**统计分析方法作业1**

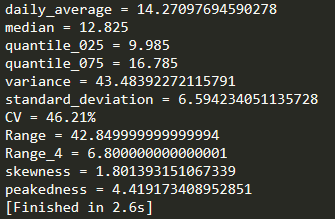
姓名：孟衍璋 学号：16337183

本次作业的前三个题目，都是用python实现的；最后一个题用matlab实现，原因是python实现计算速度太慢。代码内容在附件中，其中每个小问题用一份代码文件实现。

**题目1：**

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。

运行结果：



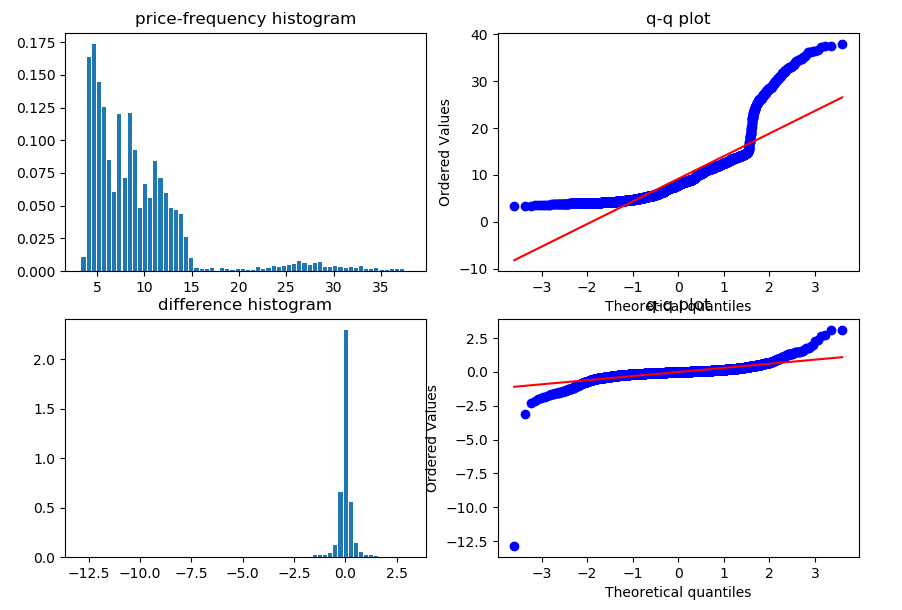
结果分析：

用pandas读取csv文件，每日股价是取当日最高价和最低价的平均值。之后的每个题目也都是用这种方法计算进行对应问题的分析。之后利用公式依次计算出题目要求求出的值。

**题目2：**

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 - t1, 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。

运行结果：



结果分析：

对股票000006的股价进行分析，依然使用pandas读取文件，用pyplot来绘图。在绘制直方图的过程中，选择样本数量的开方作为组数。从QQ图可以看出，股价的数据不是来自正态总体，因为图形没有与直线明显不拟合。而股价的差价来自正态总体，因为图形与直线基本拟合。

**题目3：**

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

运行结果：



结果分析：

导入数据之后，按照Pearson与Spearman的计算公式计算出结果。股价和成交量的Pearson相关系数为0.0298，P值为0.0472；Spearman相关系数为-0.0187，P值为0.2136。

**题目4：**

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, y2)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（100×100）。 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)。

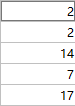
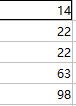
运行结果：

运行之后得到的两个系数矩阵分别保存在pearson\_matrix.mat和spearman\_matrix.mat中，这两个矩阵都是上三角矩阵，因为矩阵的下半部分是重复的数据。

之后对矩阵进行排序，分别选出绝对值最接近1和绝对值最接近0的5对数据。

**Pearson：**

可以看出绝对值最接近1的5对数据：

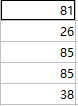
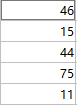
 

对应的文件是第2份与第14份、第2份与第22份、第14份与第22份、第7份与第63份、第17份与第98份。

计算结果如下：

因为这些对应的文件Pearson相关系数都接近1，所以是高度相关的，因此相关性假设的P值均为0。

之后再找到绝对值最接近0的5对数据：



对应的文件是第46份与第81份、第15份与第26份、第44份与第85份、第75份与第85份、第11份与第38份。

计算结果如下：

第一组：，相关性假设P值为0.9849

第二组：，相关性假设P值为0.9662

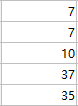
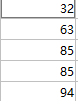
第三组：，相关性假设P值为0.9565

第四组：，相关性假设P值为0.9453

第五组：，相关性假设P值为0.9127

**Spearman：**

可以看出绝对值最接近1的5对数据：

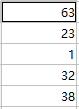
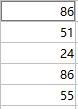
 

对应的文件是第7份与第32份、第7份与第63份、第10份与第85份、第37份与第85份、第35份与第94份。

计算结果如下：

因为这些对应的文件Spearman相关系数都接近1，所以是高度相关的，因此相关性假设的P值均为0。

之后再找到绝对值最接近0的5对数据：

对应的文件分别是：第63份与第86份、第23份与第51份、第1份与第24份、第32份与第86份、第38份与第55份。

计算结果如下：

第一组：

第二组：

第三组：

第四组：

第五组：