

## 第一章作业

16337207 石恬 智能科学与技术

- 一、求股票 000001（股票代码）的历史股价的日均值（所有天数的股价求平均）、中位数、0.25 分位数、0.75 分位数，方差，标准差，变异系数，极差，四分位极差，偏度，峰度。

日均值：所有天数的股价求平均

结果为：14.270976945902780

中位数：将所有天数股价排序后，1/2 位置的元素

结果为：12.825000000000000

0.25 分位数：将所有天数股价排序后，1/4 位置的元素

结果为：9.985000000000000

0.75 分位数：将所有天数股价排序后，3/4 位置的元素

结果为：16.780000000000000

方差：所有天数股价减去日均值的平方，再求和

结果为：43.483922721157874

标准差：方差的算术平方根

结果为：6.594234051135725

变异系数：标准差/均值 \* 100%

结果为：0.462073064523375

极差：最大值减去最小值

结果为：42.849999999999994

四分位极差：0.75 分位数减去 0.25 分位数

结果为：6.795000000000002

偏度：
$$g = \frac{n}{(n-1)(n-2)} \frac{1}{s^3} \sum_{i=1}^n (x_i - x')^3$$

结果为：1.801393151067342

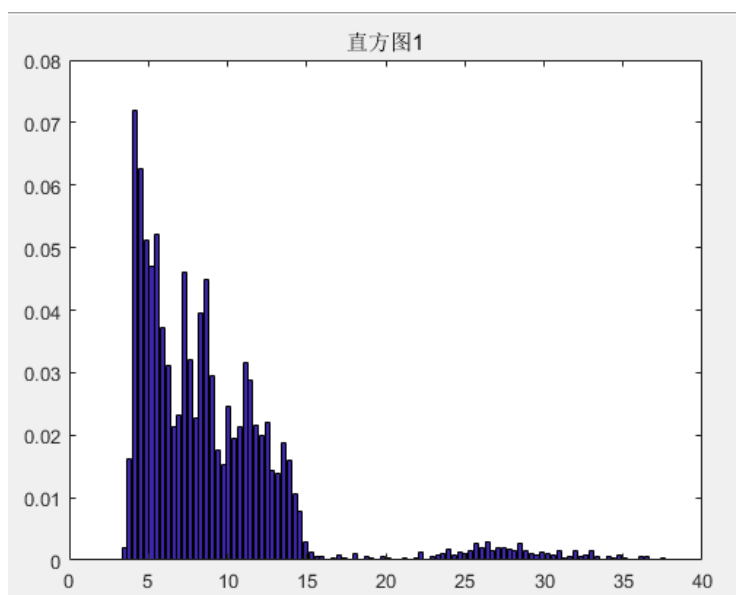
峰度：
$$g = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \frac{1}{s^4} \sum_{i=1}^n (x_i - x')^4 - \frac{3(n-1)^2}{(n-2)(n-3)}$$

结果为：4.419173408952862

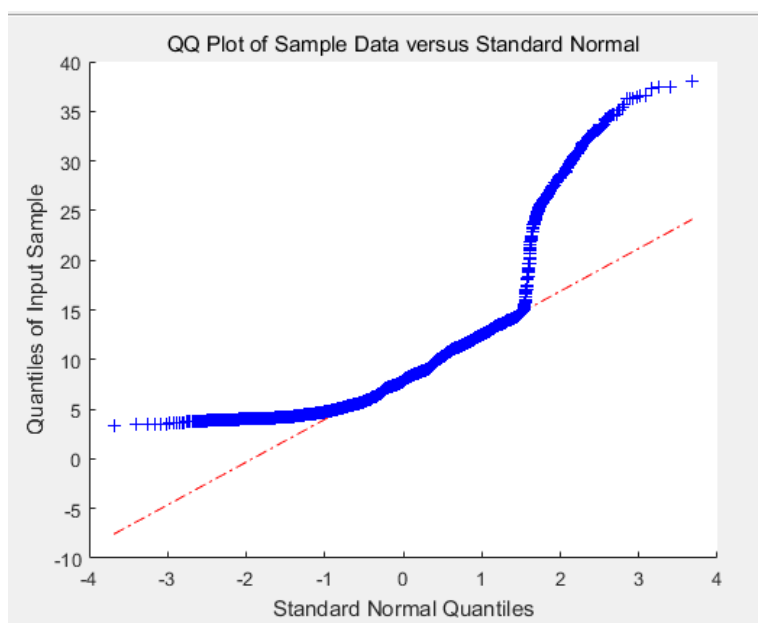
二、 对股票 000006 股价进行分析，选取合适组距，进行统计，画出的直方图（价格-频率）和正态 QQ 图，直观判断数据是否来自正态分布总体，给出简要的判断依据。如果对 000006 股价的差值（相邻两个日期的股价差值，忽略缺失日期，例如有  $t_1$ ,  $t_3$ ,  $t_4$ , 则差值为:  $t_3 - t_1$ ,  $t_4 - t_3$ ），同理计算差值的直方图和正态 QQ 图，判断差值是否服从正态分布，给出简要的判断依据。

对股票 000006 股价进行分析：

直方图：



QQ 图：

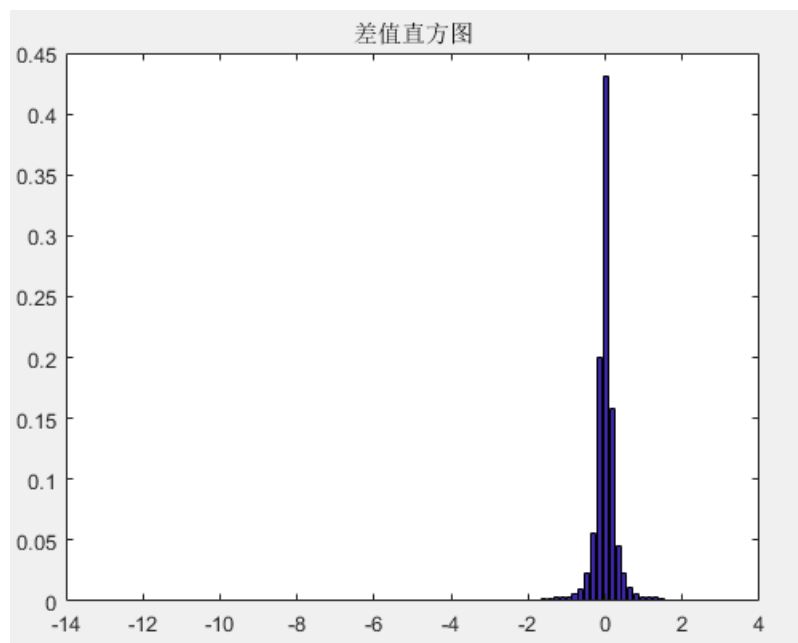


从直方图中可以看出，股价并不是正态分布的

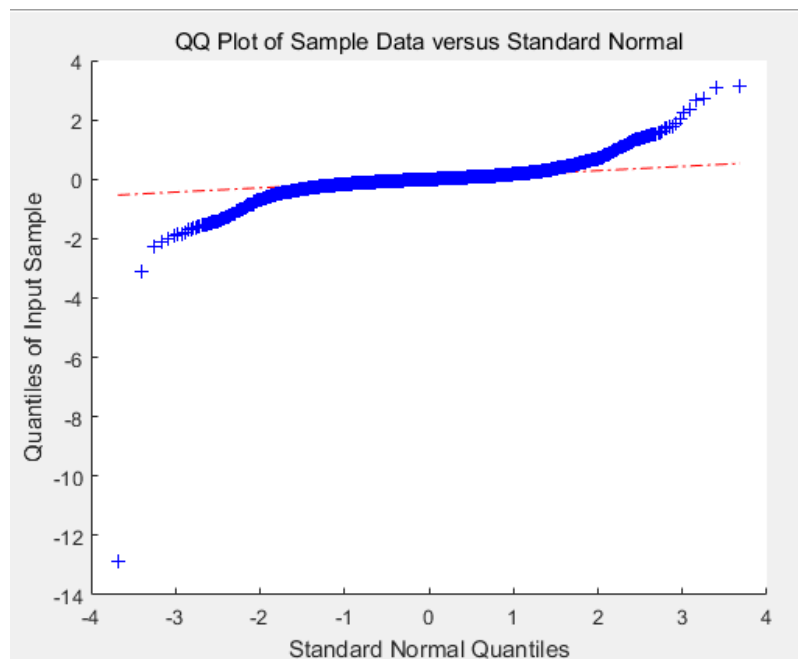
从 QQ 图中可以看出，股价是一个偏度为正，峰度为负分布

对 000006 股价的差值进行分析：

直方图：



QQ 图：



从直方图中可以直观地看出，股价服从正态分布

从 QQ 图中可以看出，股价基本是正态分布

三、 对股票 000012 进行分析，求股价和成交量的 Pearson, Spearman 相关系数。

$r1 = 0.0298$  (pearson 相关系数)

$r2 = -0.0187$  (spearman 相关系数)

可以基于相关系数判断两个变量间线性关联性的强弱。

由 Pearson, Spearman 相关系数可以看出股价与成交量之间的相关性并不强。

四、 按照日期，对股票 000001 和股票 000006 的股价进行相关分析。例如股票 000001 在  $t_1, t_2, t_4, t_5$  四个日期有记录  $x_1, x_2, x_4, x_5$ ；股票 000006 在  $t_2, t_3, t_4$  三个日期有记录  $y_2, y_3, y_4$ ，那么我们选取有共同日期记录的值， $t_2, t_4$  两个日期的记录，即  $(x_2, y_2)$  和  $(x_4, y_4)$  进行相关分析，而丢掉缺失数据（即  $t_1, t_3, t_5$  日期的数据）。推广之，对 100 支股票两两进行分析，求 100 支不同股票股价的 Pearson, Spearman 相关矩阵（ $100 \times 100$ ）。根据相关矩阵，给出这 100 只股票中，相关性最强（绝对值接近 1）的 5 对股票和相关性弱（绝对值最接近 0）的 5 对股票，根据 10 支股票，求相关性假设的 p 值。（注意，Pearson, Spearman 矩阵的元素排列依照股票代码，即，000001, 000006, 000012, ..., 000717）。

100 支不同股票股价的 Pearson, Spearman 相关矩阵（ $100 \times 100$ ）：见表格 pearson.csv 和 spearman.csv

其中，对角线元素均为 1，对称位置元素值也相同

求假设检验的 P 值：

依据 pearson 相关矩阵判断，相关性强和相关性弱的共有 10 对股票，每一对可以求出一个 pearson p 值和一个 spearman p 值，共有 20 个 P

值：

ppp = 0        0        0        0        0

pps = 0        0        0        0        0

pnp = 0.9849    0.9662    0.9565    0.9453    0.9127

```
pns =0.0000    0.0000    0.0000    0.0000    0.5159
```

同样，依据 spearman 相关矩阵判断，相关性强和相关性弱的共有 10 对股票，每一对可以求出一个 pearson p 值和一个 spearman p 值，共有 20 个 P

```
spp = 0      0      0      0      0
```

```
sps = 0      0      0      0      0
```

```
snp = 0.0000    0.0000    0.0000    0.0000    0.0133
```

```
sns = 0.9806    0.9792    0.9648    0.9612    0.9313
```

（运行时，直接编译 first\_5 即可，first\_4 和 first\_6 是其中的函数）

## 五、 附录

代码见压缩包

其中 first\_1、first\_2、first\_3 分别对应 1、2、3 问，第四问直接编译 first\_5 即可，first\_4 和 first\_6 是其中的函数

注：运行时需导入工作区 workspace