

极差：1017

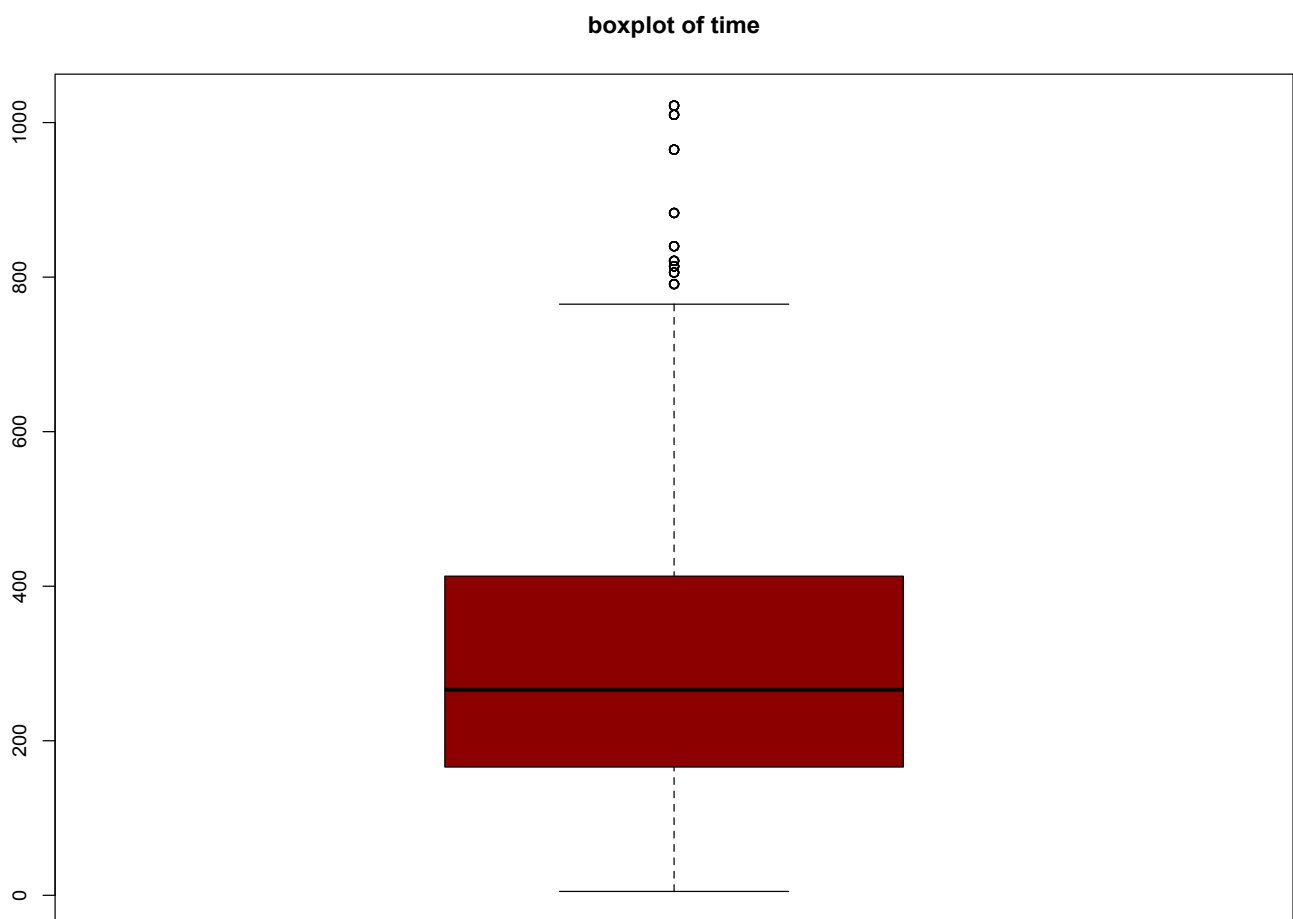
四分位差距：247

方差：44934.96

标准差：211.9787

变异系数：68.85074%

1.4 箱式图



1.5 正态性检验

以下检验假设部分为后续课程内容，第一次作业应该不会要求，在此仅作展示

建立检验假设、设立检验水准

H_0 : 总体服从正态分布

H_1 : 总体不服从正态分布

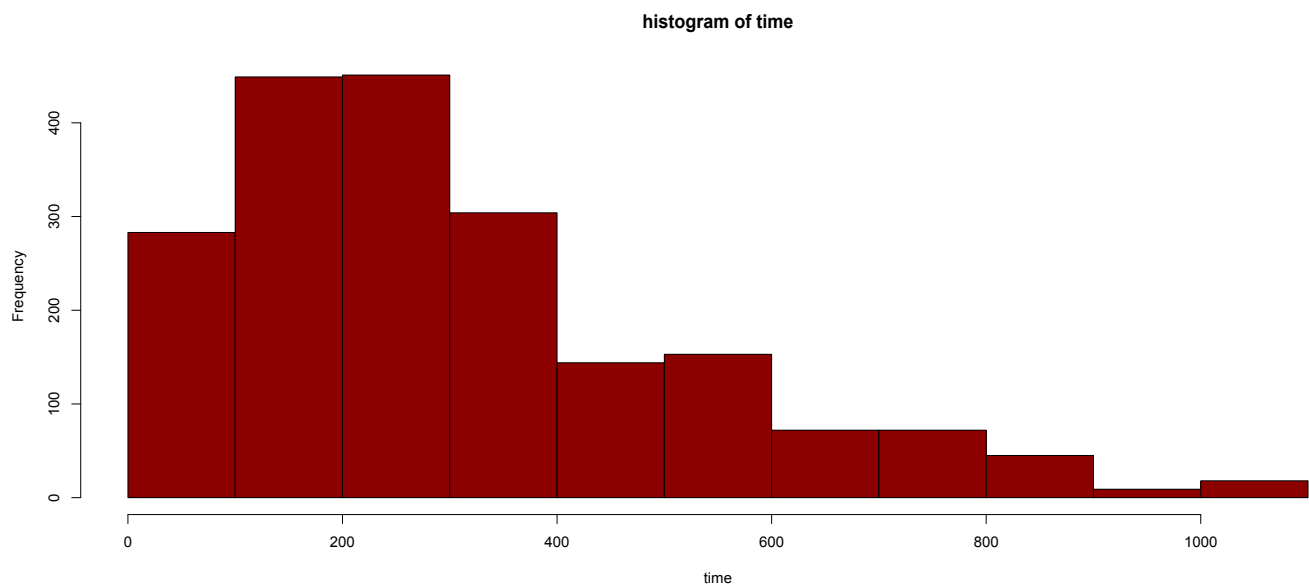
$\alpha = 0.05$

(1)

使用Shapiro-Wilk正态性检验，得到统计值 $W = 0.91915$, $p - value < 2.2e - 16 < \alpha$ ，从而拒绝 H_0 ，接受 H_1 ，即认为样本总体不服从正态分布。

1.6 正态性检验-峰度/偏度

绘制直方图，得到



从而得到数据为正偏、右偏。

引入moments包后计算得到：

偏度为 $1.057964 > 0$ ，对比正态分布的偏度为0

峰度为 $3.803648 > 3$ ，对比正态分布的峰度为3

得出结论：样本总体不服从正态分布。

2 附录

附上所有源代码：

```
# -*- coding: utf-8 -*-
# @Author :Arthals
# @File :Homework1.r
# @Time :2023/01/25 17:47:40
# @Software: Visual Studio Code

rm(list = ls()) # 清空工作空间

# 此作业源代码丢失，为重写代码/作业，为简略只做第一问，第二问类似可做。

# 第一题
# 使用R软件对time 、 value变量进行定量资料的统计描述，包括集中趋势、离散趋势、绘制箱式图、正态性检验，将软件的统计分析界面整理到A4纸上
data <- read.csv("课件&作业/作业1-定量资料的统计描述-正态分布/实习1-课后作业数据.csv")

time <- data$time
```

```
value <- data$value

# 以time为例
# 统计描述
summary(time) # 算数均数、中位数、四分位数、最大值、最小值

# 集中趋势
mean(time) # 算数均数
exp(mean(log(time))) # 几何均数
table(time)[which.max(table(time))] # 众数

# 离散趋势
max(time) - min(time) # 极差
quantile(time, probs = 0.75) - quantile(time, probs = 0.25) # 四分位数间距
var(time) # 方差
sd(time) # 标准差
sd(time) / mean(time) * 100 # 变异系数, 百分数形式

# 绘制箱式图
boxplot(time, main = "boxplot of time", col = "#8C0000")

# 正态性检验
shapiro.test(time) # Shapiro-Wilk正态性检验
# 峰度和偏度, 用直方图观察
hist(time, main = "histogram of time", col = "#8C0000")
# 峰度和偏度, 安装包moments检验
install.packages("moments") # 安装包
library(moments) # 载入包
skewness(time) # 偏度
# 1.057964>0, 正偏、右偏, 正态分布的偏度为0
kurtosis(time) # 峰度
# 3.803648>3, 正态分布的峰度为3
```