

# 统计分析作业1

---

学号：16337060

姓名：丰泽霖

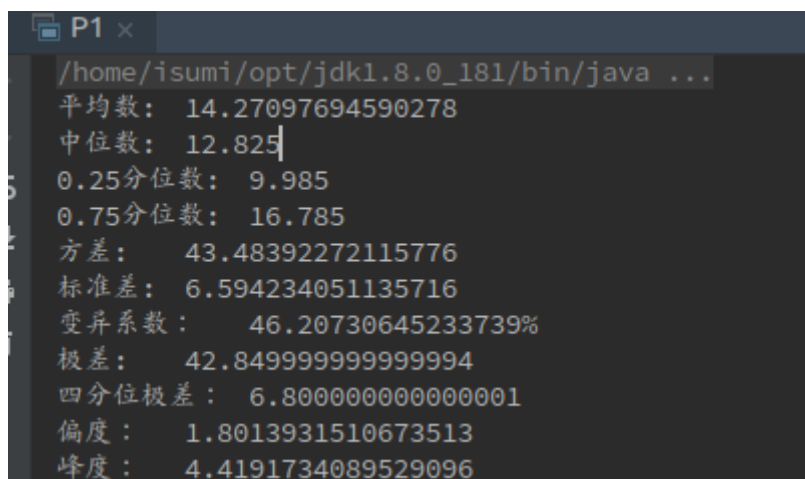
---

代码见随附的两个压缩包。

- statistics-hw 包含前三题，使用Scala + MATLAB实现
- statistics-hw-hs 包含第四题，使用纯Haskell实现

## 1

---



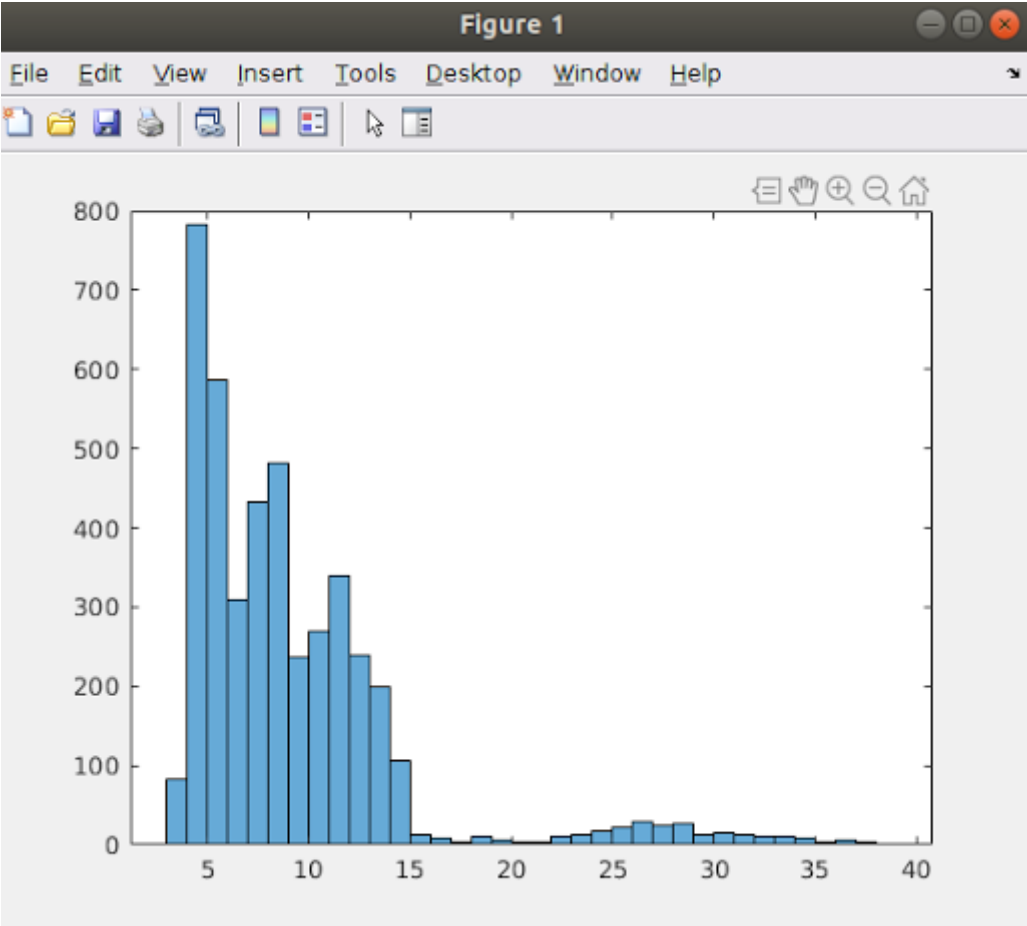
A screenshot of a Java command prompt window titled 'P1 x'. The window shows the output of a Java program, which is a list of statistical measures. The text is as follows:

```
/home/isumi/opt/jdk1.8.0_181/bin/java ...  
平均数: 14.27097694590278  
中位数: 12.825  
0.25分位数: 9.985  
0.75分位数: 16.785  
方差: 43.48392272115776  
标准差: 6.594234051135716  
变异系数: 46.20730645233739%  
极差: 42.849999999999994  
四分位极差: 6.800000000000001  
偏度: 1.8013931510673513  
峰度: 4.4191734089529096
```

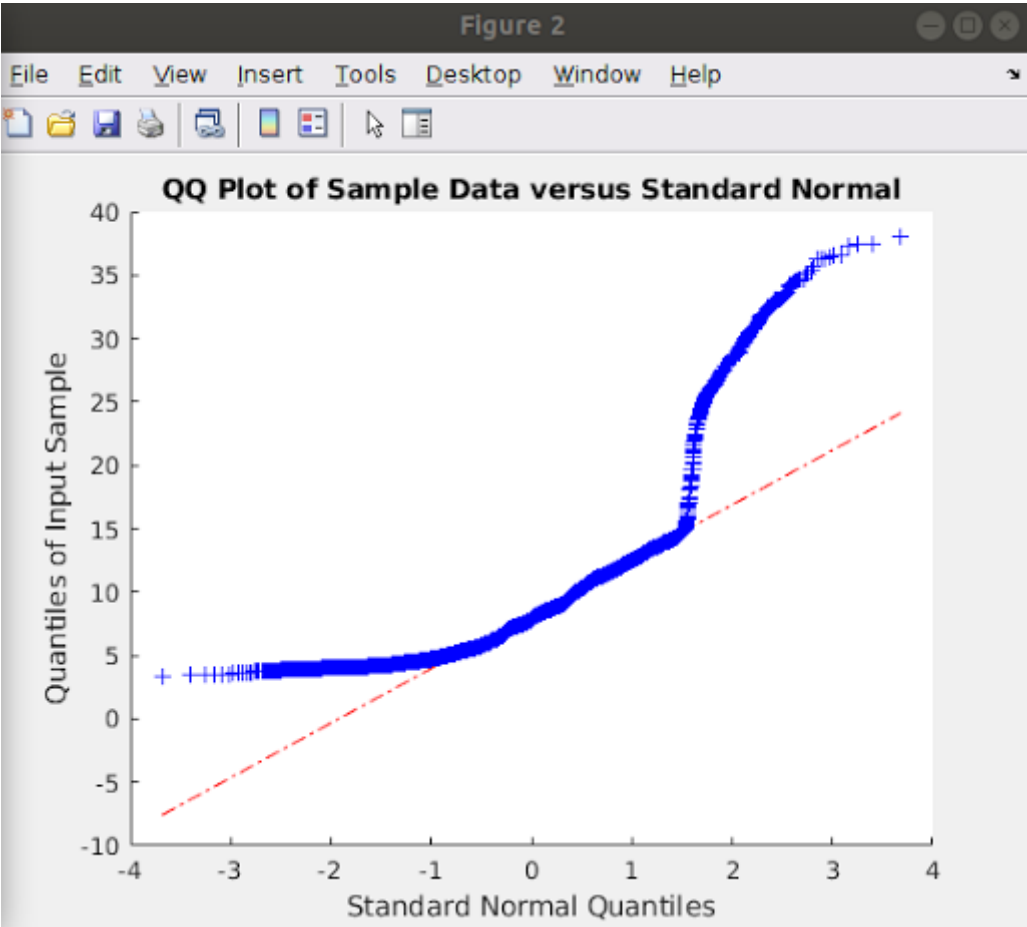
## 2

---

股价直方图

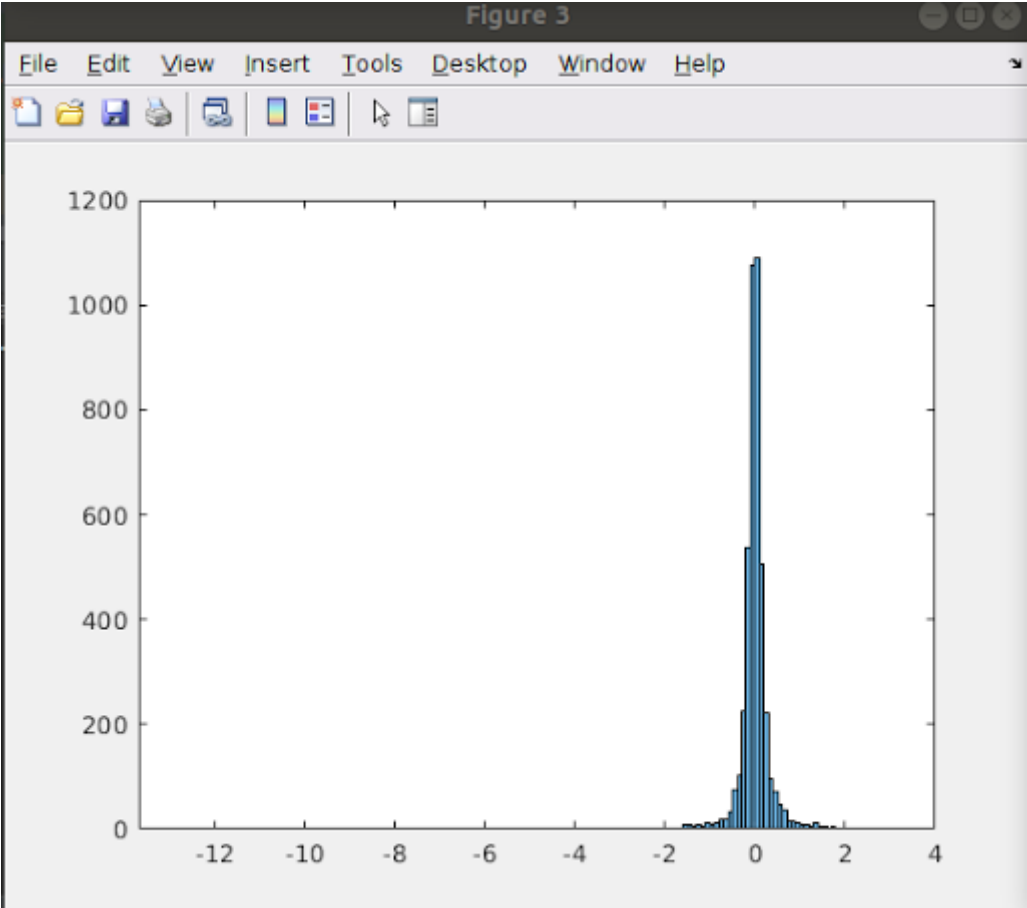


股价正态QQ图

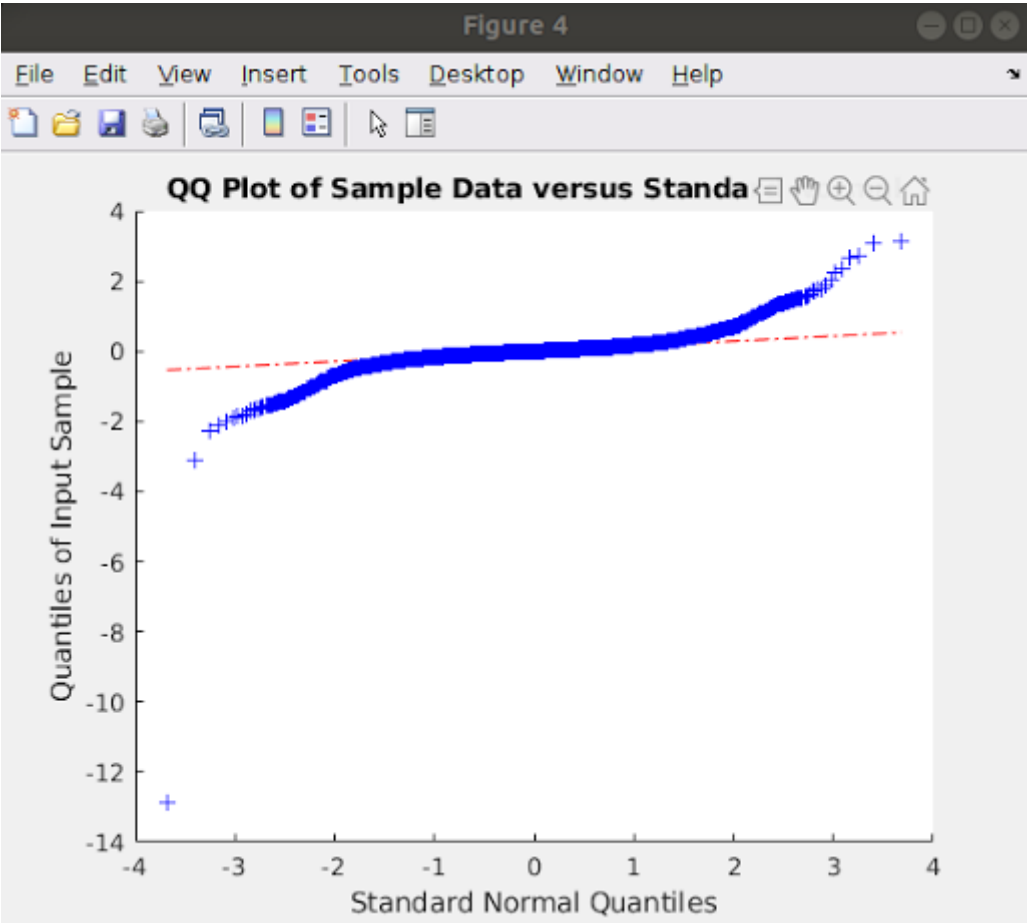


可见股价应当不服从正态分布，因为正态QQ图不在一条直线

股价差值直方图



股价差值正态QQ图



股价差值大致服从正态分布，直方图形状直观上很像正态分布，并且正态QQ图也大致在一条直线上

## 3

股票000012股价和成交量的Pearson和Spearman相关系数

```
/home/isumi/opt/jdk1.8.0_181/bin/java ...  
Pearson 相关系数: 0.029815915193986395  
Spearman 相关系数: -0.018695952028035114
```

## 4

第四题使用Haskell重写，运行结果如图

```
tsumi@tsumi-MIPRO:~/workspace/zellnf/statistics-hw-hs$ time stack exec s4 -- data  
Solution 4  
Pearson 相关性最强的5对  
股票"000046"和股票"000069"的相关系数为: 0.9494846389143637, p值为: 0.0  
股票"000046"和股票"000006"的相关系数为: 0.946448184961113, p值为: 0.0  
股票"000025"和股票"000567"的相关系数为: 0.9270517821307613, p值为: 0.0  
股票"000069"和股票"000006"的相关系数为: 0.9150686814687825, p值为: 0.0  
股票"000708"和股票"000059"的相关系数为: 0.9061597116938672, p值为: 0.0  
Pearson 相关性最弱的5对  
股票"000525"和股票"000632"的相关系数为: -2.8548602317842543e-4, p值为: 0.9849317585223495  
股票"000090"和股票"000049"的相关系数为: 6.476817190809877e-4, p值为: 0.9662156046568079  
股票"000521"和股票"000661"的相关系数为: -8.27580343252603e-4, p值为: 0.9564784342148039  
股票"000601"和股票"000661"的相关系数为: 1.0438976200818856e-3, p值为: 0.9452637333433266  
股票"000036"和股票"000425"的相关系数为: -1.68031448597759e-3, p值为: 0.9126746287569707  
Spearman 相关性最强的5对  
股票"000028"和股票"000661"的相关系数为: 0.9615927289440874, p值为: 0.0  
股票"000025"和股票"000567"的相关系数为: 0.9431301023573373, p值为: 0.0  
股票"000025"和股票"000418"的相关系数为: 0.9360251004541742, p值为: 0.0  
股票"000423"和股票"000661"的相关系数为: 0.9311030785456761, p值为: 0.0  
股票"000702"和股票"000421"的相关系数为: 0.9308496459665727, p值为: 0.0  
Spearman 相关性最弱的5对  
股票"000667"和股票"000567"的相关系数为: -3.678320318063746e-4, p值为: 0.9805995240034292  
股票"000078"和股票"000538"的相关系数为: -3.9880469825196476e-4, p值为: 0.979225317241791  
股票"000088"和股票"000001"的相关系数为: -6.760491563398692e-4, p值为: 0.9648234758699112  
股票"000418"和股票"000667"的相关系数为: 7.41872512418377e-4, p值为: 0.9611931821165076  
股票"000544"和股票"000425"的相关系数为: -1.3241252559623492e-3, p值为: 0.9313191683696738  
  
real    0m14.596s  
user    0m14.206s  
sys     0m0.424s
```

## 编译和运行说明

前三题需要使用IntelliJ IDEA，以sbt工程的形式导入，运行Problem1 ~ Problem3的main方法即可。注意添加JVM参数，参见README

第四题需要使用Haskell Stack。执行

```
stack build  
stack exec s4 -- /path/to/data/directory
```

注意通过命令行参数提供数据集所在目录