

□学曦

### 打破“正常”迷思

生活中,我们常常依赖直觉解读数据,却不知这些看似合理的判断,往往藏着致命漏洞。比如,我们从小被灌输“接近平均值就是正常,偏离均值就是怪异”的观念。身高要中等,成绩要中等,性格要中庸,仿佛只有符合“平均标准”,才是安全、正确的存在。但《数据决策》用大量真实数据和经典实验告诉我们:“平均人”只是一个虚构的神话,每个人都天生“不正常”,差异才是人类的本质。

没有任何人能符合所有“平均标准”,这个结论最早来自美国空军的经典研究。20世纪50年代,空军为了设计适配所有飞行员的座舱,测量了4000多名飞行员的10项关键身体数据,试图找到一个“标准飞行员”,即所有指标都落在平均值附近的人。结果令人震惊:没有一个人符合全部10项标准,哪怕放宽筛选范围,也找不到一个完全“平均”的飞行员。

艾伦·唐尼用更现代的人体测量数据重复了这一实验,结果完全一致。在94项身体测量指标中,几乎每个人都有至少40项偏离平均值,90%的人有超过57项“非标准”数据。不止身体,心理特征也是如此。针对87万人的“大五人格测试”显示,五项特质全部接近平均值的人,仅占千分之一,绝大多数人都有明显的性格偏向。

《数据决策》认为,这意味着我们不必为自己和平均值不同而焦虑。个子太高或太矮,性格太内向或太外向,能力某方面突出而某方面不足,都不是“怪异”,而是人类的常态。那些以“平均标准”设计的衣服、座椅、工具、制度,本质上都无法适配真实的人,这也是标准化产品总让人觉得不合身、统一规则总让人觉得不适应的根本原因。

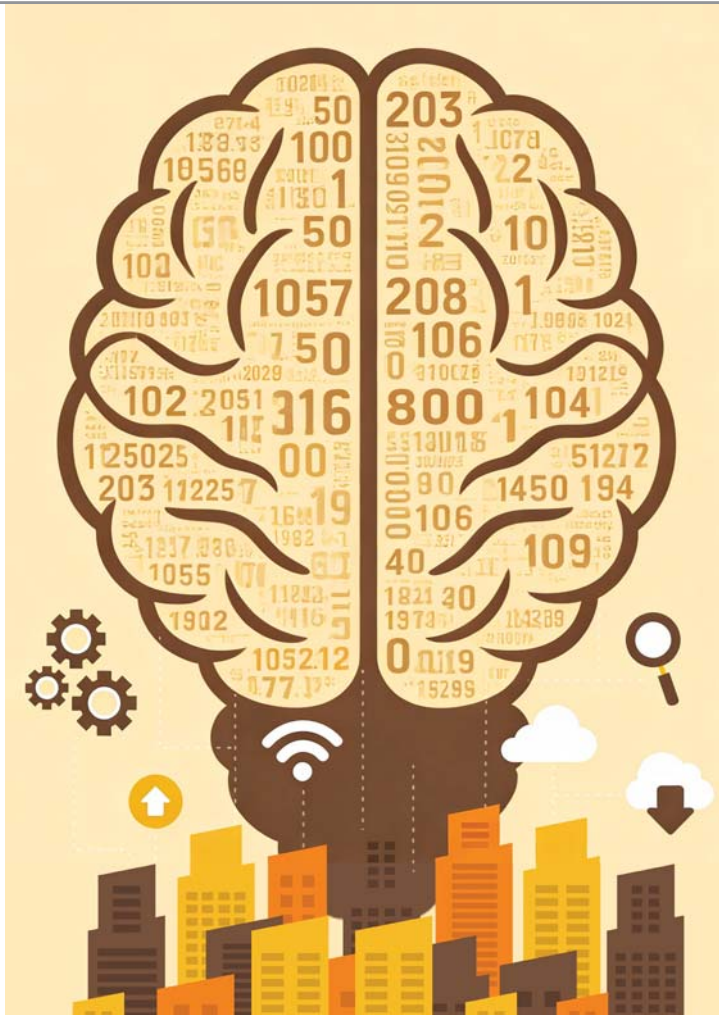
我们熟悉的身高、体重、臂长等身体数据,大多符合高斯分布(正态分布),呈现中间多、两头少的钟形曲线,大多数人集中在平均值附近,极端值很少。这种分布下,差异是有限的,最高的人不会比平均值高出太多,最矮的人也不会低得离谱。

但还有一类数据属于对数正态分布,则与人们的直觉认知完全不同。体重、跑步速度、国际象棋水平、音乐能力、事业成就,都遵循这种分布。它们的曲线不对称,右侧有一条长长的“尾巴”,意味着极端优秀的人,会比平均值高出数倍、数十倍,甚至上百倍。这就能解释为什么顶尖运动员能远超普通选手,国际象棋大师能层层碾压高手,商业天才的成就就能甩开常人。

那些“顶尖牛人”不是天选之子,而是对数正态分布的自然结果。能力不是简单相加,而是多重因素相乘。天赋、环境、训练、专注、机遇,任何一项拉满都会放大优势,所有因素协同,就会走到分布的极端尾部。在《数据决策》看来,这个规律也打破了“一万小时定律”的绝对化认知。单纯的重复练习无法突破分布限制,只有让优势因素形成乘数效应,才能走向顶尖。

### 统计悖论

生活中很多看似矛盾、无法理解的现象,其实都是统计悖论在作祟。



## 让数据“说”实话

我们生活在一个被数据包围的时代。商家用数据说服我们消费,专家用数据解释社会现象,就连日常做选择,也总被各种数字、比例、趋势所影响。艾伦·唐尼的《数据决策》,是一本帮我们看穿数据假象,避开统计悖论,用理性做选择的通俗读物。它没有复杂公式,没有枯燥推导,而是用生活里的真实故事,经典实验,社会现象,把统计学变成人人能用的生活智慧。这本书像一把解码数据的“钥匙”,提醒我们保持清醒思考,做出理性决策。

书中讲了一个有趣的故事。国王为了控制人口过快增长,颁布法令,要求每位女性生育的孩子数量,必须比自己的母亲更少。本以为这样家庭规模会缩小,人口增长会放缓,可25年后,平均家庭规模从3人涨到3.3人,人口增长反而更快。

这是著名的普雷斯顿悖论。核心原因是长度偏倚抽样:大家庭的孩子数量更多,下一代中,来自大家庭的女性占比会呈指数级增长。哪怕每位女性都比母亲少生一个孩子,来自大家庭的女性基数太大,依然会拉高整体平均生育数。

现实中这一规律同样成立。20世纪70年代,美国女性平均生育3个孩子,即便下一代女性每人少生1个,平均家庭规模依然短暂上升,直到生育数量大幅减少,才真正下降。这揭示了人口统计的基本逻辑:人口变化不是简单的个体减法,而是群体结构的动态博弈。单纯限制个体生育数量,无法直接控制人口增长,群体结构的变化才是关键。

统计悖论的影响无处不在。

这两个悖论教会我们最重要的数据思维是,关联不等于因果。看到数据结论,先别急着相信,要追问背后的原因,理清变量之间的真正关系,否则就会得出完全错误,甚至有害的结论。

肥胖悖论也是同理。肥胖是心脏病的诱因之一,但遗传、衰老、慢性病等诱因对心脏的危害更大。肥胖患者的心脏病,大多是单纯肥胖导致,而正常体重患者的心脏病,往往由更严重的病因引发,所以肥胖患者存活率反而更高。

《数据决策》不是为了普及统计学知识,而是为了让普通人拥

你一定有过这样的感受:商业街里的网红餐厅,味道往往平平;偏僻小巷里的小店,反而惊艳味蕾;名校里的学生,好像语文好的数学差,外向的人不够细腻。《数据决策》告诉我们,这些看似成立的规律,其实都是伯克森悖论制造的虚假关联。

以大学录取为例,精英大学只录取总分达标的学生,经过筛选后,校园里的学生就可能呈现出“语文好则数学差”的负相关。因为总分不够的人被淘汰,只剩下一科极强、一科偏弱的学生能进入名校。

生活中这样的陷阱无处不在。偏僻餐厅能长期存活,必然靠味道取胜,否则早就被市场淘汰;能进入职场核心岗位的人,要么能力强,要么情商高,很少有两者都顶尖的人;我们看到的“性格互补”“能力偏科”,不是真实规律,而是筛选规则的结果。

这一悖论的警示是,做判断时,一定要警惕被筛选过的样本。找餐厅不看位置,选人才不看单一特质,看数据不看局部结果,否则只会被虚假关联误导,做出错误判断。

医学和健康领域,也有两个长期误导人的悖论。曾经有研究发现,吸烟母亲生下的低体重儿,死亡率比不吸烟母亲的低体重儿更低,甚至有人据此宣称“吸烟对低体重儿有益”。还有研究发现,肥胖的心脏病患者,存活率比正常体重患者更高,让人误以为肥胖对心脏病有保护作用。

《数据决策》直言,这些结论都是完全错误的,根源是因果混淆。低体重儿有两种成因:一是母亲吸烟,危害相对较小;二是先天缺陷,危害极大。吸烟母亲的低体重儿,大多是吸烟导致,而非先天缺陷,所以死亡率更低。不吸烟母亲的低体重儿,大概率是先天缺陷,死亡率自然更高。吸烟始终有害,只是数据混淆了真正的因果。

在教育政策制定中,抽样偏差的识别也至关重要。美国普渡大学的班级规模争议告诉我们,制定教育资源配置政策时,不能仅仅依赖官方统计的平均班级规模,还需要考虑学生的实际体验。如果大量学生处于大班教学中,即使官方平均数据达标,也可能存在教学质量问题。通过加权再抽样的方法,可以从学生调查的有偏数据中,估算出真实的班级规模分布,为政策制定提供可靠的依据。

用数据规律找准发展方向,则是《数据决策》给每个读者的人生建议。

### 让理性主导选择

同时,书中“年龄—时代—世代”的分析,打破了“人老必保守”的偏见。人的观念不是随年龄增长而改变,而是由成长的时代环境决定,且大多数人会随着社会进步变得更开放、更包容。年龄从不是思想僵化的理由,持续学习、接触新事物,就能始终保持清醒的认知。

直觉很容易出错,数据不会骗人,但解读数据的方式会骗人。正如艾伦·唐尼在书中所写:“数据决策的目标,不是追求绝对正确,而是通过证据和理性,提高决策的胜率。”生活中没有绝对的确定性,但学会识别抽样偏差、因果混淆、极端值偏差等常见陷阱,就有望在复杂的信息环境中保持清醒,做出更明智的选择。

### 【相关阅读】



《数据决策》  
[美]艾伦·唐尼 著  
何生 译  
中信出版集团



《简单统计学》  
[美]加里·史密斯 著  
刘清山 译  
后浪 | 九州出版社



《女士品茶:统计学如何变革科学和生活》  
[美]戴维·萨尔斯伯格 著  
刘清山 译  
后浪 | 九州出版社