

基于Python的 数据分析与可视化

第五讲 数据分析中的数学、统计描述与假设检验

刘子瑜 副研究馆员

liuziyu@cau.edu.cn 中国农业大学图书馆情报研究中心

<https://github.com/CAU-LiuZiyu/Data-Analysis-with-Python>

下列代码的输出结果是？

```
students = ["zhang", "lin", "chen"]  
students.sort()  
print('students')
```

- ☒ A students
- ☐ B 'chen meng', 'lin song', 'zhang ziqian'
- ☐ C 'zhang ziqian','lin song','chen meng'
- ☐ D 报错

提交

下列说法中错误的是()

A.当利用import导入模块，显示出“func导入成功”后，将func文件从硬盘中删除，则不会再显示“func导入成功”。

B.定义一个函数的基本格式如下：()

def 函数名：

函数体

return 返回值

C.利用random库可以实现简单的金额随机分配，由于每次各位置产生的金额相差过大，所以多次试验(如1万次)每个位置的平均金额值相差也很大。

D.如果在利用random库时不设置随机种子，直接print(random())，则会以当前的时间为默认值，每一次重新选取都会根据时间产生不同的值。

☒ A☒ B☒ C☐ D

提交

在数据输出时，可以运用一段代码进行修饰，控制输出数据的样式。下面代码与输出结果对应正确的是()：

- ☐ A `print("{0:@<25}".format(9.87654321))`
9.87654321@@@@@@@@@@@@@@@@@@
- ☒ B `print("{0:,}".format(6000000))`
6,000,000
- ☐ C `print("{0:.2f}".format(PI))`
3.1
- ☐ D `print("{0:.2%}".format(PI))`
314.15%

提交

以下说法错误的是

- ☐ A 迭代器没有长度
- ☐ B range()不会被耗尽
- ☐ C 闭包中内层函数可以引用外层函数的变量
- ☒ D 装饰器不能传递额外的参数

提交

第二次作业（4.5提交）

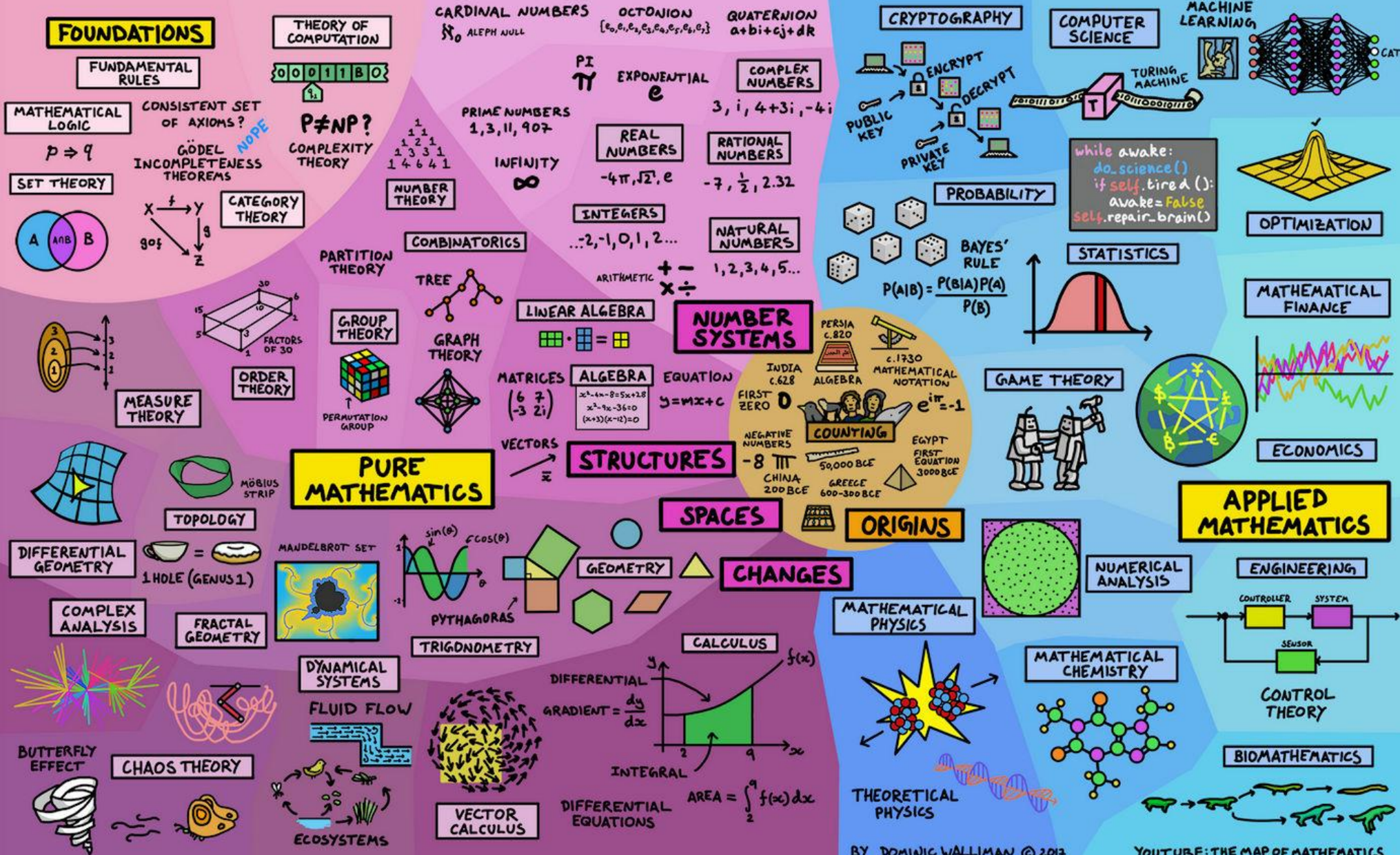
选项1：在“了不起的数学”项目中，选择一项与数据分析相关的学习资源进行自学，并提交资料来源（如网页链接、pdf文档、实体书照片等，建议大家充分利用我校图书馆资源，包括读秀、电子书、各类西文中文数据库等）和学习笔记（学习一点点也可以，毕竟时间很短）。

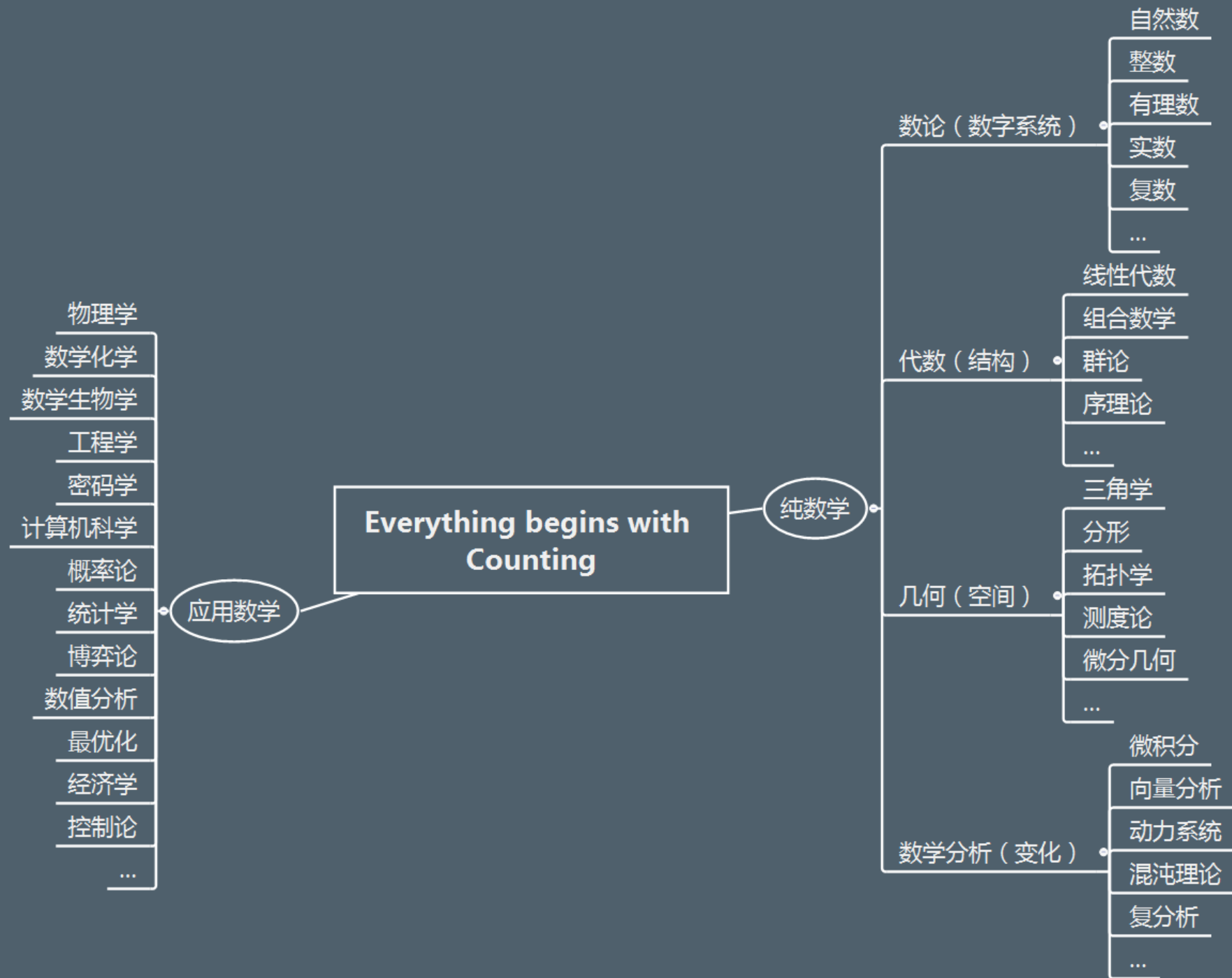
<https://github.com/rossant/awesome-math>

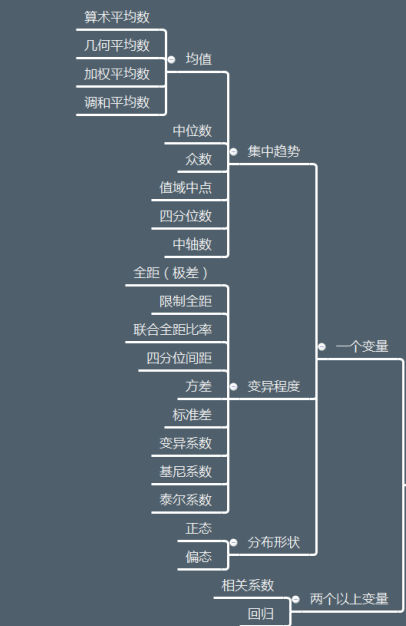
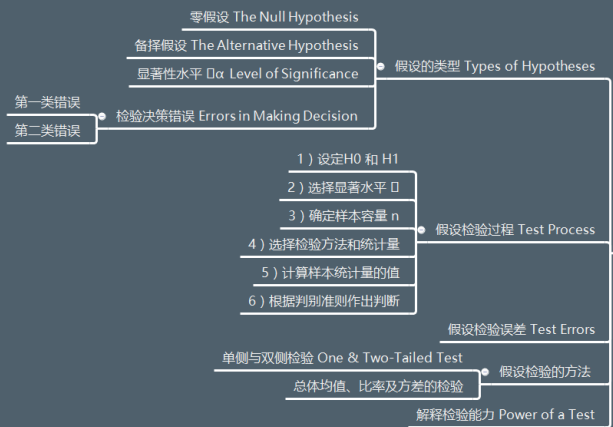
选项2：寻找我国2013-2017年城镇居民收入数据，并通过编程进行简单的统计描述，数据、程序、结果都保留在ipynb文件中。

选项3：基于<https://setosa.io/ev/>网站，学习普通最小二乘回归、主成分分析、条件概率（三者之一），并录一个5-10min的解说视频。

THE MAP OF MATHEMATICS



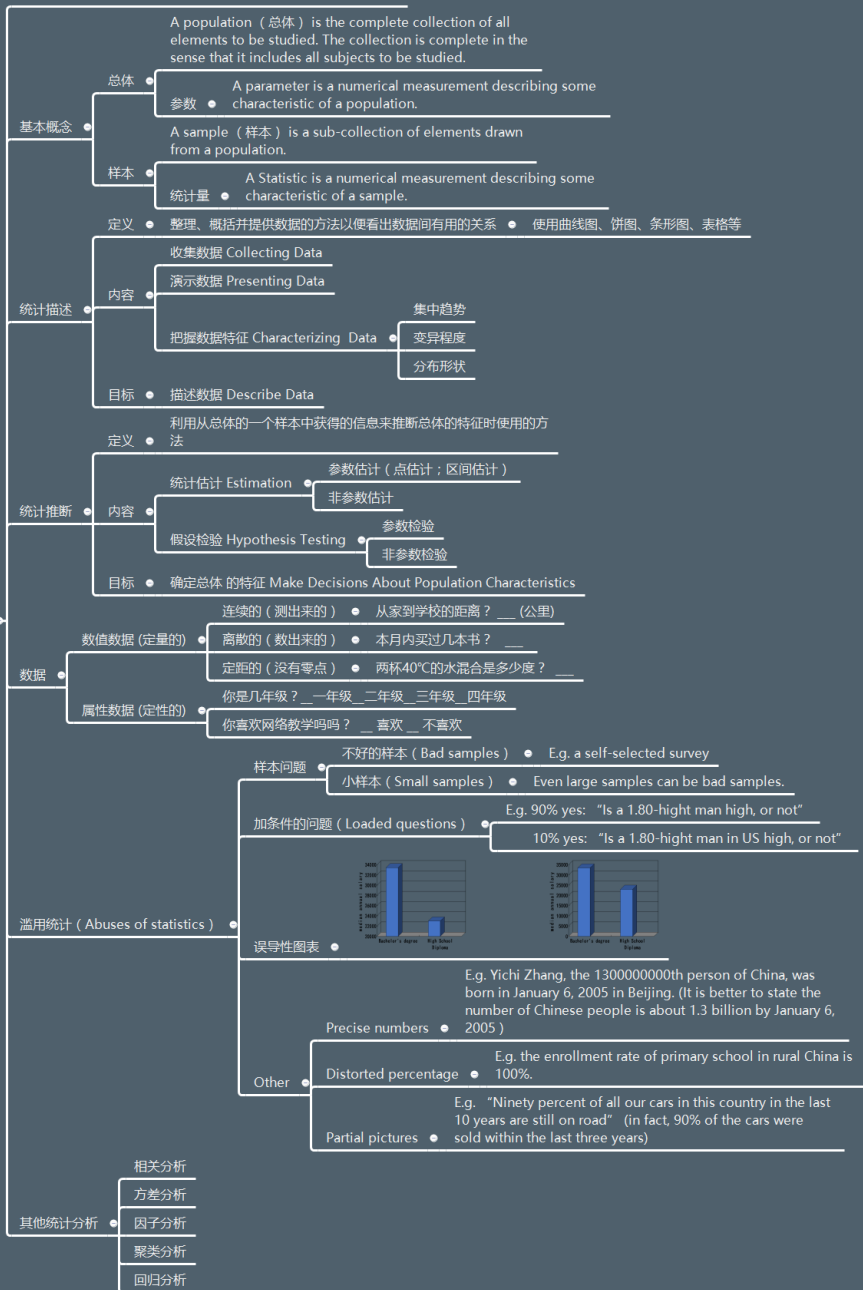




概率统计

什么是统计学

统计学是从一组数据中抽象出有用信息以支持决策的原理和方法。
统计学的研究领域或方法分为两部分：统计描述 Descriptive Statistics和统计推断 Inferential Statistics



此题未设置答案，请点击右侧设置按钮

一个学生阅读2页书，读第一页时的速度折合为每小时20页，读第二页时的速度折合为每小时40页，问该生平均每小时阅读速度是多少？（单位：每小时页数）

算术平均数 [填空1]

调和平均数 [填空2]

正常使用填空题需3.0以上版本雨课堂

作答