



优达学城
UDACITY

业务中的描述统计学

商业数据分析 VIP 晋升班公开课



目录 CONTENT

01

数据与业务

工作中的描述统计学

02

03

统计和非事实

拓展书籍

04



01

数据与业务

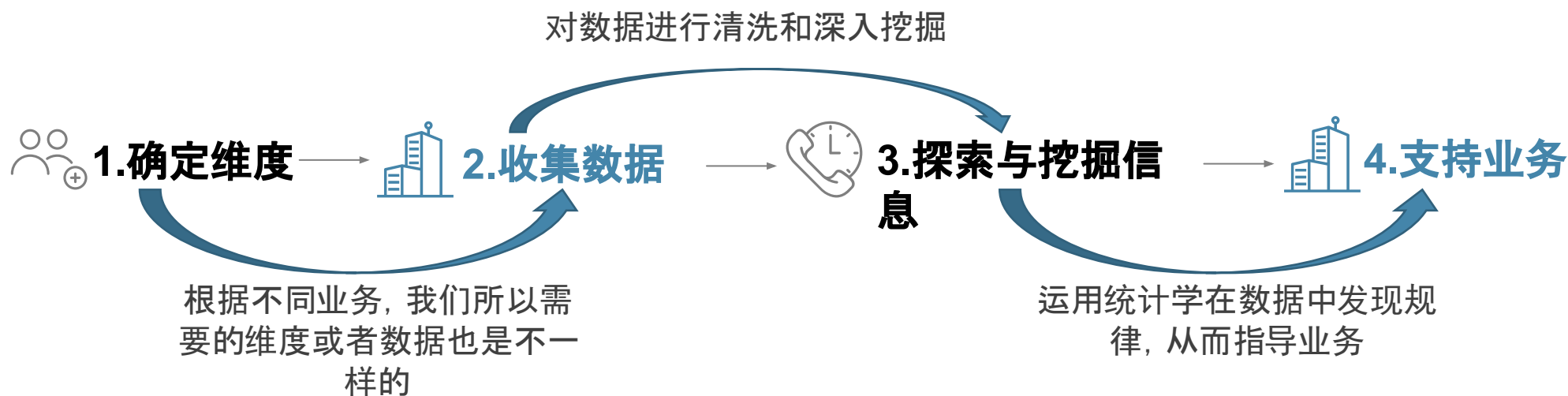
工作中会产生非常多的数据，我们需要收集并解读数据从而对我们的业务工作起到指导作用



01 数据与业务

什么是数据

数据是指不同的信息片段，不仅仅是表格上的简单数据，还有可能以多种多样的形式存在：





01 数据与业务

数据在业务中起到的作用：



普通员工：

市场：根据数据优化营销、投放策略

产品经理：根据数据优化产品，寻找用户痛点

中层管理：

使用数据进行表现、KPI的追踪，并统一绩效的统计口径

高层管理：

通过数据看整个公司的运营和管理；

展示商业回报，为融资做足数据准备等



02

工作中的描述统计学

数据在我们的工作和生活中无处不在，我们需要学习如何把握手中的数据，从数据中总结我们所需要的的指导方向



02 工作中的描述统计学

数据类型：

数值：	连续	离散
	身高、年龄、收入	书中的页数、院子里的树、咖啡店里的狗
分类：	定序	定类
	字母成绩等级、调查评级	性别、婚姻状况、早餐食品

- 了解手上的数据是什么类型，再根据不同类型的数据采取相应的分析方式或可视化图形
- 针对于数值型数据，我们一般分析其四个主要方面：**集中趋势**、**离散程度**、**形状**和**异常值**
- 对于分类数据一般探讨的比较少，分析方法多采用查看每个组的独立个体的数量或比例，用到比较多的是**众数**



02 工作中的描述统计学

集中趋势：平均数、中位数、众数

根据不同的业务场景，我们需要采用不同的值来描述手上的数据集



平均数

根据本月业务线日销售均值了解整体情况



02 工作中的描述统计学

集中趋势：均值、中位数、众数



中位数

部门领导根据组员的KPI完成程度指定下个月的KPI



02 工作中的描述统计学

集中趋势：均值、中位数、众数

根据不同的业务场景，我们需要采用不同的值来描述手上的数据集



众数

了解自己的用户中的核心用户是哪个年龄层级



02 工作中的描述统计学

离散程度：极差、四分位差(IQR)、标准差、方差

定义：每个观察值与均值之间的平均差异

标准差

最常见的数据离散程度度量之一



学到秃顶

公式

$$\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

方差

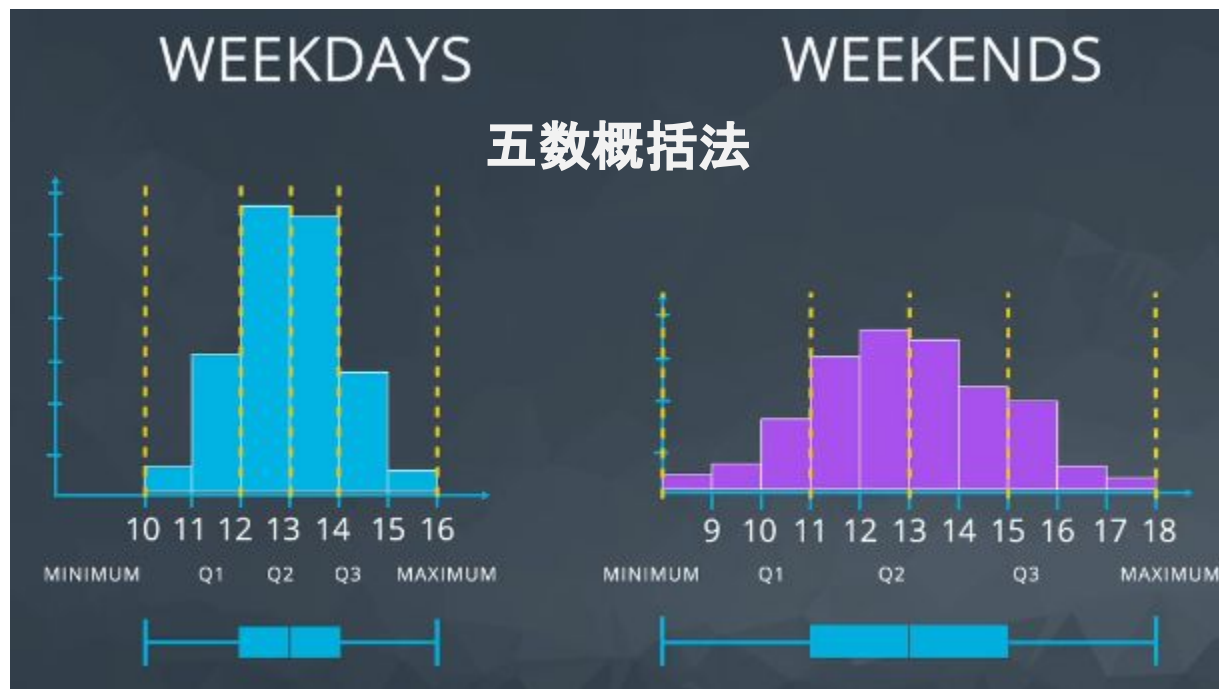
$$\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$

标准差



02 工作中的描述统计学

离散程度：极差、四分位差(IQR)、标准差、方差



标准差的好处

1. 只用一个数值比较两个数据集的离散程度
2. 具有推论统计学方面的优势
3. 具有与原始数据集相同的单位



02 工作中的描述统计学

举个栗子：

标准差越小越好

衡量组内人员对目前业务知识的掌握情况

(针对某些业务知识需要达成一致的理解, 比如一些审核规范)

通过对业务知识的小测, 收集到组内成员的成绩



举个栗子
FOR SAMPLE

集中

明确不清楚的业务点, 对症下药, 围绕展开培训

离散

大家掌握参差不齐, 很难找重点予以培训

进行质量对比

甚至可以对比培训前后标准差的大小, 查看离散程度有没有降低, 拉平大家对业务的水平



02 工作中的描述统计学

举个栗子：

标准差越大越好



举个栗子
FOR SAMPLE

公司内人员工资

对于相同岗位甚至不同岗位童鞋的工资，一般情况我们都希望其离散程度大一点比较好

相同职位

对于相同职位如果工资离散程度比较大，说明大家同工不同酬，大家凭实力说话，更容易提高积极性，拉开距离

不同职位

对于不同职位如果工资离散程度比较大说明了该公司职位很多，层级明显，侧面反映了公司的规模很大



03

统计和非事实

有的时候统计得出的结果和事实自我感觉并不一致，难道是统计出了问题么，这里我们来找具体例子探讨一下



03 统计和非事实

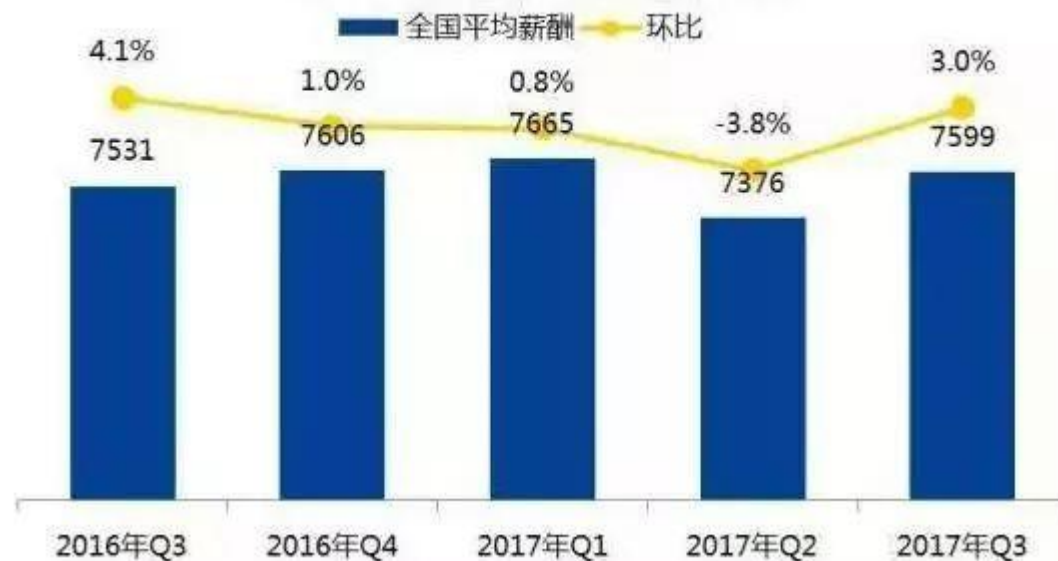
栗子1

我的工资被平均了：被平均的锅，平均数不背。

2017全国平均工资为7599元，相信大家的也应该是蒙的不行，这个平均数怕是不准哦



2017年全国平均薪酬比较



A房间有小明和姚明，平均身高1米8，小明要不要美滋滋？

极端值同学，教导主任想找你谈谈

03 统计和非事实

栗子2

为什么被抽样代表了:好的抽样,不是你想的那么简单。

采用抽样而不是使用全样本分析,因为有的样本量大到难以操作:了解全国学生的视力情况

- 样本量经过科学的计算,不是随口来说
- 抽样调查方法繁多,根据具体实例会采用不同方法
- 抽样调查的随机性,可以很大程度上反应总体情况
- 一般在95%的置信度上,我们能得到一个范围区间

抽样



普查



一个良心的抽样调查,你大可以放心被代表



03 统计和非事实

栗子3

美味的冰淇淋会导致溺水:严谨的统计结论只说相关性, 不说因果性

这个荒谬的结论是因为错误地将相关性当作因果性。夏天温度升高会带来“吃冰淇淋人数增多”和“游泳溺亡的人数”增多, 但并不代表二者之间有因果性。

- 统计学探究的多是相关性
- 因果性的探究需要用到更严谨的探究方法
- 尽量控制变量, 让探究更加严谨



优秀的数据分析师, 报告言辞严谨才显得专业



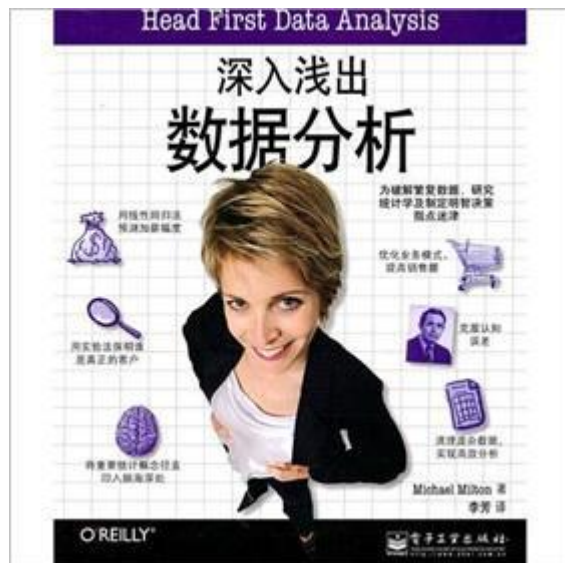
04

拓展书籍

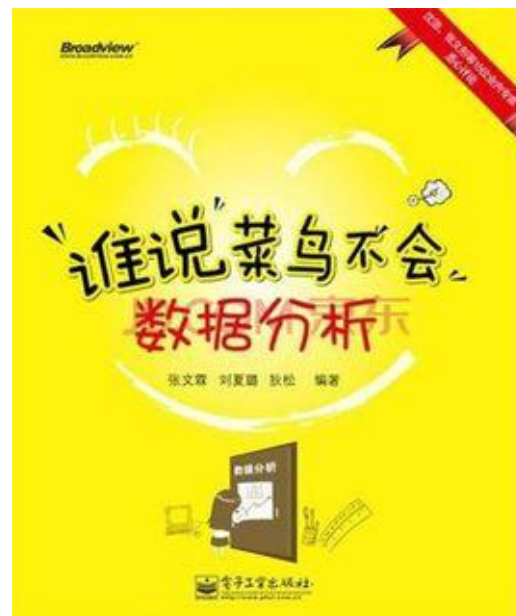
本章会罗列出可供学习的书籍以及资料



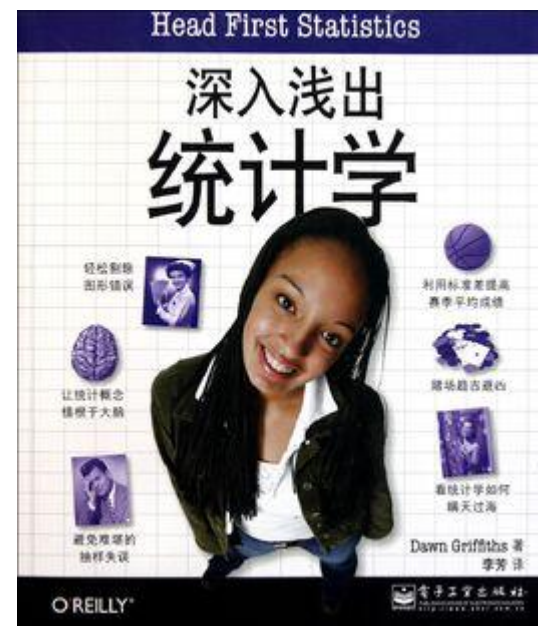
04 拓展书籍



《深入浅出数据分析》



《谁说菜鸟不会数据分析》



《深入浅出统计学》

互动环节

在你的工作场景中，是否用到描述统计数字呢？他们是如何帮助到你的工作呢？是否所有的场景都运用正确？





优达学城
UDACITY

Thank you for listening!

Q & A