

# 业务中的描述统计学 使用说明



<2019年1月17日更新 V1.0>

---

如果有更新迭代的建议，请发送邮件至 [kylie@udacity.com](mailto:kylie@udacity.com) 并抄送 [april.chen@udacity.com](mailto:april.chen@udacity.com)。谢谢。

## 0 概览

本公开课需要包含互动环节和 PPT 讲解环节：

### 【互动环节】

在讲解 PPT 之前，请先确保本公开课中的互动环节。为了确保课堂效果，助教可以先与班级里较为活跃的同学进行私下的沟通，鼓励分享与互动。为什么要进行互动的说服同学技巧：

1. 在商业分析工作中，我们需要有条理地向别人讲解自己的工作中的表现和数据成果，这是对个人能力的锻炼；
2. 可以听到其他同学的分享，与其他同学进行信息交流和思想交换；
3. 体现公开课的价值，不然公开课与视频课无差别。
4. ....（自由发挥）

请尽量确保本课程中同学全员参与到本互动环节。

### 【讲解部分】

PPT 主要包含三个部分 - 数据与业务、工作中的描述统计学、统计和非事实。

数据与业务：理解工作中数据的来源，以及会对我们有什么样的帮助。

工作中的描述统计学：在工作中，哪些场景我们会用到数据，数据会如何指导我们的工作；我们如何使用这些数据指导工作呢。

统计和非事实 - 用统计学的地方多种多样，哪些场景我们是用错了的。

请尽量按照时间分配和内容来讲解每部分内容。

## 1 互动环节 - 20 分钟

互动话题：

**“大家既然来学习数据分析，相信大家都是有分析数据的需求的。我们今天先来聊一下工作场景中的数据吧。大家的工作场景中会使用到哪些数据呢？这些数据都是什么样的形式/格式呢？相信大家已经通过课程了解到了描述统计学，大家也可以谈谈一些描述统计数字的使用。”**

同学分享建议包含：（需提前告诉同学进行思考）

1. 我工作中会用到哪些数据？
2. 这些数据的来源？是什么格式？
3. 这些数据会如何帮助到我的工作？
4. 在数据的处理和分析中，是否会用到统计学？用到了哪些？
5. 是否有描述统计学解决不了的部分？

同学分享完助教需进行点评，可以“描述统计学”课程进行点评。如果“有描述统计学解决不了的部分”，请助教老师给出建议。

（也可以将互动环节放到公开课部分 1 之后，建议不要放到最后，可能会影响同学的分享热情。）

## 2 公开课部分 - 30 分钟

同学的分享完毕。接下来进入公开课环节。

转折语：感谢大家的踊跃回答。我们本节课的课程内容并不难，主要目的在于帮助大家学会在工作场景中如何应用课程中学到的知识。**如果大家在本科期间学习过概率论和数理统计的话，相信大家对课程中的内容并不陌生，但是在工作中，我们是否知道该如何利用这些知识来指导我们的具体工作呢？**

（划重点：👉这是本课程目标，虽然不难，但是帮助大家扫清使用中的障碍，如果有问题同学可以随时提问）

### ● 数据和业务 - 10 分钟

请根据讲义，并结合自己的理解和同学前面的分享讲解。本部分主要讲解：

- 数据的常见来源
- 数据的类型

- 数据对于不同员工的作用和价值  
(可以结合同学的分享案例来讲解, 拉进与同学的亲密度。在每一种具体的数据来源、数据格式、以及业务中的场景, 都建议采用案例展开来讲, 使内容充实。)

## ● 工作中的描述统计学 - 14 分钟

这部分的内容与课程相关性比较大, 讲解方式, 建议先于同学进行概念回顾, 然后再结合具体的场景和案例进行展开。

数据的类型 - 3 分钟

- 数值(离散、连续)、分类(定序、定类)
- 举例子介绍这些类型。
- 工作中(页面访问人数 - 数值离散; 订单的金额 - 连续数值; 工作岗位的评级 - 分类定序; 产品品类 - 分类定类, 等等, 可自由拓展)
- 每种类型对应使用的分析统计值是什么

平均数 - 2 分钟

- 如何确定某天销售额的表现? 与昨天相比, 与前天相比, 还是与一周前相比? 我们可以通过与月销售均值(平均值)来客观地评定绩效。

中位数 - 2 分钟

- 再次以销售为例, 销售一般都是有任务考核指标(KPI)。如果你是销售部的主管, 会如何给员工制定绩效目标呢? 平均值或许可以, 但是如果有的员工特别特别优秀呢? 比如完成了一个大单。使用这个值会大大提高平均值, 在这种情况下, 我们使用中位数会比较好一些呢。

众数 - 2 分钟

- 上面我们讲的主要是连续数据, 假如是分类数据, 我们会较多地使用到众数。例如: 我们需要根据现有的用户找到我们的核心用户群, 来得到用户画像, 将来进行更精准的投放。我们应当如何哪种性别的用户最多? 哪个年龄段的用户最多? 哪种工作类型的用户最多? 这里利用了众数了解多数群体的特征。

离散程度 - 5 分钟

如何用一个数字表现两个数据集的波动情况? 离散程度可以有极差、四分位差、标准差、方差等不同的方式进行表现。

工作中比较常用的是标准差和四分位差、极差。方差与标准差的含义是很相近的, 不过由于标准差具有与原始数据集相同的单位使用更加普遍一些。

实际使用中, 离散程度/标准差是否越大越好呢? 这里我们来看几个例子🍷。(根据讲义和自身经验详细展开。)

## ● 统计与非事实 - 6-10 分钟

生活和工作中常见的统计误区有哪些? 请结合讲义和自己的理解讲解。

(可以提问同学: 是否有哪些地方用错了统计学? 使用的数据是否真实可信? 大家进行思考, 是否有案例可以分享呢? 点评: 如果用的不对, 如何处理会更好?)

### 3 资料参考 - 2 分钟

本小节的目的在于向同学推荐自己拓展的资源。除了下面中的一些拓展资源，也可以自由发挥。

深入浅出数据分析  
深入浅出统计学  
谁说菜鸟不会数据分析

### 4 Q & A - 5 分钟

同学的自由提问环节，为防止同学不够活跃，可以提前准备几个同学问到较多的常见问题进行讲解。

### 5 补充学习建议 - 1 分钟

推论统计学课程难度比描述统计学难度要大许多，学有余力的同学建议可以在学习本课程内容之后提前学习哦！

结语 - 大家如果仍然不知道如何运用描述统计学，请大家私下与我（助教）讨论。

日期	版本	修改人	修改原因
2019年1月17日	V1.0	Kylie	初稿