

STATISTICAL REASONING FOR EVERYDAY LIFE

妙趣横生的 统计学

培养大数据时代的统计思维

杰弗里·班尼特 (Jeffrey Bennett)

[美] 威廉·L.布里格斯 (William L. Briggs) ◎ 著

马里奥·F.特里奥拉 (Mario F. Triola)

徐 斌◎译

拒绝枯燥乏味的
理论与公式

揭示数据之下隐藏的
统计学之美

第四版
Fourth Edition

STATISTICAL REASONING FOR EVERYDAY LIFE

妙趣横生的 统计学

培养大数据时代的统计思维

杰弗里·班尼特 (Jeffrey Bennett)

[美] 威廉·L.布里格斯 (William L. Briggs) 著

马里奥·F.崔奥拉 (Mario F. Triola)

胡晖 徐斌 译

第四版
Fourth Edition

人民邮电出版社
北京

目 录

1

第一章 浅谈统计学

- 1.1 什么是统计 // 2
- 1.2 抽样 // 7
- 1.3 统计研究的类型 // 13
- 1.4 统计研究可信吗 // 20

2

第二章 统计测量

- 2.1 数据类型和测量尺度 // 30
- 2.2 误差处理 // 33
- 2.3 百分数在统计学中的应用 // 37
- 2.4 指数 // 41

3

第三章 数据的可视化展示

- 3.1 频数分布表 // 50
- 3.2 绘制数据分布图 // 53
- 3.3 媒体中的图表 // 63
- 3.4 关于图表的一些注意事项 // 70

4

第四章 描述数据

- 4.1 什么是平均 // 80
- 4.2 分布形态 // 87
- 4.3 离散程度的测量 // 91
- 4.4 统计悖论 // 101

5

第五章 正态的世界

- 5.1 什么是正态 // 112
- 5.2 正态分布的性质 // 117
- 5.3 中心极限定理 // 125

6

第六章 统计中的概率

- 6.1 概率在统计学中的作用：统计显著性 // 136
- 6.2 概率基础知识 // 138
- 6.3 大数定律 // 145
- 6.4 风险观念和预期寿命 // 150
- 6.5 组合事件的概率 // 155

7

第七章 相关性和因果关系

- 7.1 探索相关性 // 166
- 7.2 解释相关性 // 175
- 7.3 最佳拟合线和预测 // 181
- 7.4 因果关系分析 // 188

8

第八章 从样本到总体

- 8.1 抽样分布 // 196
- 8.2 估计总体均值 // 204
- 8.3 估计总体成数 // 210

9

第九章 假设检验

- 9.1 假设检验的理论基础 // 218
- 9.2 总体均值的假设检验 // 225
- 9.3 总体成数的假设检验 // 236

10

第十章 t 检验、列联表和方差分析

- 10.1 均值的 t 分布 // 246
- 10.2 列联表假设检验 // 252
- 10.3 方差分析 (单因素方差分析) // 260

附 录

- 附录 1 标准分数表 // 269
- 附录 2 随机数字表 // 271
- 附录 3 专业术语释义 // 272

第一章 浅谈统计学

你现在喝的水健康么？有多少人赞同总统的预算案？我们在医疗方面的投入是否得到了相应的回馈？上述问题只有通过统计学来进行研究。在第一章中，我们将讨论统计研究的基本概念，为后续章节更深入的统计研究奠定基础。在本书中，我们将结合大量的实际案例来论证统计学的精妙作用，即统计研究结果能为社会政策和私人决策提供方向上的指导，同时也讨论少部分案例以说明误用统计学将会导致错误的决策。

对于追求效率的公民而言，统计思维总有一天会和读写能力一样必要。

——H.G. 威尔斯

学习目标

1.1 什么是统计

理解统计的两个定义以及常用的基本概念，研究统计中总体、样本、样本统计量和总体参数之间的关系。

1.2 抽样

理解选择代表性样本的重要性，熟知几种常用的抽样方法。

1.3 统计研究的类型

了解观察研究和实验研究之间的区别；理解实验研究的重点注意事项，包括实验组和对照组的分配、安慰剂效应及盲法。

1.4 统计研究可信吗

评估在媒体中见到的统计研究方法，判断其统计研究结果是否有意义。

热点话题

聚焦心理学：你开车时沉迷于手机吗

聚焦公众健康：你的生活方式健康吗

>>1.1 什么是统计

统计学这门学科通常比较枯燥、一成不变或者专业性很强，但现代社会的方方面面几乎都与之相关。它可以告诉我们一种新研发的药品对治疗癌症是否有效；还能帮助农产品检验人员判断我们的食品是否安全；它还可以指导和解释民意调查；在商业上，我们运用统计方法进行市场调查和广告推广；在体育领域，统计学经常作为运动员和团队排名的方法。实际上，你会发现在某些重要领域几乎任何话题都与统计学有关。

本书最基本的目标是帮助读者学习一些统计方法背后的核心思想。值得称道的是，即使本书包含很少的统计理论，书中大量的案例也能帮助你理解在新闻、课堂、工作地点，或者在日常生活中所遇到的统计问题。

在统计信息中，虽然有时候把单词 **data** 看作 **information** 的同义词，但实际上它是复数。在只有一则信息的时候叫作 **datum**，当有两则或两则以上信息时才称作 **data**。

我们从统计自身的定义开始学习。统计这个单词在英语中用作单数和用作复数时，意思是不一样的。当被看作单数时，统计是帮助我们理解如何去收集、整理和分析一些话题中的数据或其他信息的科学方法，在这里把数和其他信息作为参考数据。当被看作复数时，统计是描述某些事物属性的实际数据。例如，你的班里有 30 个学生，他们的年龄是

17~64 岁。在这里，“30 个学生”“17 岁”和“64 岁”，从不同角度来说都是描述你所在班级某些属性的统计数字。

统计的两个定义

- 统计是对某一现象有关数据的搜集、整理和分析。
- 统计是描述、总结某种现象的数据或其他信息。

1.1.1 如何运用统计学

根据新闻报道，1.113 亿美国人看到纽约巨人队取得了超级碗橄榄球比赛的冠军，这就解释了，为什么一条 30 秒的商业广告的价格为 300 多万美元。那么是谁统计了这几亿人的数据？

统计学起源于一些国家事务，如人口普查数据和税金数据的采集。这就是为什么统计 (**statistics**) 这个词的词根是国家 (**state**)。

答案是没有统计这些数据。1.113 亿美国人观看超级碗橄榄球比赛的这个数据来自于一家名为尼尔森媒介研究 (**Nielsen Media Research**) 的调查公司的统计研究结果。这家公司仅仅通过监控 5 000 个住户的收视习惯，就编制出了著名的尼尔森收视率。

如果你是统计学的初学者，那么会认为尼尔森所得的结论更像是一个延伸。只研究几千个数据怎么能得出关于上亿人的结论？然而，统计科学表示，只要统计研究设计精准，这个结论就相当准确。下面我们以超级碗的尼尔森收视率为例，研究一些关键问题来论证通常情

况下的统计研究如何进行。

1. 统计研究的目标是什么

尼尔森的研究目标是确定美国观看超级碗橄榄球比赛的总人数。用统计学的专用术语表示,尼尔森所研究的所有美国人是一个总体;尼尔森所希望确定的数据,即观看超级碗橄榄球比赛的总人数,是这个总体的一个主要特征,在统计学中,这个总体特征被称为总体参数。

我们通常认为总体表示一群人,其实它可以是任意一个群体——人、动物或者事物。例如,在一个有关汽车安全的研究中,总体可以是在路上行驶的所有车辆。同样,总体参数这个术语,可以解释为总体的任何一个特性。在上例中,总体参数可以是某一时间段内所有行驶车辆的数量、行驶车辆的事故率或者行驶车辆的载重范围等。

总体: 在统计研究中,总体是所研究的人或事物的完整集合。

总体参数: 总体中对某变量的概括性描述。

例1 总体和总体参数

根据下列情况,描述研究的总体,并确定感兴趣的总体参数。

a. 你在农夫保险公司工作,公司要求你确定在没有侧面撞击安全气囊的车祸事故中,对受害者赔付的平均金额。

b. 你被麦当劳录用,确定每周用于炸薯条的土豆用量。

c. 你是美国基因泰克公司下属的一个商业记者,正在调查一个新的治疗办法对儿童白血病是否有成效。

答案 a. 总体包括所有的没有侧面撞击安全气囊的车祸事故中已经拿到保障金的受害者。相关的总体参数是对受害者赔付的平均金额(在之后的章节中,“average”将会被一个更准确的术语“mean”代替)。

b. 总体包括每周运输过来的用于炸薯条的土豆重量。相关的总体参数是土豆的平均重量,以及其重量的差异(例如,大多数每周用量是接近还是远高于平均值)。

c. 总体包含所有患白血病的儿童。其中重要的总体参数是没有用新治疗办法就痊愈的儿童百分比以及使用新治疗办法后痊愈的儿童百分比。

2. 实际研究的内容是什么

如果尼尔森的研究是无所不能的,它就能通过调查每个美国人来确定观看超级碗橄榄球比赛的人数。但现实中没有人能做到这样的全面调查,所以他们尝试研究相关的小集体来估计观看比赛的人数。尼尔森尝试通过监控美国人的一个较小样本群体,来研究所有观看比赛的美国人这个总体。尼尔森在大约5 000个家庭中安放了记录装置,所以这些家庭的人们就成为尼尔森所研究的美国人样本。

尼尔森从5 000个家庭中收集的个体测量值组成了原始数据。它收集了很多原始数据,比如每个家庭在什么时候收看电视,收看多久,播放什么节目,谁在看等。尼尔森把这些原始数据组合成一系列描述某个

亚瑟·C. 尼尔森(Arthur C. Nielsen)在1923年建立公司并发明市场调查。1942年,他推出了尼尔森广播指数为广播节目定级,并在20世纪60年代把该方法扩展到电视节目领域。该公司现在也监测其他媒介(互联网、智能手机等),并不断地改变方法来适应新的媒介技术。

样本特征的数据，比如样本中观看个别节目的观众百分比或者样本中观看超级碗橄榄球比赛的总人数。这些数据被称作样本统计量。

样本是总体的一个子集，它的数据是进行实际测量而获得的。

原始数据是对样本进行实际测量或观测所收集的数据。

样本统计量是描述从原始数据中筛选总结的样本特征的数据。

美国劳工部定义，一些没有工作的人并不是失业。例如，家庭主妇或者家庭主夫都不计入失业人群，除非他们在很积极地尝试寻找工作。那些试图寻找工作最后却因挫折而放弃了的人也不算在失业人群中。

例2 失业调查

美国劳工部把城市劳动力定义为那些已就业或正在积极寻找就业机会的人。每个月，劳工部会报告失业率，其是指在全部城市劳动力中积极寻找工作的人的比率。为确定失业率，劳工部调查了 60 000 个家庭，在失业率报告中，对以下术语进行描述：

a. 总体 b. 样本 c. 原始数据 d. 样本统计量 e. 总体参数

答案 a. 总体是美国劳工部想要研究的构成城市劳动力的群体。

b. 样本是指 60 000 个家庭调查中的所有人。

c. 原始数据是指调查中收集到的所有信息。

d. 样本统计量是对样本中原始数据的总结。在本例中，相关样本统计量是指积极寻找工作的人所占的百分比（劳工部同样也计算其他样本统计量，如青年、男性、女性和退伍军人的失业率）。

e. 总体参数是指与样本统计量对应的总体特征。在本例中，相关总体参数是实际失业率。需要注意的是，劳工部并没有实际测算总体参数，只是从样本中收集数据来估计总体参数。

3. 样本统计量如何与总体参数相关联

假设尼尔森调查发现 5 000 个家庭的样本中，有 31% 的人观看超级碗橄榄球比赛。“31%”是样本统计量，它是对样本的描述。但尼尔森真正想了解的是相应的总体参数，是观看超级碗比赛的人占有美国人的百分比。

尼尔森的调查者没有办法准确了解总体参数的数值，因为他们只研究了一个样本。然而，他们希望所做的工作可以保证样本统计量是总体参数很好的估计值。换句话说，他们希望得出结论，因为样本中 31% 的人观看超级碗，所以总体中也会有近 31% 的人观看。统计学的一个主要目标就是帮助调查者评估这类推论的有效性。

思考时刻

假设尼尔森的结论是 30% 的美国人观看超级碗橄榄球比赛。这表示有多少人观看比赛？（美国的总人口接近 3.1 亿。）

统计科学提供了一种方法，使得调查者能够确定样本统计量可以很好地估计总体参数。例如，调查或投票结果经常涉及误差幅度的概念。通过加减误差幅度，可以得到样本统计量的区间（即置信区间），它很可能包含总体参数。在大多数情况下，误差幅度被定义为该范围包含总体参数的置信度为 95%。我们将在第八章讨论有关“可能”和“95% 置信度”的精确定义。《纽约时报》（*The New York Times*）上有一个十分有用的解释（如图 1-1 所示）。在尼尔森案例中，误差幅度是 1%。因此，如果样本中 31% 的人观看超级碗橄榄球比赛，那么

总体中有 30%~32% 的人观看超级碗比赛的这个结论有 95% 的置信度。

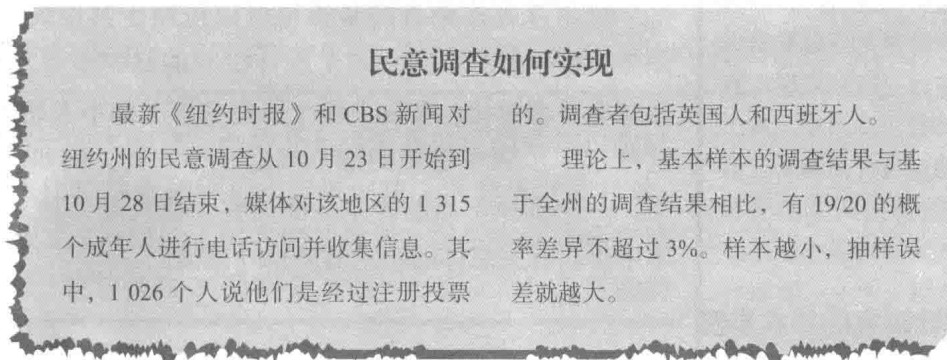


图 1-1 民意调查中的误差幅度

统计学最重要的一个标志性发现是，可以从非常小的样本中得到有意义的结果。然而，样本容量越大越好（当可行的时候），因为通常情况下，样本容量越大，误差幅度就越小。例如，在一个设计良好的投票选举中，在 95% 的置信区间内，当样本容量为 400 时，误差幅度通常为 5%；当样本容量为 1 000 时，误差幅度下降到 3%；而当样本容量为 10 000 时，误差幅度为 1%（参考第八章，了解怎样计算误差幅度）。

统计研究中的误差幅度描述了一个很可能包含总体参数的值域或置信区间。置信区间可以通过样本统计量加减误差幅度获得。也就是说，很可能包含总体参数的值的范围是：从（样本统计量 - 误差幅度）到（样本统计量 + 误差幅度）。

误差幅度通常给定 95% 的置信区间，这意味着在研究中有 95% 的样本会得出此区间将包含实际的总体参数（5% 的样本不会）。

例 3 政治性丑闻

皮尤人口和压力研究中心共计采访了 1 002 个美国成年人，调查政治性丑闻在最近的官方选举中有所增加的原因。57% 的被采访者声称，其增加的原因是丑闻都暴露在媒体更强有力的监督下，而 19% 的人则认为增加的原因是道德水准的下降。投票的误差幅度为 3%。请描述这次调查的总体和样本，并对 57% 的样本统计量作解释。我们如何推断出总体的百分比？即总体人群中“相信政治性丑闻的增加是由于媒体愈加强有力的曝光性”这部分人群的百分比。

答案 总体是所有成年的美国人，样本是 1 002 个被采访的美国成年人。样本统计量的 57% 是样本中实际回答“政治性丑闻的增加是由于媒体愈加强有力的曝光性”的百分比。根据 57% 的样本统计量和 3% 的误差幅度，可以确定取值范围：从 $57\% - 3\% = 54\%$ 到 $57\% + 3\% = 60\%$ 。这个范围很可能包含总体参数（95% 的置信区间）。

思考时刻

例 3 所描述的调查中，被采访者只在两个可能的答案——媒体高曝光率和日益下降的道德准则——中做选择。如果被采访者可以自己给出答案，你认为结果会改变吗？请解释。

统计学家通常把这个学科分为两个主要的分支：描述统计，即通过图表和样本统计量对原始数据进行处理；推断统计，即利用样本数据来推断总体特征。本书中，第二至第五章介绍描述统计，第六至第十章介绍推断统计。

1.1.2 总结：统计研究的步骤

尼尔森媒体调查的步骤也可以应用于其他统计研究中。图 1-2 和下面的框图总结了统计研究的基本步骤。注意这些步骤有些理想化，实际上每个研究的情况都不太相同。而且，这些基本步骤中隐藏的细节问题十分重要。例如，在步骤 2 中，一个选择不当的样本可能使整个研究毫无意义。从小样本中得出有关总体的结论需要格外谨慎。

统计研究的基本步骤

步骤 1：明确表述研究的目标。确定想研究的群体以及想了解的内容。

步骤 2：从总体中选择有代表性的样本。

步骤 3：从样本中收集原始数据，通过计算感兴趣的样本统计量来总结数据。

步骤 4：使用样本统计量对总体进行推断。

步骤 5：得出结论，确定是否达到目标。

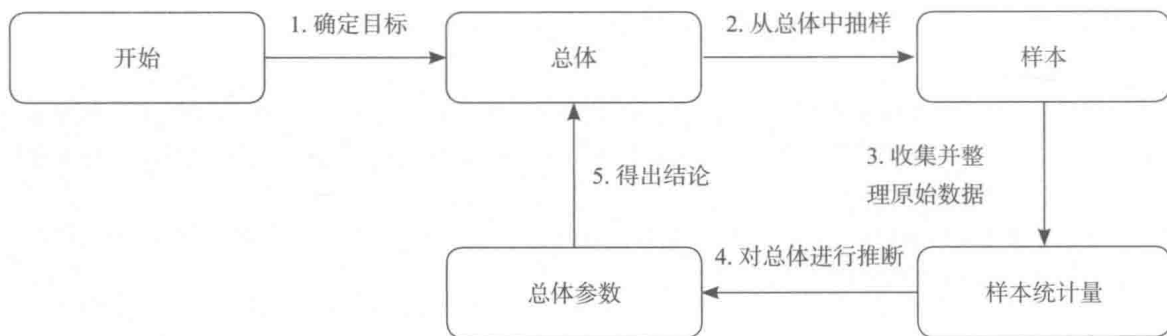


图 1-2 统计研究的基本步骤

例 4 确定步骤

确定例 3 中调查者进行统计研究的五个步骤。

答案 a. 调查者以“美国人如何看待最近政治性丑闻增加的原因”为目标，以所有美国成年人为总体，谨慎地将儿童排除在外。

b. 选择 1 002 个美国成年人作为样本。虽然并未说明样本是如何选取的，但我们可以假设这 1 002 个美国成年人包含总体美国人的所有类型。

c. 调查者通过投票收集原始数据。所有的原始数据都是独立的回答。通过计算样本统计量总结归纳数据，如每个问题中回答是或否的人数所占的百分比。

d. 调查者可以通过统计技术推断总体特征。在本例中，包含推断相关总体参数、计算误差幅度。

e. 调查者通过计算确保正确研究和解释估计的总体参数，得出有关全体美国人对近期丑闻态度的结论。

1.1.3 统计：在不确定中做出决策

到目前为止，我们讨论的大多数案例涵盖了调查和投票，而统计学所包含的内容不仅仅是这些，比如用实验检验新药物的治疗效果，分析全球变暖的危害，或者评估大学教育的价

值。实际上，统计学的主要目标是帮助我们在面对很多可能的选择时做出恰当的决策。

统计学的目的

统计学有很多用处，而它最重要的目的是帮助我们在面对不确定的问题时做出决策。

我们讨论的大多数案例都基于以上目的，但偶尔也会讨论一些看似比较抽象的理论。如果记住了所有统计的目的，你最终会通过这些理论了解我们的世界。以下案例将会带给你一些启示，它包括在 20 世纪公共卫生方面取得巨大成就的一些重要理论。



案例研究 索尔克脊髓灰质炎疫苗

在 20 世纪四五十年代，如果你是一位家长，你最害怕的就是小儿麻痹这种疾病。在这个小儿麻痹肆虐的时期，每年都会有数以千计的儿童因为这个病而瘫痪。1954 年，乔纳斯·索尔克（Jonas Salk, 1914—1995 年）医生对新研制出的疫苗进行有效性实验。实验的样本为从美国所有儿童中选出的 40 万名儿童，其中的一半儿童注射了索尔克疫苗，而另一半只注射了生理盐水（生理盐水是安慰剂，见 1.3 节）。在注射索尔克疫苗的儿童中，只有 33 人患了小儿麻痹症。而未注射疫苗的儿童中，有 115 人患上了小儿麻痹症。通过统计分析，调查者认为疫苗对抑制小儿麻痹症有效。因此他们决定大力支持并改进索尔克疫苗，并为所有儿童注射。感谢索尔克疫苗，让恐怖的小儿麻痹症成为过去的一个回忆。

在研发了索尔克疫苗后，小儿麻痹症在美国迅速减少，但在一些不发达国家这种病还是很常见。1998 年，全球开始大力支持为儿童注射索尔克疫苗。尽管现在还没有达到完全根除这个疾病的目标，但对小儿麻痹症的抑制已经非常成功了。

工作的最大报酬，乃是获得可以做得更多的机会。

——乔纳斯·索尔克

>>1.2 抽样

我们仅能依靠观察总体中的每个成员来了解总体参数的真值。例如，想知道你所在学校学生的精确平均身高，你需要对每个同学进行测量。这种对总体中每个成员进行数据收集的方法称为普查。然而，进行一次普查通常很不切实际。在某些情况下，总体太大，收集每个成员的数据既耗时又费力。在另一些情况下要排除普查，因为它会干涉研究目标。例如，涉及检验销售前的棒棒糖的质量，我们不能运用普查的形式，因为那样的话，每个棒棒糖都要检验，就没有完整的棒棒糖可以出售了。

不是所有的价值都能被计算，不是所有能计算的都具有价值。

——阿尔伯特·爱因斯坦

(Albert Einstein)

普查是指对总体中所有样品进行数据收集。

大多数统计研究都可以不进行普查。一般情况下，我们从样本中收集数据（而不是对总

体中的每个个体进行收集), 并通过样本统计量推断总体参数。当然, 仅当样本中的个体对总体具有代表性或依据研究的特性进行数据选取时, 这个推断才是有意义的。也就是说, 我们要寻找总体的代表性样本。

代表性样本是具有总体本质属性的样本。

例 1 有关身高的代表性样本

假设你要统计你所在学校全部学生的平均身高, 以下哪个更可能是该研究的代表性样本? 男子篮球队成员还是你统计课班级里的同学?

答案 男子篮球队不是研究身高的代表性样本, 因为它只包含男生且男子篮球队成员一般比同龄人高。你所在统计课班级的同学的平均身高更接近所有同学的平均身高, 所以你所在统计课班级里的同学比男子篮球队成员这个样本更具有代表性。

许多药物研究是测试新研发药物是否有效的。美国医学协会杂志发表过一篇文章, 其中表明研究结果为药物有效的论文比药物无效的论文更容易发表。这个公共偏差使得所发布的药效结果比实际药效更容易让人相信。

1.2.1 偏差

假设在电视收视率案例的 5 000 个家庭样本中, 尼尔森仅选择常常夜间轮班的工薪家庭。它会发现样本中的家庭很少观看深夜节目(因为上夜班的人不能在家里观看深夜节目)。很明显, 这个样本不能代表全部的美国家庭, 据此得出“深夜节目在美国人中不受欢迎”的结论是错误的。我们说这个样本是有偏差的, 因为它并不包含所有典型的美国家庭(在

实际的样本选择中, 尼尔森尽力避免这些明显的偏差)。偏差在设计和进行统计研究时会引起许多问题, 并使结果趋向于我们偏好的特定结果。我们不能相信一个有偏差的研究结果。

如果统计研究的设计和实施趋向于一个偏好的特定结果, 那么它可能存在偏差。

偏差产生的途径有如下几种。

- 如果样本中的个体在某些方面与总体的个体有区别, 则说明该样本存在偏差。在这种情况下, 研究结果将反映样本的特有性质, 而不是总体的普遍特征。
- 如果研究者倾向于某一特定结果, 则会产生偏差。在这种情况下, 研究者会有意无意地曲解数据含义。
- 如果在收集数据时, 有意无意地选择了不具有总体代表性的数据, 则说明数据设定本身存在偏差。
- 即使研究已经完成, 也可能存在报告偏差。例如, 一个数据图表可能只反映了部分事实, 或歪曲了数据(见 3.4 节)。

减少偏差是统计研究中的一项巨大挑战。因此, 寻找偏差成为评估统计研究或媒体报道的重要步骤。

例2 为什么使用尼尔森数据

尼尔森媒体研究通过向电视台和网站收取服务费来赚钱。例如，NBC 向尼尔森公司支付费用来获取电视节目收视率数据。为什么 NBC 自己不统计数据，而是购买另一家公司的数据呢？

答案 电视节目所能赚取的广告费，取决于广告的收视率。收视率越高，广告费就越高。这意味着如果 NBC 自己统计其收视率，将会存在很大的偏差。因此，广告客户不会相信 NBC 本公司统计的收视率。通过租用相对独立的数据资源，如尼尔森数据，NBC 公司可以提供让广告客户更能信任的收视率信息。

思考时刻

实际上 NBC 购买尼尔森公司数据的同时，也给尼尔森一个机会让 NBC 的收视率看起来还不错。如果你是广告代理，尼尔森收视率还有其他别的因素能帮助你获得信心吗？

1.2.2 抽样方法

一个好的统计研究必须具有一个代表性样本，否则样本存在偏差会导致结论毫无意义。我们依次检验一些普通的抽样方法，理论上至少能筛选出一个代表性样本。

1. 简单随机抽样

在众多案例中，获得代表性样本的最好方法是在总体中进行简单随机抽样。随机抽样使总体中的每一个对象都有平等的机会被选入样本。例如，可以通过掷骰子选择总体中掷到 6 的人作为一个随机抽取的样本。但选择高于 6 英尺^①的人作为样本并不是随机抽样，因为每个人被选择的机会不同。

在统计中，我们预先决定所需的样本容量。在简单随机抽样中，具有特定样本容量的可行样本都有均等的被选择的机会。例如，从你所在学校的全部学生中用简单随机抽样选取 100 个学生作为样本，你可以给学校中的每个同学编号，然后抽签选出 100 个人，只要确保每个学生的编号是唯一的，那么由 100 个学生组成的每一个可能样本被选中的机会就相同。你也可以通过计算机或者有内置随机数字生成器的计算器选择学生编号。

思考时刻

寻找计算器上的随机数字键（几乎每个科学计算器都有）。当你按下它时会发生什么？如何用随机数字键选择 100 个学生的样本？

因为对于特定容量的样本，简单随机抽样使每个样本有均等的被抽取机会，所以只要样本容量足够大，就能找到具有代表性的样本。

例3 当地居民抽样

你想在城镇的居民中进行投票选举，能否从当地的财产税收记录中通过简单随机抽样抽取名字？

答案 来自财产税收记录的样本不是该城镇总体人口的简单随机抽样样本，因为记录中只包含该城镇有财产的人，它不包括这个城镇的所有人，它可能还包括不在该城镇生活但在这里有不动产的人。

① 1 英尺 = 0.3048 米。——译者注

2. 系统抽样

在一些案例中，尽管简单随机抽样十分有效，但是我们可以通过更简单的方法获得同样的结果。假设你要检验英特尔公司生产的微机芯片的质量。由于芯片靠流水作业线产出，你可以每隔 50 个抽取一个芯片作为样本。此样本是一个代表性样本，因为每隔 50 个抽取的芯片与其他芯片相比没有任何特殊的规律特征。这种形式的抽样方法被称作系统抽样。

例 4 博物馆评估

美国航空航天博物馆想策划一个新的太阳系展览，工作人员对一个系统抽样的参观者样本进行采访调查。他选择每 15 分钟正好进入参观太阳系展览的人作为系统抽样的样本。他为什么使用系统抽样的方法而不是简单随机抽样来确定样本呢？在该案例中，系统调查更能抽取到有代表性的样本吗？

答案 简单随机抽样偶尔会间隔很短地选择两个参观者，这样的话，工作人员没有足够的时间去采访他们。而系统抽样每隔 15 分钟抽选一个参观者，很好地避免了这个问题。因为参观者在特定的时间进入，与早一点或晚一点进入没有任何不同，所以我们可以通过系统抽样得到有关总体的代表性样本。

例 5 什么情况下系统抽样会失败

你准备调查男女混住的宿舍，其中男生被分配在奇数号房间，女生被分配在偶数号房间。当你每隔 10 个房间抽取 1 个房间时，能获得代表性样本吗？

答案 不能。如果你从奇数号房间开始抽取，每隔 10 个抽取的房间依旧是奇数号（比如房间号 3, 13, 23, …）。同样地，如果你从偶数号房间开始每隔 10 个抽取一个房间，它依然是偶数号房间。因此，你获得的样本要么全是男生，要么全是女生，不具有男女混合的代表性。

思考时刻

在例 5 中，如果你每隔 5 个房间抽取 1 个，而不是每隔 10 个抽取，所得的样本是否具有代表性？

3. 任意抽样

系统抽样虽然比简单随机抽样简单，但在某些情况下仍然不切实际。假设你想知道你在学校惯用左手学生的比例，你将花费很大的精力去进行简单随机抽样或系统抽样，因为两者都要考虑到你所在学校的所有学生。然而，以你统计课班级的同学作为样本就会变得容易得多，你只需让班里惯用左手的同学举手以作统计。这种抽样方式叫作任意抽样。为了确定惯用左手学生的比例，以你所在的统计班级作为任意样本是不错的选择。因为该班级里惯用左手学生的比例和整个学校范围内惯用左手学生所占的比例没有什么不同。但如果你是研究不同专业同学的比例，该样本就会存在偏差，因为一些专业不需要学统计。总的来说，任意抽样比其他形式的抽样调查更容易出现统计偏差。

例 6 沙拉酱口味测试

超市在考虑是否要引进一种新品牌的沙拉酱，所以提供免费试吃，以收集顾客意见。这里使用了哪种抽样方法？这个样本对总体购买者来说具有代表性吗？

答案 顾客试吃沙拉酱的样本是一个任意样本，因为这些顾客刚好在这个店里，并愿意尝试新产品（这种类型的任意样本是人们自己选择是否成为该样本的一部分，也称为自主选择样本，我们将在 1.4 节中详细介绍）。该样本不能准确地代表顾客总体，因为不同的人会在不同的时间进行购物。例如，在家休息的人通常比上班的人更趋向于在中午购物，并且仅仅是喜欢吃沙拉酱的人才有可能试吃。当然，数据仍然很有用。因为在这个调查中，喜欢吃沙拉酱的顾客的意见才最重要。

4. 整群抽样

整群抽样是指将总体中的个体归类为随机的组或群。假设你在农业部门工作，并且想知道使用有机农业技术的农民所占的比例。如果用简单随机抽样或系统抽样去选取样本，会耗时耗力，因为需要去许多相距甚远的独立农场采集数据。使用某个县的农民作为任意样本又会出现偏差，因为不同地区的农业生产方式会有很大差异。你可以在美国范围内随机选择几十个县，对其中的每个农民进行调查。我们认为每个县都是一个农民群体，样本包含了随机选取的群体中的农民。

例 7 汽油价格

你想知道离机场出租车站 1 英里^① 远的加油站的汽油平均价格。说明在本案例中使用的抽样调查方法。

答案 你可以在全美随机选取几个机场，调查离机场出租车站 1 英里远的所有加油站的油价。

5. 分层抽样

假如你在策划一个预测下届美国总统选举结果的投票。研究总体为所有可能的投票者，你可以从中选择一个简单随机抽样样本。然而，由于总统选举以州为单位计算选举结果，如果你知道每个州投票者的偏好，就可以更准确地预测结果。因此，样本应该包括 50 个州中每个州各自的随机样本。在统计术语中，50 个州代表不同子组或层级。因为样本是从每个层级中随机选择得到的，所以该方法叫作分层抽样。

例 8 失业数据

美国劳工部每月调查 60 000 个家庭来编制失业率报告（见 1.1 节，例 2）。劳工部首先将县和市分成 2 000 个不同的地区，然后在这些地区随机选择一些家庭进行调查。这是分层抽样吗？它的层级是什么？为什么在本案例中分层抽样很重要？

答案 失业率调查是分层抽样的一个案例，因为它首先将总体分成各子组。子组或层级是 2 000 个地区的人。分层样本在本案例中十分重要是因为地区不同，其失业率也不同。例如，堪萨斯州农村和硅谷的失业率会存在很大差距。通过分层抽样，劳工部可以保证样本公平地代表所有地区。

1.2.3 抽样方法总结

图 1-3 总结了之前讨论的五种抽样方法。没有单独的哪种方法是最好的，每种方法都有其使用范围（某些研究甚至采用两种或者多种抽样调查方法）。但不管怎样抽选样本，其一定要符合以下三点：

- 只有样本对总体具有代表性，研究才有意义；
- 有偏差的样本不可能是代表性样本；
- 由于在实际选择样本时可能运气不好，即使是一个精心选择的样本也可能不具有代表性。

^① 1 英里 = 1 609.344 米。——译者注

常用的抽样方法

简单随机抽样：选取的是容量相同且每个样本有同等概率被选择的样本。

系统抽样：通过简单的系统抽取样本，比如在总体中每隔 10 个或每隔 50 个抽取样本。

任意抽样：使用一个碰巧很容易被选择的样本。

整群抽样：先将总体分成不同的组或群，然后从中随机挑选几个群作为样本。

分层抽样：当我们注重总体中不同的子组或层级时，选择该方法进行抽样。首先定义层级，然后在每个层级中随机抽取样本。总样本包括所有从每个独立层中抽取的样本。

**简单随机抽样：**

每个容量相等的样本都有同等被抽取的机会，计算机常被用于形成随机的数字

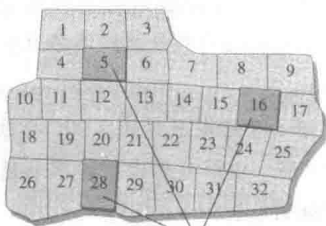
**系统抽样：**

依次选取第 k 个成员

**任意抽样：**

使用方便现成的结果

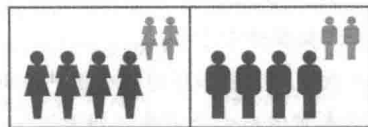
卡森县的选举地区



采访阴影地区的所有投票者

整群抽样：

将总体分为不同组或群，随机选择其中的一些群，选择所选群中的所有个体作为样本

**分层抽样：**

将总体至少分为两个层级，从每个层级中抽取样本

图 1-3 常见的抽样方法

例 9 抽样方法

辨别下列案例中所使用的抽样方法。

- 某一果园采摘了 1 200 篮苹果。一位农业检查员随机抽取 25 篮，检查其中的每个苹果是否有虫子。
- 一个教育调查者想知道在某个大学中，男生还是女生在课上问的问题多。在这所大学的 10 000 名学生中，随机抽取了 50 名男生和 50 名女生。
- 为了研究行星系，天文学家进行了一项在 100 颗邻近的星球中找寻行星的调查。
- 为了决定谁将获得球星亲笔签名的足球，使用一个计算机程序从体育场的观众中随机选择 11 个人的门票号码。

答案 a. 对苹果的检查是整群抽样调查，因为调查者最初随机选择了几个组群（篮子），之后检查所选群中每个苹果的质量。

b. 男生和女生代表研究中两个不同的层，所以这是一个分层抽样的案例。