
绪 言

◆



我岳父刚从爱荷华州到加利福尼亚州便对我说：“你们这儿治安不好。”在他所阅读的报纸中的确是这样描述的。但是，这种报纸以不轻易忽略当地发生的任何犯罪行为，并且以对于爱荷华州发生的谋杀案用比爱荷华州本地的主要报纸更多的笔墨进行大肆渲染而闻名。

我岳父得出的这个结论是一个建立在明显有偏样本基础之上的随意的统计结论。类似于其他某些更为精致的统计结论，它也存在着证据与结果不匹配的问题。它假定报纸专栏中对犯罪行为的报道是测量犯罪率的工具。

几年前，十来个调查人员分别独立地发表了关于抗组织胺药物的数据。所有的数据都证明，在经过抗组织胺药物治疗后相当大比例的感冒能够治愈。这引起了传媒的大肆宣传，至少在广告中也是如此，并兴起了药物生产的热潮，这是因为人们长期以来对药物具有强烈的需求。但奇怪的是，人们拒绝越过统计去注意一下早就了解的事实。正如一位幽默的非医学权威人士，亨利·G. 菲尔森（Henry G. Felsen）所指出的那样，服用上述药物的确能在一个星期内治愈感冒，但人们却忽略了另一个事实——即使不服用上述药物，一般而言，感冒也能在七天内痊愈。

你所见所闻的许多东西也和上面的事例一样。平均

数、作用关系、趋势和图表与看上去的并不总是一致。虽然经验告诉我们“眼见为实”，但眼睛告诉我们的也许隐瞒了或者夸大了一部分事实。

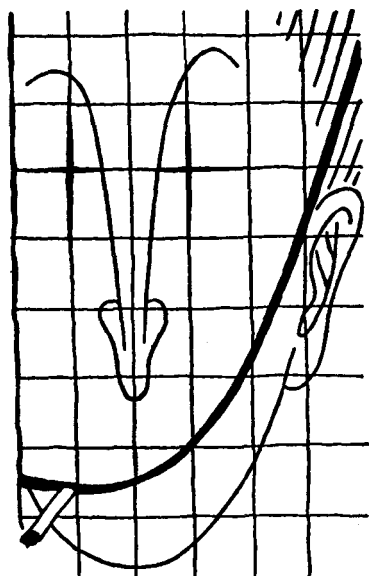
统计这种神秘的语言，在一个用事实说话的社会里是如此的吸引人，但有时它却被利用并成为耸人听闻、恶意夸大或简化事实、迷惑他人的工具。在报告社会经济趋势、商业状况、民意测验和普查的大量数据时，统计方法或者统计术语是必不可少的。但如果作者不能正确理解并恰当地使用这些统计语言，而读者又并不能真正懂得这些术语的含义，那么，统计结果只能是一堆废话。



在通常的科学写作中充斥着被滥用的统计数据，以至于人们很难想到这些统计数据的背后有一个“白衣侠

客”在昏暗的灯光下没日没夜地工作，就像一小撮药粉、一小瓶颜料，统计把一些重要的事实弄得似是而非。巧妙伪装的统计资料胜过希特勒的弥天大谎，虽然它也会引起人们的误解，却不会因此受到指责。

本书是一本如何利用统计以达到瞒天过海目的的入门读物。看上去，它很像骗子的行骗指南。但或许我可以模仿一个已退休的窃贼——在如何神不知、鬼不觉地撬开一把锁的方面，他的回忆录达到了研究生课程的水平——替这本书说句公道话 毕竟 骗子早就掌握了行骗的技巧，而诚实的人出于自卫也应该学会它。



第一章

内在有偏的样本



“1924 级的耶鲁毕业生平均年收入有 25 111 美元。”《时代》(Time 杂志在评论纽约《太阳》(Sun 报的某篇报道时曾经指出。

哇，他们干得真不赖！

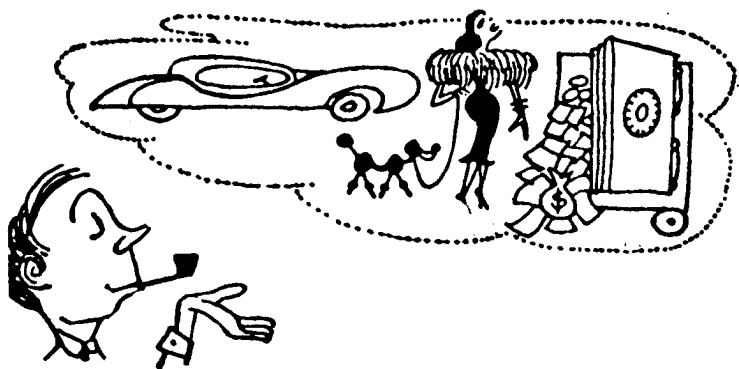
可是，等一等，这个令人印象深刻的数据到底意味着什么？是否像表面看到的那样，足以证明如果你把你的孩子送进耶鲁大学，那么在老年时你就不用辛苦地上班，甚至他年老时也不用上班？

在充满怀疑的惊鸿一瞥后，我们发现这个数据的两点可疑之处 它惊人地精确 它大得令人难以置信。

任何一群分布很广的人其平均收入能精确到以元为单位都不太可能。就算是自己去年的收入，除非全部来自薪水，否则也很难知道得如此准确。但对于年收入 25 000 美元的阶层而言，多种投资渠道使得收入不可能完全来自于薪水。

而且毫无疑问，这个可爱的平均数出自耶鲁人之口。即使 1924 年他们在纽哈芬接受过良好的教育，也很难保证四分之一个世纪后，他们还能坚持说真话。当问及收入时，有些人出于虚荣或天生乐观而夸大数据；有些人却故意缩小数字，特别当涉及计征所得税问题时，往往会犹豫不决 生怕与其他文件填报的数据不符 谁知道税务员又看到了什么？也许这两种趋势——夸大与缩小将相互

抵消 但这种可能性极小。一般而言，一种趋势总会强于另一种趋势，但我们无从猜测哪种趋势较强。

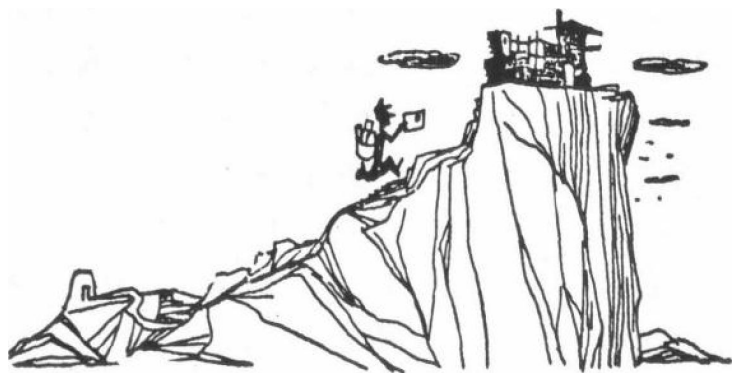


常识告诉我们单凭某一数据很难反映实情，这是我们得到的结论。那些实际收入也许只有 25 111 美元一半的人们最终会形成如此高的平均收入，最大的误差来源在哪儿 接下来 让我们来揭开这神秘的面纱。

这是一个抽样过程。在你所遇到各种各样的课题中，大部分统计问题的核心便是抽样。抽样的原理本身很简单 但实践中对其进行的加工导致了许多副产品 有些是不正确的。举个例子，如果你有一桶豆子，有红色有白色 那么 红色的豆子到底有多少呢 解决的办法只有一种 数豆子。然而 用一种更简单的方法 你也可以得到红豆数目的近似结果 抓一把豆子 计算其中红豆的比例，这把豆子中红豆的比例与一桶的比例基本相同。

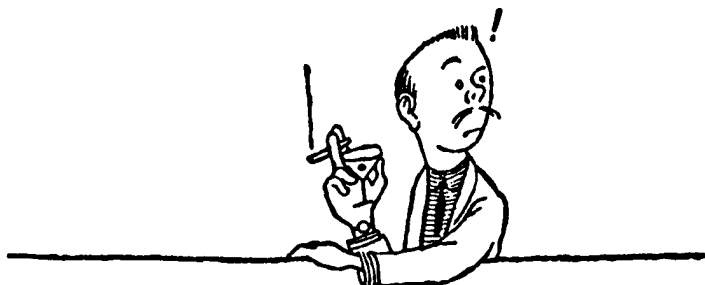
只要样本足够大 并且具有代表性 在多数情况下 样本的信息可以很好地代表总体。但如果条件不满足，抽样的结果比一个臆想的结果好不到哪儿去——除了会形成一种十分科学精确的虚假印象以外，这种方法根本不值得推荐。然而 令人悲伤的事实是 在我们看到的或自以为知道的事物中 充斥着许多从类似样本 即有偏的或过小的或两者都满足的样本得出的结论。

耶鲁毕业生的收入也来自于样本。出于理智，我们可以很肯定地认为，没有人能够掌握所有仍活着的 1924 级学生的情况 25 年后，许多人已经地址不详。



并且 在能够取得联系那些人中 许多人根本不会回答问卷，特别是一个内容涉及隐私的问卷。一般的问卷 5%~10% 的回收率已经是相当可观了。也许这个调查的回收率会高些，但不可能达到 100%。

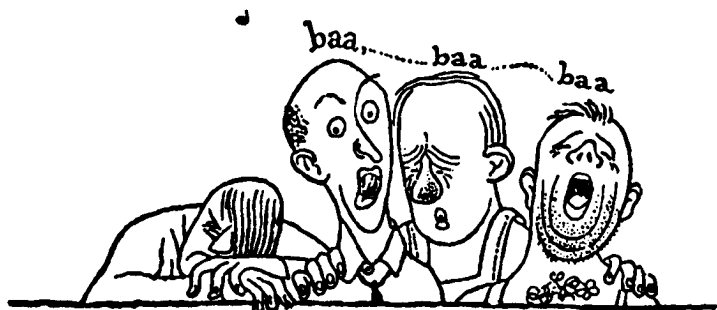
因此，这个收入数据是建立在一个由能够取得联系并愿意回答问卷的耶鲁学生组成的样本之上的。那么，这个样本具有代表性吗？也就是说，能否假设这个样本与样本未被包括的人具有同样的收入水平？



那些在耶鲁大学毕业生通讯录上被注明“地址不详”的迷路小羔羊是谁呢？他们是高收入阶层吗？他们是华尔街的金融家、公司领导层，抑或是制造企业或公用事业的执行总裁？不，富人的地址是不难找到的。这个班级最富有的人，即使忽略了与校友办公室联系，他们的地址也可以通过查《美国名人录》(*Who's Who in America*)或其他参考资料找到。因此，我们可以较合理地猜测，那些被遗漏的人是获取耶鲁文学学士学位之后的25年来没能实现自己光辉梦想的人，他们是小职员、技工、流浪者、失业的酒鬼、仅仅得以糊口的作家或艺术家……六七个甚至更多这样的人将其收入相加才可能达到25 111美

元 他们不经常在班级的联谊会上注册 仅仅因为他们支付不起路费。

我们是可怜的快速的小羔羊



又是谁会将调查问卷丢进最近的废纸篓？我们不太肯定，但是猜想这种人中大部分都没有赚到足以自夸的数目是合理的。这有些类似于第一次领取薪水的小职员，当他发现薪金条上粘着一张建议对同事保密工资的数额并不要以此作为谈资的小纸条时的心态“别担心，”他对老板说“我与你一样 对这么低的工资感到羞愧。”

很明显，样本遗漏了对平均收入起降低作用的两种人。现在我们可以了解 25 111美元的庐山真面目了，如果它是一个真实的数据 它也仅仅代表了 1924 级耶鲁学生中可以联系到的，并愿意站出来说出所赚数目的这个特殊群体。当然，它的真实性还建立在这样一个假定基础之上——这些绅士说的都是真话。

是否能过于轻率地做出这样的假定呢？来自抽样理论的一个分支——即市场研究的经验告诉我们——人们会说真话的假定往往是不可靠的。以前曾经搞过一项旨在了解杂志阅读量的上门调查，其中的一个主要问题是：你和你的家人阅读什么杂志？当将调查结果制表并分析后发现，大部分的人喜欢《琴师》(*Harper's*)，而没有多少人喜欢《真实故事》(*True Story*)。但出版商提供的数据却很明显地表明：《真实故事》的发行量是几百万份，而《琴师》只有几十万份。正如这项调查的设计者所疑惑的，或许他们问错了对象？但这不可能，因为上门调查走访了美国各式各样的居民区。惟一合理的解释是许多被调查者没有说实话，导致调查结果偏离了事实。

最后你将发现，当你想了解人们到底在读什么时，询问是无济于事的。直接上门收购旧杂志看看他们能提供什么，或许能得到更多的信息。你所要做的就是点一点《耶鲁评论》(*Yale Reviews*)和《爱情罗曼史》(*Love Romances*)各自的份数。即便是这种方法也并不确定，它只能说明人们曾经有过什么，而不是现在有什么。

同样，当你下次在阅读时看到普通美国人，这个词最近频繁出现，但大多数情况下是不现实的。每天刷牙1.02次——这个数据是我瞎编的，但它与其他任何数据一样有用——请自问一个问题：任何一个人怎么可能发现这

个事实？一个妇女在看了无数宣传不刷牙者是对社会冒犯的广告之后还会向陌生人承认自己不经常刷牙吗？这个统计数据只能对那些关心人们如何谈论刷牙的人才有价值，却根本不能反映牙齿接触牙刷的频率。

我们知道，一条河永远不可能高于它的源头。但如果在河的某处藏有水电站，却可以做到。同理，对样本研究后得到的结论不会好于样本本身。当数据经过层层统计处理，最后简化成一个小数形式的平均数时，结论似乎被确定的光环所笼罩，但只要再仔细留心整个抽样过程，这个光环就会破灭。



癌症的早期发现能否挽救生命？也许吧。但通常用于证明这一点的数据却更适合支撑相反的结论。这要追溯到 1935 年，根据康涅狄格肿瘤研究所挂号处（The Connecticut Tumor Registry）的记录从 1935 年到 1941

年手术后 5 年的存活率大量上升。但实际上这些记录是从 1941 年才开始登记的，在此之前的数据则是通过跟踪的形式得到的。许多病人离开了康涅狄格州，其生死与否无从得知。正如医务记者雷纳得·恩格尔(Leonard Engel)所说“所存在的内在偏差已足以解释存活率上升的真相”。

一个以抽样为基础的报告如果要有价值，就必须使用具有代表性的样本，这种样本排除了各种误差。这就是耶鲁毕业生的收入数据失真的原因，也是许多你在报纸或杂志上读到的报道毫无意义的原因。

一位心理医生曾经报道：实际上所有的人都是神经质的。暂且不去管这种提法是否破坏了“神经质”一词的含义，我们来看看这个医生的样本。也就是说，他观察了哪些人？结果证实，他是在对他的病人进行研究后得到了这个发人深省的结论，这和代表全体人的样本可差得太远、太远了。如果一个人心理健全，他永远不可能接受心理医生治疗。

对你所读到的东西多思考一下，你将避免接受许多似是而非的结论。

记住下面这点是有益的：无形的误差与有形的误差一样容易破坏样本的可信度。也就是说，即使你找不到任何破坏性的误差来源，但只要有产生误差的可能性，你



就有必要对结果保留一定的怀疑。如果你还有一丝的疑惑 想想 1948 年和 1952 年的总统大选 它们已足够证明这一点。

更远的例子可以追溯到 1936 年《文学文摘》(*Literary Digest*)的惨败。曾经准确预测了 1932 年美国大选的 1 000 万个《文学文摘》的订阅者对 1936 年的大选进行了预测 通过电话 他们向这个倒霉的杂志编辑信誓旦旦地保证 兰登 London 将在竞选中脱颖而出 并且与罗斯福 Roosevelt) 所得的票数比是 370 : 161。这样一个久经考验的调查群体怎么可能产生误差呢？但的确有误差，正如后来许多大学论文和报社评论员发现的，1936 年就有能力购买电话和订阅杂志的人并不能真正代表选民 至少在经济上 他们是极特殊的 是有偏的 后来证实他们中许多人是共和党的选民。该样本选择了兰登，而

选民心里却想着罗斯福。



最基本的样本是随机样本，它是指完全遵循随机的原则从总体中选出样本。总体即形成样本的母体。从索引卡片档案中将每隔 10 个的名字抽出来 从许多纸张中任意抽出 50 张；在马克特街上每遇见的第 20 个人作为访问的对象（但需要注意的是 在最后一个例子中 总体并不是全世界的人，也不是全美国人或者全旧金山人，而只是当时在马克特街上的人。一个进行民意调查的访问员说 她选择在火车站进行调查的原因是“那儿能遇到所有类型的人”。但应该向她指出的是 有些人比如婴儿母亲的代表性不足。）

随机样本的检验标准是：总体中的每个名字或事物是否具有相同的几率被选进样本？

纯随机抽样是惟一一种能有足够把握利用统计理论进行检验的抽样方法。但它同样存在着缺陷。在许多情况下 获得这种样本难度很大 并且十分昂贵 以至于单纯的经济考虑就会剔除这种方法。一个更经济的替代品是分层抽样，它在市场研究和民意调查等领域中得到了广泛的应用。

为了获得分层抽样下的样本，你需要将总体按照事先已知的优势比例划分成不同的组。这时你就可能遇上麻烦：关于分组比例的信息可能并不正确。你对访问员进行指导，以确保他们调查到一定比例的黑人，按照这样或那样的比例调查属于不同收入阶层的人，一定数量的农民 等等。而且 每一组人都要确保 40 岁以下和 40 岁以上的人数相同。

这听上去很不错，但实际上呢？在黑人还是白人的问题上，大部分时候访问员能准确判断。但在收入分组时，他会出很多错。至于农民，你如何划分一个在城镇上班又有部分时间种地的人？甚至连岁数的问题也会引起差错 为了确保准确性 访问员会挑选那些看上去明显小于 40 岁或大于 40 岁的人进行调查。在这种情况下，由于缺少 40 岁左右的人而导致样本有偏。你不可能获胜！

除此之外，如何在各层内部获得随机样本呢？最有效的办法是准备好每一层中所有单位的名单，并以随机

抽中的名单构成样本。当然，这耗资不菲。于是又转为街头调查，但由于遗漏了呆在家中的人而变得有偏。白天挨家挨户上门调查，又遗漏了上班族；转而改为晚上访问，但又不能包括那些看电影和去夜总会的人。

民意调查最终将演变为一场与误差的遭遇战。所有信誉良好的调查公司将不可避免地投入到这场战斗中。调查报告的读者应谨记这点：这场战斗永远不会取得胜利。在看到“67%的美国人反对”或其他类似的字眼时，应保留这样一个问题：67%的哪些美国人？

阿尔弗雷德·C. 金西 (Alfred C. Kinsey) 博士的《妇女卷》(*Female Volume*) 也和任何建立在抽样基础之上的读物一样。关键是如何阅读这本书 (或其通俗读本) 才能避免学习很多不必要的东西。这里至少包括了 3 次抽样。金西博士从全部妇女中抽出的样本 (第一次抽样) 远不能称为随机样本，不具有显著的代表性，但与该领域以前所做的任何工作相比，这个样本的容量较大，他的数据虽然不一定是最好的，但该数据比较重要并有一定启发意义。重要的是任何一个调查问卷都只不过是所有相关问题的样本 (第二次抽样)，而女士们给的答案也只是她们对于这个问题看法和态度的样本 (第三次抽样)。

由谁组成调查人员也会对调查结果产生微妙的影响。第二次世界大战期间，国家民意调查中心 (The Na-