

# Python 数据分析第 1 次作业\*

金 林

中南财经政法大学统计系

jinlin82@gmail.com

2020 年 4 月 2 日

## 1 作业内容

1. 分别创建一个列表, 元组, 字典, 并提取其子集
2. 利用 numpy 生成一个一维数组和二维数组
3. 0-100 之间生成等步长的 201 个数; 把列表 [1,4,9] 重复 10 遍
4. 随机生成 8 乘以 20 的整数 array, 并得到其转置数组, 对其转置数组提取前 10 行, 第 2,5,8 列数据
5. 利用 pandas 把 data 文件夹中的数据.xls 导入, 要求使用相对路径
6. 显示数据.xls 数据集的结构, 前 5 条数据, 后 5 条数据, 所有变量名, 及其维度
7. 对数据集增加一个变量 HRsq, 其为 HR 数据的平方
8. 对数据.xls 数据集进行子集的提取:
  - (a) 删除有缺失值的年份
  - (b) 提取逢 5, 逢 0 年份的 Year, GDP, KR, HRsq, CPI 变量的数据
  - (c) 提取逢 2, 逢 8 年份的 Year, GDP, KR, HRsq, CPI 变量的数据
  - (d) 对 2 和 3 得到的数据集按行进行合并, 并按年份进行排序
  - (e) 提取 1978 年之后的数据
  - (f) 提取 1978 年之后且 KR 变量在 1~1.2 之间的数据
9. 保存数据为 csv 和 excel
  - (a) 写出第 8 题中第 4 问得到的子集到 data 文件夹
  - (b) 写出第 8 题中第 6 问得到的子集到 data 文件夹

---

\*基金项目:

## 2 作业要求

- 提交 1 个 Python 程序文件和 data 文件, 压缩, 发送到邮箱 znufestat@qq.com
- 文件名和邮件名都按照作业要求命名 (具体见课程说明)
- 截至日期: 2020 年 4 月 7 日 (星期二) 晚上 12 点之前