

应用统计案例

案例 1

《文学文摘》与抽样调查

《文学文摘》是美国一个很有名的刊物，1936 年预测美国总统选举结果时发生了重大失误。当年的两位总统候选人，一位是民主党的罗斯福，一位是共和党的兰登，当时大多数民意测验、新闻机构和政治观察家都预测罗斯福会获胜，但《文学文摘》与众不同，它预测兰登会以 57% 的优势战胜罗斯福。最后的结果是：罗斯福以 62%：38% 的压倒优势当选。由于出现了这个重大失误，这家杂志社不久即宣告破产。

《文学文摘》的预测，并非一种主观臆断，而是依据对 240 万人的民意测验作出的。附带说明，盖洛普咨询有限公司在第二次世界大战以后做过多次关于总统大选结果的民意测验，不仅与实际结果接近（当选者预测无误，得票率估计略有误差），且调查的人数也不过几千人。

为何《文学文摘》做了这么大规模的调查，反而没有取得满意的结果呢？回顾一下该刊的抽样过程，该刊的访问对象大多数是从电话号码簿和俱乐部会员名册上挑选的，这样做在工作上带来了方便。如果要在全国范围内用随机的方法挑选访问对象，则麻烦要大得多。但在 1936 年美国家庭装的电话机只有 1100 万部左右，因此有家用电话者，尤其是有条件参加某种俱乐部的人，大多数是经济上较富有、政治上保守而倾向共和党的选民。也就是说，较贫困的阶层，包括当时多达 900 万的失业者，在样本中缺乏其应有的代表性。当时正值 1929

1933 年经济大萧条过去不久，较贫困的阶层人数不少，与兰登相比，罗斯福推行的新政策较多地考虑了这些人。该刊起初拟访问 1000 万人，相信在这个庞大的样本中，美国社会各阶层的代表性会好些，但这 1000 万人中只有 240 万人寄回了对问题简单的回答。较富有的人以及文化程度较高的人，对当时的现实比较满意，作出回答的可能性要大一些，这个倾向有利于共和党。这一点曾在芝加哥地区得到证实：该刊向芝加哥地区 1/3 的登记选民发了问卷，有 20% 的人做了回答，其中半数以上有利于兰登。但实际结果是：在芝加哥是罗斯福以 2:1 的优势战胜了兰登。

了解这个案例后，大家不禁会问，为什么《文学文摘》会预测失误？怎样运用抽样调查进行预测才能得到好的结果？下面我们的论述会澄清大家的疑问！

案例 2

由于战争，德国有一个时期物资特别紧缺，对面包实行配给制：政府把面粉发给指定的面包房，面包师傅烤好了面包再发给居民。有一个统计学家，怀疑他所在区域的面包师

傅私扣面粉，于是就天天称自己的面包。几个月以后，他去找面包师傅，说：“政府规定配给的面包是400克，因为模具和其他因素，浮做的面包可能是398、399克，也可能是401、402克，但是按照统计学的正态分布原理，这么多天的面包重量平均应该等于400克，可是你给我的面包平均重量是398克。我有理由怀疑是你使用较小的模具，私吞了面粉。”面包师傅承认确实私吞了面粉，并再三道歉保证马上更换正常的模具。又过了几个月，统计学家又去找这个面包师傅，说：“虽然这几个月你给我的面包都在400克以上，但是这可能是因为你没有私吞面粉，也可能是因为你从面包里特意挑大的给我。同样根据正态分布原理，这么多天不可能没有低于400克的面包，所以我认为你只是特意给了我比较大的面包，而不是更换了正常的模具。我会立刻要求政府检查你的模具。”面包师傅只好当众认错道歉，接受处罚。

案例3

为什么庄家只赚不赔呢？因为庄家是根据中奖的概率来确定中奖额的。简单地说，如果中奖的概率是十分之一，也就是说十张彩票中有一张是中奖的。假设每张彩票卖2元钱，十张彩票可以卖出二十元钱。这十张彩票分别被十个人买走。其中有一个人中奖了，他中奖了，他会中多少呢？他最多只能中二十元，庄家最多只奖给他十倍于彩票价格的钱，再奖下去庄家就要亏了。也就是说，当庄家在制定中奖规则的时候，首先就要运用统计学原理，来计算一下老百姓中奖的机会有多大，如果机会大的话，庄家就给你少奖一些钱，机会小的话就给你多奖一些。老百姓中奖的机会是百分之一，庄家就最多给老百姓发放100倍的奖金，通常来说，发个七八十倍，老百姓自己就觉得很了不起了，殊不知另外的二十倍让庄家自己赚走了。但是有几个人会那么细心地计算中奖的几率呢？况且再分他个特等奖、一等奖、二等奖、三等奖、鼓励奖，就更加让人眼花缭乱了，谁还算得那么清楚？结果是庄家赚得更多了。而且，就算是计算了又怎么样？每一个人都相信自己的运气好，只要中奖了，哪怕该奖100倍的奖了80倍，中奖者也赚了呀！中奖者和庄家都赚了，赔的是那些没有中奖的老百姓。所以说，所谓彩票，是庄家和少量中奖者，共同分享广大非中奖者金钱的一场社会性活动。