

数据分析基础

平均值

1. 从概念上看，平均值有很多种。单从数学上来说，就有算术平均值、几何平均值、平方平均值、调和平均值、加权平均值等等。

平均值在什么情况下才有价值

1. 平均值是用所有的样本数据进行计算的，容易受到极端值的影响。在不少情况下，平均值是没有价值的，它无法客观准确的反应数据的整体情况，更进一步来说，整体平均值实在数据成分布均匀或者正态分布的情况下才有意义，如果忽略整个数据的分布情况，只提平均值其实是没有意义的。
2. 整体平均值不能代表各分组情况，分组结论和整体平均值结论可能会大相径庭。
3. 辛普森悖论：**在分组比较中都站优势的一方，有的时候在总评中反而是失势的一方**，用一句话来讲，就是“质”（命中率）与“量”（投球数）是两个维度的数据，如果全部合并成“质”（命中率）这个维度的数据，那就会出错了。每次小范围内的输赢，其实和你在整体上的输赢没有太大直接的关系。这也是为什么在打麻将或者打德州真正赢的那些人，不是那些小牌把把赢的人，而往往是赢一把大的人。
4. 看到一个平均值的时候，需要查看看它的数据构成情况，而不是简单地用平均值去代表所有的整体。

总结

整体平均值是在数据均匀分布或者正态分布下才会有意义，如果忽略整个数据的分布情况，只提平均值是没有价值的。分组结论和整体平均值结论可能会大相径庭。

大数定律和小数陷阱

什么是大数定律

1. 大数定律是由瑞士数学家雅各布·伯努利提出来并验证的，它的核心逻辑是说当随机事件发生的次数足够多时，发生的频率才会趋近于预期的概率。

什么是小数陷阱

1. 当我们了解了大数定律后，大概率认为随着数据的增多，整体趋势会趋向于50%。那么我们在赌场玩轮盘赌大小的时候，然后前面开的都是大，那么我们接下来应该向小的去加倍下注。因为理论上长期来看出现“大”和“小”的概率应该是趋于一致的，所以未来出现“小”的概率应该增大。
2. 每次的事件其实和上一次的事件是独立且随机的，并不是前面都是“大”，后面开“小”的概率就会高。

大数定律和小数陷阱给我们的启示

对于大数定律来说，我们做事不要轻易跟风，开始某件事之前要想好自己是否有持续投入的能力。

对于小数陷阱来说，我们要保持一颗平常心，当我们在做某一件事的时候，如果失败了，要有平常心。气馁或者赌徒心理都是不可取的。你需要意识到，我们在生活中的数，绝大多数都不够大，所以我们的偶然只是偶然，并不是我们这段时间就该倒霉。

总结

多利用前人的经验，站在前人的实验结果和规律上，多去学习、阅读。多去经历、总结自己想从事的事业。相信绝大多数人是做不到在赌博中拥有足够大的“大数”，但是没关系，你可以根据自己的生活去建立属于自己的“大数定律”，赋予偶然的生活必然性。

大数定律说的是当随机事件发生的次数足够多时，发生的频率就趋近于预期的概率。对于一件事情，你需要持续不断努力，才可以达到你的期望值。而“小数陷阱”则告诉你，每个事件都是独立的事件，“否极泰来”需要足够多的次数才可能出现，做事情要少一些“赌徒心态”，多一些平常心，不要盲目跟风和下注才能获得最后的成功。

数据的期望值

墨菲定律

如果有两种或两种以上的方式去做某件事情，而其中一种选择方式将导致灾难，则必定有人会做出这种选择。

- 任何事情都不会像它表面上看起来那么简单；
- 所有任务的完成周期都会比你预计的长；
- 任何事情只要有出错的可能，就会有极大的概率出错；
- 如果你预感某件事可能出错，它一定会出错。

期望值

1. 期望值就是对可能出现的结果以概率为权做加权平均。
2. 当样本量 N 趋近无穷大的时候，样本的平均值无限接近数学期望。
3. 期望就是反映在大数定律下多次执行某件事情之后，得到的一个最可能的收益结果。

解释墨菲定律

人都有一个特殊的心理机制，那就是倾向于记住一些不好的事情。就像航班晚点的概率对每一个人而言都是一样的，但对我来讲，每次飞机晚点的等待就会让我印象深刻，而平时没有晚点的时候，我的注意力都集中在其他事情上。

抽象一下，你经常担心一个坏事的发生，比如说事情 A，我们假设 A 发生时对你产生的心理作用是 X，事情 A 不发生时对你产生的心理作用是 Y。这个时候显然 X 是大于 Y 的，所以当发生 A 这件事情的时候，你受到的心理影响 X 就比 Y 大。

其实，你在担心一件事情的发生之时，这件事情已经具备了发生的大多数条件。我们假设事件 A 发生的概率是 R，那么当你担心这件事情发生（R1）和不担心这件事情发生（R2）的时候，两个概率是不一样的，R1 一定大于 R2。

如何规避墨菲定律

对于心理影响来说，我们要做的就是不断调整事情对你心里影响的预期，让它们趋同。特别是遇到坏事情的时候，你可以通过增加 B 计划等方式，调整预期以降低坏事情发生对你的心理影响。

对风险的概率来说，你可以优化流程，提高自身能力，尽可能减少事件出错的概率。

“为大概率坚持，为小概率备份”——创业的时候，我们要努力为好的期望（N）去坚持，同时考虑为坏的影响（X）备份，应该尽力降低坏期望（R1）的风险。

“已知的是成本，未知的才是风险”——如果坏的影响（X）为已知，那么即使你按照坏事件发生概率（R1）100% 来准备资金，这批资金也算是你付出的成本；但如果坏影响（X）未知，那么无论坏事件发生概率（R1）为多少，都是风险，因为你不知道这个坏事件究竟会造成多大的影响。

“项目风险控制”——项目管理当中有各种风险管理和预防措施，把风险分为很多类，例如静态风险、动态风险、局部风险、整体风险，同时也会把风险应对措施细分为很多类，其实背后的核心是为了去避免墨菲定律的发生，让整体项目在项目经理的期望值下正确运行。

“生活中的风险控制”——我们在生活中，其实也是可以借用这种风险控制的方法论，识别生活中的风险并做好准备，这样才能够在墨菲定律发生的时候不会手忙脚乱。例如提前看看天气预报，查看这趟航班的过往准点率，预估自己乘坐航班情况。在去重要会议的时候，多提前一些时间。这些生活里的小事你或许平时不会太过于在意，但请相信，一旦你把这些小事落到实处，你对生活的掌控力会大大提升。

总结

通过墨菲定律介绍了一个有意思的概念叫期望值（Mean），它是对可能出现的结果的概率加权平均，期望值完全是由概率分布决定。而我们之前讲的平均值一般是指算术平均值，也就是一组数据中所有数据之和再除以数据的个数。某个事情长期不断发生，次数足够多后会达到我们预设的期望值，这就是大数定律。

对平均值来说，你要学会为不同事物去分组，用更细分的数据来看待问题。对于大数定律来说，要成事，其实需要 we 不要有赌徒心态，要学会持续投入。而对于期望值来说，平衡预期和未雨绸缪这两个词，希望你能够在生活中灵活运用。

其实，这几个数据分析的概念都告诉我们一个最朴实的道理：没有事情可以一蹴而就（平均值），我们需要努力足够多的次数（大数定律），学会规避风险（期望值）。这样最终在若干年后，企业和个人才能有一份满意的企业 / 个人数据报表。