**课程实践作业六**

编写数据统计分析交互计算IPython Notebook

**问题：**

16.4 Statistical Measures Don't, Tell the Whole Story

**要求：**

1. 四组数据的数据文件：纯文本格式或MS Excel表格
2. 从数据文件读取数据，计算均值、方差和相关系数等统计指标
3. 对数据做线性回归
4. 绘制数据点图（含回归结果）

**提交：**

1. 文件压缩包：学号\_姓名\_作业六.zip；
2. 电邮：[cmh@seu.edu.cn](mailto:cmh@seu.edu.cn), 邮件主题：学号\_姓名\_作业六；
3. 推送到GitHub仓库。

**提示：**

可以基于课程IPython Notebook修改实现。

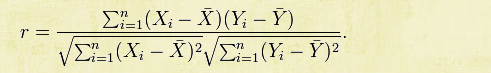
实践中遇到并解决的问题：

1 从excel获取第一个sheet从第0个起数而不是第1个开始

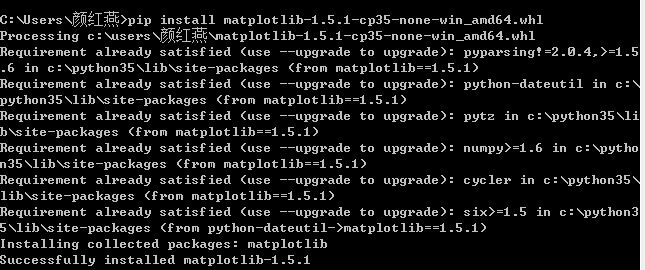
2 SyntaxError: invalid syntax 错误原因：python3.5不同于python2.x版本，print后面要加（）

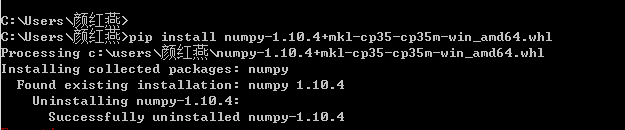
3 皮尔逊相关系数：在概率论和统计学中，相关（Correlation，或称相关系数或关联系数），显示两个随机变量之间线性关系的强度和方向。在统计学中，相关的意义是用来衡量两个变量相对于其相互独立的距离。

在推荐系统中，我们常用皮尔逊相关系数来衡量两个用户兴趣的相似度，它是判断两组数据与某一直线拟合程度的一种度量。它在用户对物品的评分数据差别大时（如有些用户评分普遍较高，有些用户评分普遍偏低）时的效果更好。也即它修正了“夸大分值”的情况，如果某个用户总是倾向于给出比另一个人更高的分值，而两者的分值之差又始终保持一致，则两者间依然可能存在很好地相关性。



4 查看pip list发现numpy和matplotlib没有安装完好，重新安装了一下





5 数据点图

