

统计分析方法第一章作业

16337203 屈博雅

1. 股票 000001 分析

结果：日均值= 14.2709769459

中位数= 12.825

0.25 分位数= 9.985

0.75 分位数= 16.785

方差= 43.4839227212

标准差= 6.59423405114

变异系数= 0.462073064523

极差= 42.85

四分位极差= 6.8

偏度= 1.80139315107

峰度= 4.41917340895

分析：首先将每日股价通过 每日股价 = (每日最低价 low+每日最高价 high) /2 算出来，然后通过 python 的函数算出各个数据值。

代码： 运行 1.py

2. 股票 000006 分析

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

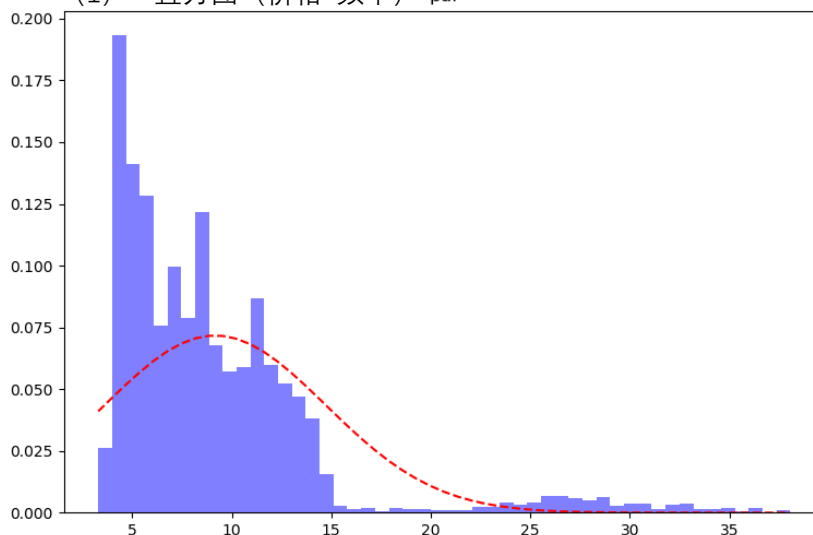
(价格-频率) 和正态 QQ 图，直观判断数据是否来自正态分布总体，给出简要的

判断依据。如果对 000006 股价的差值,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。

分析:

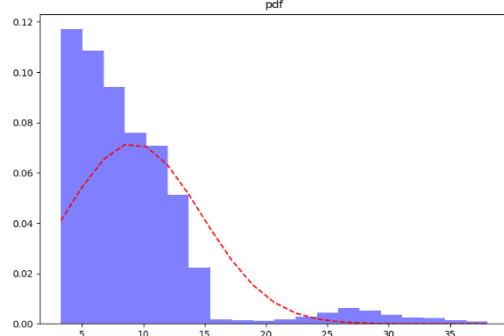
(1) 直方图 (价格-频率) pdf



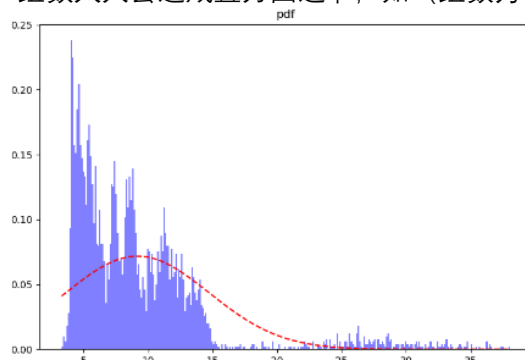
分为了 50 组，组距为 0.7，数据不来自正态分布，因为图中红线是根据 06 股票股价数据得到的正态分布函数曲线，明显不能与直方图相合。

选择这个组数是因为，

组数选小后最左边的信息会损失，如（组数为 20），

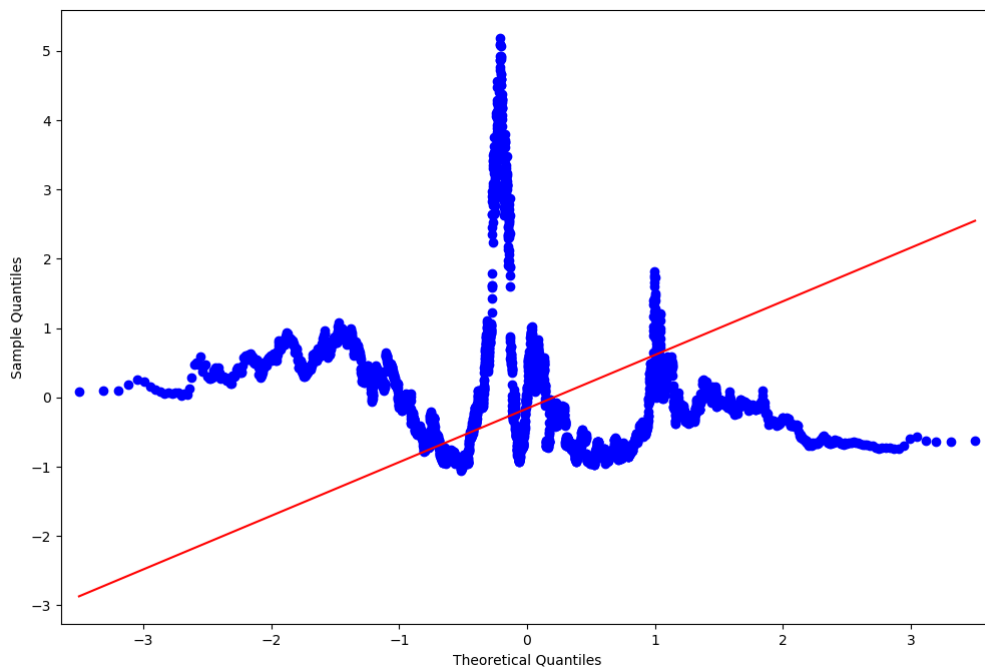


组数太大会造成直方图过窄，如（组数为 300），



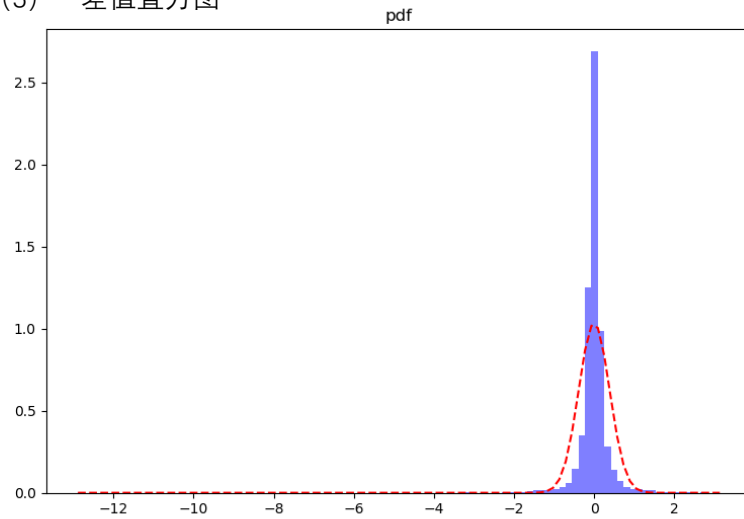
所以对比之下，50 是个合理的组数。

(2) 正态 QQ 图



这是股票 06 的每日股价的正态 QQ 图，蓝点没有在红线周围，看上去无特别关系，明显，数据不来自正态分布总体。

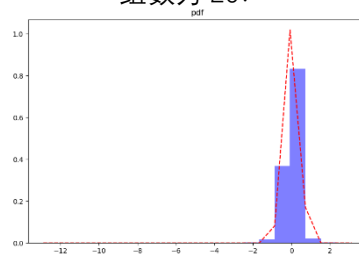
(3) 差值直方图



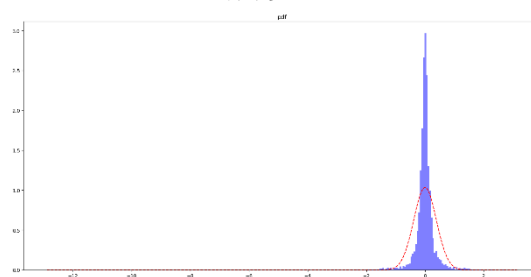
同样，分为 100 组，组距为 0.16。差值数据来自正态分布，因为图中红线是根据 06 股票股价数据得到的正态分布函数曲线，能与直方图相合。

选择 100 组的原因同上，

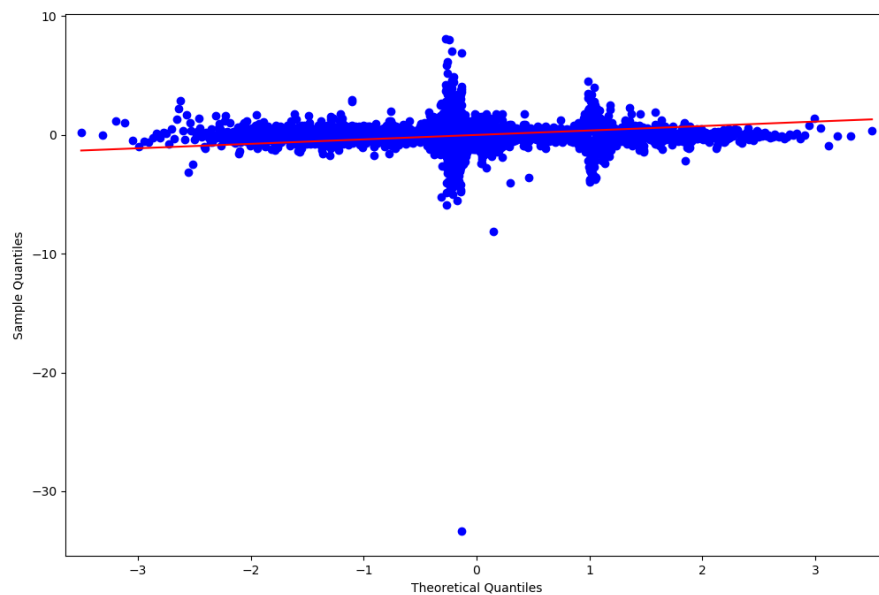
组数为 20:



组数为 300:



(4) 差值正态 QQ 图



这是差值的正态 QQ 图，蓝色的点基本拟合红线，只有少数几个点明显偏离，所以数据来自正态分布总体。

(5)代码：运行 2.py

3. 股票 000012 分析

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

结果：

pearsonr= 0.029815915194

spearsonr= -0.018695952028

分析：对于股票 000012，它的股价与交易量的相关性比较差。

代码：运行 3.py

4. 数据分析

(1) 股票 01 与 06 的分析

结果：

PearmanrResult= (0.77719365433573362, 0.0)

SpearmanrResult=(correlation=0.50989796095176776,pvalue=7.7006772434924242e-276)

(2) 全部股票的分析

结果见 2.xls，一共有 8 个列表。

介绍如下：

p_0_sheet —— 存放 pearmanr 的 r 值

s_0_sheet —— 存放 spearmanr 的 r 值

p_all_sheet —— 存放 pearmanr

s_all_sheet —— 存放 spearmanr

p_all_p_10_sheet —— 存放 p 算法十对的 pearmanr 值

p_all_s_10_sheet —— 存放 p 算法十对的 spearmanr 值

s_all_p_10_sheet —— 存放 s 算法十对的 pearmanr 值

s_all_s_10_sheet —— 存放 s 算法十对的 spearmanr 值

p_0_sheet 节选

	A	B	C	D	E	F	G	H	I	J	K	L
1	1	0.77719	0.63277	0.26197	0.27204	0.38157	-0.14402	0.39945	0.76994	-0.07587	0.15656	0.28248
2	0.77719	1	0.61296	0.52844	0.44222	0.56878	0.10743	0.53651	0.8593	-0.06112	0.49364	0.54752
3	0.63277	0.61296	1	0.38509	0.54345	0.62801	-0.04283	0.63501	0.60721	-0.06146	0.45496	0.52154
4	0.26197	0.52844	0.38509	1	0.28183	0.24739	0.67648	0.55679	0.41013	0.63478	0.54394	0.30835
5	0.27204	0.44222	0.54345	0.28183	1	0.88254	-0.00618	0.41993	0.40196	-0.04431	0.7102	0.57649
6	0.38157	0.56878	0.62801	0.24739	0.88254	1	0.00556	0.45058	0.52496	-0.15854	0.72916	0.57285
7	-0.14402	0.10743	-0.04283	0.67648	-0.00618	0.00556	1	0.25781	-0.09251	0.80068	0.47238	0.05546
8	0.39945	0.53651	0.63501	0.55679	0.41993	0.45058	0.25781	1	0.45757	0.21755	0.48489	0.61409
9	0.76994	0.8593	0.60721	0.41013	0.40196	0.52496	-0.09251	0.45757	1	-0.14196	0.28678	0.50036
10	-0.07587	-0.06112	-0.06146	0.63478	-0.04431	-0.15854	0.80068	0.21755	-0.14196	1	0.28543	0.00365
11	0.15656	0.49364	0.45496	0.54394	0.7102	0.72916	0.47238	0.48489	0.28678	0.28543	1	0.57514
12	0.28248	0.54752	0.52154	0.30835	0.57649	0.57285	0.05546	0.61409	0.50036	0.00365	0.57514	1

s_0_sheet 节选

	A	B	C	D	E	F	G	H	I
1	1	0.509898	0.652607	0.252553	0.280422	0.334924	0.205404	0.435712	0.518636
2	0.509898	1	0.729001	0.644775	0.704019	0.746312	0.536311	0.61688	0.69039
3	0.652607	0.729001	1	0.466016	0.540928	0.668094	0.367936	0.698213	0.716902
4	0.252553	0.644775	0.466016	1	0.443587	0.374184	0.827764	0.56795	0.313295
5	0.280422	0.704019	0.540928	0.443587	1	0.876847	0.444065	0.569927	0.636528
6	0.334924	0.746312	0.668094	0.374184	0.876847	1	0.354789	0.585423	0.734969
7	0.205404	0.536311	0.367936	0.827764	0.444065	0.354789	1	0.630546	0.119584
8	0.435712	0.61688	0.698213	0.56795	0.569927	0.585423	0.630546	1	0.425444
9	0.518636	0.69039	0.716902	0.313295	0.636528	0.734969	0.119584	0.425444	1
10	0.174512	0.040994	0.061656	0.597067	-0.09464	-0.18945	0.750418	0.271781	-0.21723
11	0.119535	0.71253	0.452584	0.635865	0.752539	0.723498	0.709487	0.554358	0.339387
12	0.211867	0.511819	0.467603	0.274733	0.701169	0.623559	0.389293	0.59775	0.455625

以下各十对（显示的是股票的 code），前五个为相关性小的，后五个为大的。

Pearsonrs:

- 第 1 对股票是(525, 632)
- 第 2 对股票是(49, 90)
- 第 3 对股票是(521, 661)
- 第 4 对股票是(601, 661)
- 第 5 对股票是(36, 425)
- 第 6 对股票是(59, 708)
- 第 7 对股票是(6, 69)
- 第 8 对股票是(25, 567)
- 第 9 对股票是(6, 46)
- 第 10 对股票是(46, 69)

Spearmanrs:

- 第 1 对股票是(567, 667)
- 第 2 对股票是(78, 538)
- 第 3 对股票是(1, 88)
- 第 4 对股票是(418, 667)
- 第 5 对股票是(425, 544)
- 第 6 对股票是(421, 702)
- 第 7 对股票是(423, 661)
- 第 8 对股票是(25, 418)
- 第 9 对股票是(567, 25)
- 第 10 对股票是(661, 28)

按照排序（0-99），10 支股票的序号为

Pearsonrs:

```
ij_p0_pos= [[45, 80], [14, 25], [43, 84], [74, 84], [10, 37], [16, 97], [1, 21], [6, 62], [1, 13], [13, 21]]
```

Spearmanrs:

```
ij_s0_pos=[(62, 85), (22, 50), (0, 23), (31, 85), (37, 54), (34, 93), (36, 84), (6, 31), (62, 6), (84, 9)]
```

所需十对的 p 值:

Pearsonrs 算法求出的 10 对股票:

Pearsonrs 算法的 p 值:

pea_10_pea_pvalue=

```
[ 0.9849317585223375,    0.96621560465683753,
  0.95647843421478063,    0.94526373334347769,
  0.91267462875677452,    0.0,    0.0,    0.0,    0.0,    0.0]
```

Spearmanrs 算法的 p 值:

pea_10_spea_pvalue=

```
[ 6.4444131291881671e-13,  8.7987478779004736e-41,
  6.904237548268161e-111,  6.1196742427551438e-06,
  0.51593188492860498,    0.0,    0.0,    0.0,    0.0,    0.0]
```

Spearmanrs 算法求出的 10 对股票:

Pearsonrs 算法的 p 值:

spea_10_pear_pvalue=

```
[ 4.9790929293911055e-08,  4.2848767275261499e-08,
  1.501553477406247e-11,    1.0339494366222513e-26,
  0.013307418053292877,    0.0,    0.0,    0.0,    0.0,    0.0]
```

Spearmanrs 算法的 p 值:

spea_10_spear_pvalue=

```
[ 0.98059952399803763,    0.97922531724940554,
  0.96482347586905515,    0.96119318211375626,
  0.93131916837179762,    0.0,    0.0,    0.0,    0.0,    0.0]
```

(3) 实现方法简单介绍

代码 4_2.py 为主体实现程序。首先处理得到每日股价，筛选时间，通过库函数得到两个 100x100 的矩阵，然后通过排序得到相关性最强（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，因为求相关性最强时对角线的值 1 和下三角的对称值会进行干扰，以及我的排序实现的形式，加上数据不大，所以我人工筛选了数据，去重去 1，得到 Pearmanr、Spearmanr 各十对股票，然后得到其两种 p 值，得到结果。

(4) 代码

运行代码 4_2.py