

概率论（二）

（数理统计学）

程东亚

苏州大学数学科学学院

dycheng@suda.edu.cn



什么是统计学？

统计学的英文 **statistics**，最早代表对国家的资料进行分析的学问，也就是“研究国家的科学”。

抽样调查、试验设计、.....

Statistics, the science of collecting, analyzing, presenting, and interpreting data.

统计学是**收集**、分析、展示和解释数据的科学。

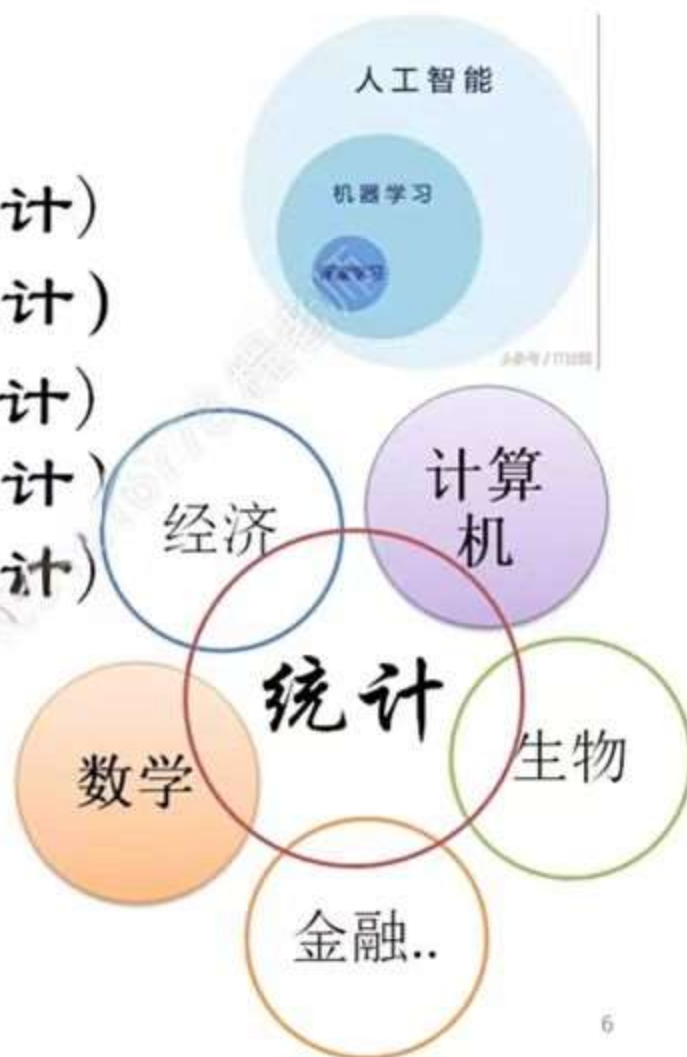
——不列颠百科全书(2008-2020年)

<https://www.britannica.com/science/statistics>



统计学分支

数理统计 (数学、统计)
 经济统计 (经济、统计)
 金融统计 (金融、统计)
 生物统计 (生物、统计)
 工业统计 (工科、统计)
 旅游统计



什么是数理统计学？

数理统计学 (Mathematical Statistics)

- ◆ 对应西方的Statistics，加上mathematical这个字眼，以与在我们这里存在的被视为一门社会科学的统计学加以区别
- ◆ 关于统计学的数学理论部分（西方）

◆ 近代统计学



数理统计学的发展

◆20世纪以前的萌芽阶段

奠基人：

英国统计学家F. Galton (1822-1911) 初创生物统计学，提出回归分析方法。

英国统计学家K. Pearson(1857-1936)矩估计法

德国地测学家F. Helmer 1875发现 χ^2 分布（由阿贝(Abbe)于1863年首先提出？）

数理统计学的创立者



◆20世纪到第二次世界大战结束，成熟阶段奠基人：

- 英国统计学家R.A. Fisher(1890-1962) “女士品茶”
 - 1)正态分布下各种统计量的抽样分布;
 - 2)相关，回归分析和多元统计等分支的建立;
 - 3)最大似然估计法理论建立;
 - 4)创立试验设计，发展了方差分析。
- K. Pearson:检验拟合优度的 χ^2 统计量;
- W.S. Gosset: t分布;
- Neyman和E.S.Pearson: 1928-1938间假设检验理论
- J.Neyman:1934-1937间建立置信区间理论



◆ 战后时期，进一步发展阶段

计算机技术促进数理统计发展，战前的许多初步结果战后得到完善，许多相关学科崛起。如时间序列分析、随机过程等成为重要分支，贝叶斯思想主导的贝叶斯学派的崛起等。



统计学的地位？



一般来说，人们将统计分为两类：

➤ 描述性统计 (Descriptive Statistics)

运用制表和分类，图形以及计算概括性数据来描述数据特征，主要包括数据的频数分析、集中趋势分析、离散程度分析、分布以及一些基本的统计图形。

➤ 数理统计（推断统计） (Mathematical Statistics)

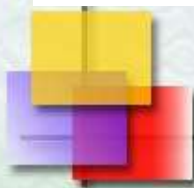
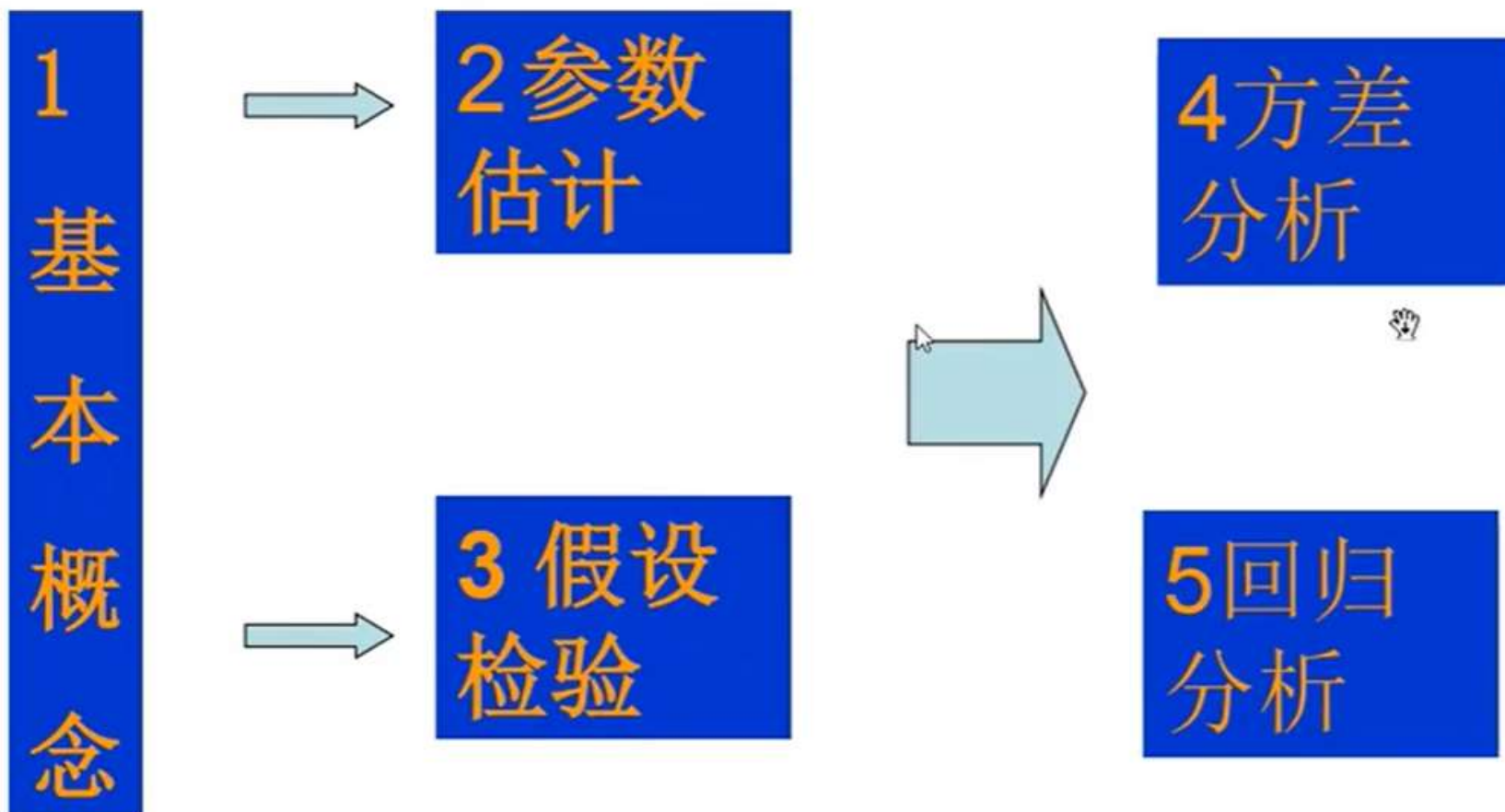
本课程主要研究数理统计。



数理统计是以概率论为理论基础，根据试验或观察得到的带有随机性的数据，来研究随机现象，对研究对象的客观规律性作出种种合理的估计和判断。

主要包括：如何有效收集、整理数据资料；如何对所得的数据资料进行量化分析、研究，从而对所研究的对象的性质、特点作出推断和预测。后者就是我们所说的统计推断问题。本书只讲述统计推断的基本内容。





例 某地环境保护法规定:倾入河流的废水中某种有毒物质的平均含量不得超过3ppm($1\text{ppm}=10^{-6}$)。该地区环保组织对某厂倾入河流的废水中该有毒物质含量连续进行20天测定,记录了20个数据(单位:ppm):

$$X_1, X_2, \dots, X_{20}$$

现要用这20个数据作如下统计推断:

- 该有毒物质含量 X 的分布是否为正态分布?
- 若是正态分布 $N(\mu, \sigma^2)$,其参数 μ 和 σ^2 如何估计?
- 对命题“ $\mu \leq 3.0$ ”(符合排放标准)作出判断:是或否。



常用软件

- (1) SPSS: 菜单方式和程序方式
- (2) SAS: 一切围绕编程设计
- (3) S-Plus: 是S语言(AT&T贝尔实验室)的后续发展,对编程能力要求极高
- (4) **R软件**: 一种免费软件, 是基于S语言的统计软件包, 运行的稳定性缺乏保证.
- (5) Minitab: 和SPSS非常相似, 是北美大学教学中的常用软件
- (6) Eviews: 这是一种计量经济学软件, 主要针对时间序列分析。
- (7) **Matlab**: 这是一种计算软件, 以编程为主。
- (8) Excel: 这是一种数据表格处理软件, 可作简单统计分析。
- (9) **Python**: 免费, 已经成为最受欢迎的程序设计语言之一

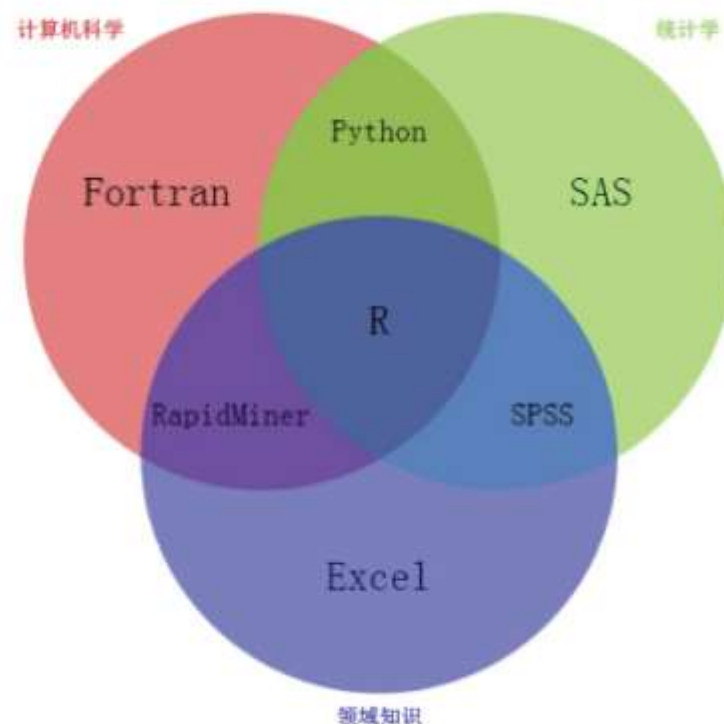


R vs. Python, ?

- Both are popular programming languages in data science

◦ R: 倚天剑

◦ Python: 屠龙刀



Employment: 2017年11月美国招聘岗位需求统计

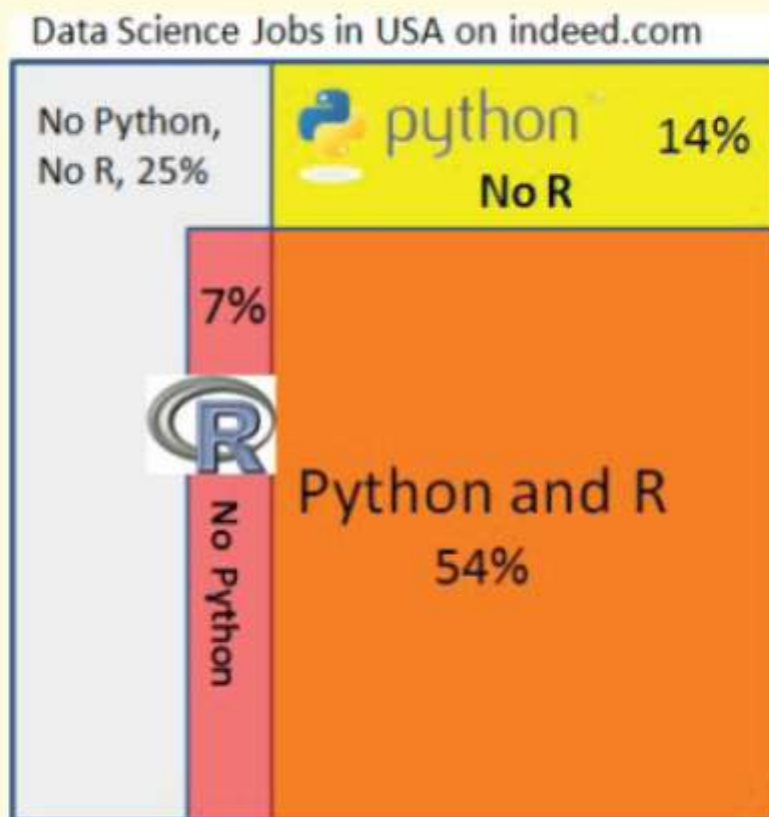


Figure 10: R vs Python: jobs

参考书推荐

1. 茆诗松，吕晓玲. 《数理统计学》. 中国人民大学出版社，2016（第2版）
2. 茆诗松等， 《概率论与数理统计教程》， 高等教育出版社
3. 陈希孺， 《数理统计引论》， 科学出版社， 1981
4. 陈希孺， 《数理统计简史》， 2002
5. 《女士品茶》
6. R语言
7. Python语言

.....

多看书、多思考、多实践



总评成绩：235 or 334

平时成绩：出勤+课堂表现+作业 请假要有假条

作业本要求：2本薄本子

抄作业扣分严重！



目标：好成绩+真知识

