



用 AI 做经营分析， 三个常见的误判

从毛利、回款、费用三个案例，看 AI 的结论错在哪

Three Common Misjudgments in AI-Assisted Business Analysis

适用对象：在职财务 BP；希望转岗 BP 的会计与财务分析人员

配套：视频课约 17 分钟。本讲义含完整数据、拆解步骤、练习与参考文献

作者：杨浩（August） | 1215868565@qq.com

版本 v3 | 2026 年 10 月 | 案例公司与数据均为教学虚构

先看这一页

AI 写的经营分析，数字往往都核得上底稿，结论却可能是错的。错在算之前的三个决定：怎么合并，算了谁，跟谁比。本讲义讲这三个决定最容易做错的地方。

误判一

合计跌了，分项都在涨

-4.5

个点

毛利率下滑里，结构变化拖累的部分

怎么合的？

看到就警觉

合计数和分项数方向相反；“平均毛利率”“平均费用率”

怎么拆

拆成构成效应和率效应

误判二

只算了回来的钱

43.1

天

补回未回款票据后的回款周期下限，不是 38 天

算了谁，漏了谁？

看到就警觉

两期数据成熟度不同；空白、缺失、“已完成”口径

怎么拆

固定分母，固定观察窗口

误判三

把自然回落当成管控效果

8

个点

22 个点的降幅里，管控真正做出来的部分

跟谁比？

看到就警觉

“实施后改善了 X”；对象是按最差或最好挑出来的

怎么拆

找对照组，算双重差分

怎么用这份讲义

建议按这个顺序：先读第 1 章的案例，对三条结论各押一个答案；再看视频；最后回到讲义看完整拆解，做第 6 章的练习。先押答案再看解析，即使押错了，也比直接读答案记得牢，学习科学里叫预测试效应（Richland et al., 2009; Kornell et al., 2009）。

视频只讲每个误判的主线。讲义多了四样东西：完整的数据和拆解步骤，概念的学术出处，实务里的对应做法，以及常用工具的默认设置。

目录

1 一份 AI 写的经营分析	3
2 误判一：合计跌了，分项都在涨	5
3 误判二：只算了回来的钱	9
4 误判三：把自然回落当成管控效果	12
5 工具的默认设置与 AI 复核指令	15
6 练习与解析	16
7 延伸阅读	18
附 术语表与参考文献	20

案例

一份 AI 写的经营分析



禾川饮品（虚构）是一家区域性饮料公司，卖两类产品：功能饮料和包装水。全国 40 个销售区域，走经销商渠道，按票据回款。

周一上午的经营分析会，大屏上是 AI 助手生成的《Q3 经营分析》，摘要写了三条结论：

AI 助手生成 · 《Q3 经营分析》摘要

结论一 Q3 整体毛利率 34.5%，同比下降 3.5 个百分点，主因渠道价格战侵蚀单品毛利。建议申请 Q4 价格补贴预算。

结论二 Q3 平均回款周期 38 天，较 Q2 的 41 天缩短 3 天，回款管理持续改善。

结论三 Q3 对差旅超支最严重的区域实施专项管控，超支率由 36% 降至 14%，降幅 22 个百分点。建议 Q4 推广至全国。

三条结论里的每个数字都核得上底稿。问题出在算之前的三个决定：数字按什么方式合并，哪些对象被算了进来，拿什么作对比基准。财务把这些决定叫“口径”。AI 会替你把这三个决定做掉，而且不会主动告诉你它是怎么定的。

这类问题在用 AI 做分析时格外容易出现。从取数到清洗合并，这套 ETL 流程 AI 几句指令就能跑完，数字上的异常也能很快标出来。难的是下一步，解释数字为什么变。两件事在同一期出现，AI 很容易把它们写成果因，可相关不代表因果。毛利率下跌和渠道价格战落在同一个季度，不等于前者由后者造成。这一步不加分辨，结论就会误判。

人因工程研究把这种风险叫作自动化偏差（automation bias）：人把自动化系统的输出当作省力的替代，系统出错的时候，要么漏看了问题，要么照着错误的建议去做（Parasuraman & Riley, 1997; Parasuraman & Manzey, 2010）。报告写得越顺，越容易被放过去。

先押一个答案

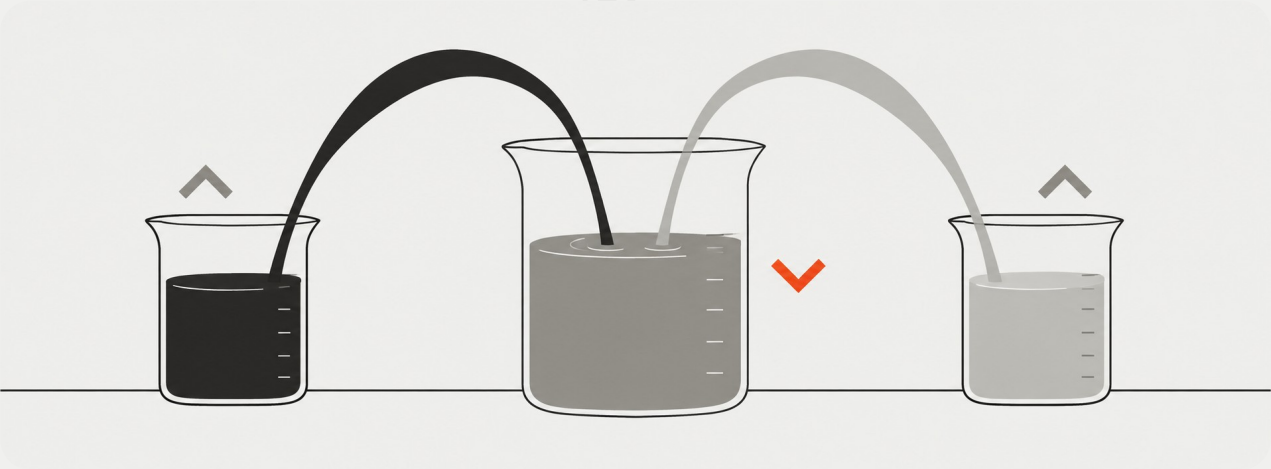
往下读之前，对三条结论各做一个判断，在右栏写一句理由。押错没关系。

结论	信不信	一句理由
一 价格战侵蚀单品毛利	信 / 不信	
二 回款周期缩短 3 天	信 / 不信	
三 管控让超支率降了 22 个点	信 / 不信	

底稿数据放在后面三章的开头，每章拆一条结论。

误判一

误判一：合计跌了，分项都在涨



对应结论一：“毛利率同比下降 3.5 个百分点，主因渠道价格战侵蚀单品毛利。”

表 1 禾川饮品品类毛利（单位：万元）

品类	去年 Q3 收入	去年 Q3 毛利率	今年 Q3 收入	今年 Q3 毛利率
功能饮料	6,000	50%	5,400	51%
包装水	4,000	20%	6,600	21%
合计	10,000	38.0%	12,000	34.5%

两个品类的毛利率都涨了 1 个点，合计却跌了 3.5 个点。三个数都对，放在一起互相矛盾。

同一张表，三个答案

算法	结果	回答的问题
按收入加权的毛利率	38.0% → 34.5%，跌 3.5 个点	公司整体每 100 元收入赚多少毛利
两个品类毛利率简单平均	35.0% → 36.0%，涨 1 个点	每个品类平均做得怎样，不管卖了多少
毛利额	3,800 万 → 4,140 万，多 340 万	公司一共多赚了多少钱

AI 报告只用了第一个数，然后把它归因到单品毛利上。三个数各自回答一个问题，用哪个，取决于会上要做什么决定。

拆开看：构成效应和率效应

毛利率的变化可以拆成两部分：卖什么变了（结构），每样东西赚多少变了（效率）。功能饮料的收入占比从 60% 降到 45%，包装水从 40% 升到 55%。

图 1

收入涨了 2,000 万，涨的全是低毛利的包装水

左档 = 去年 Q3 · 右档 = 今年 Q3 · 一档 = 200 万元 · 品类名下方是毛利率与收入占比

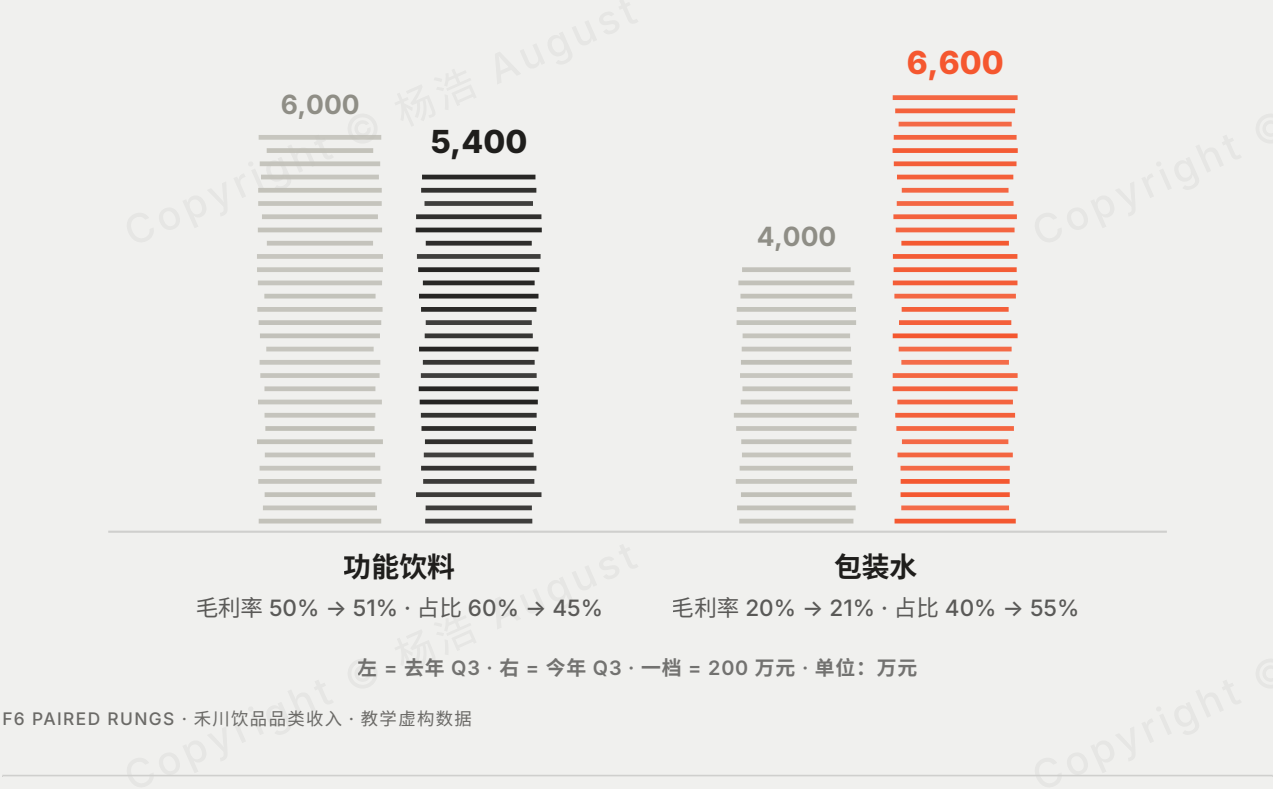
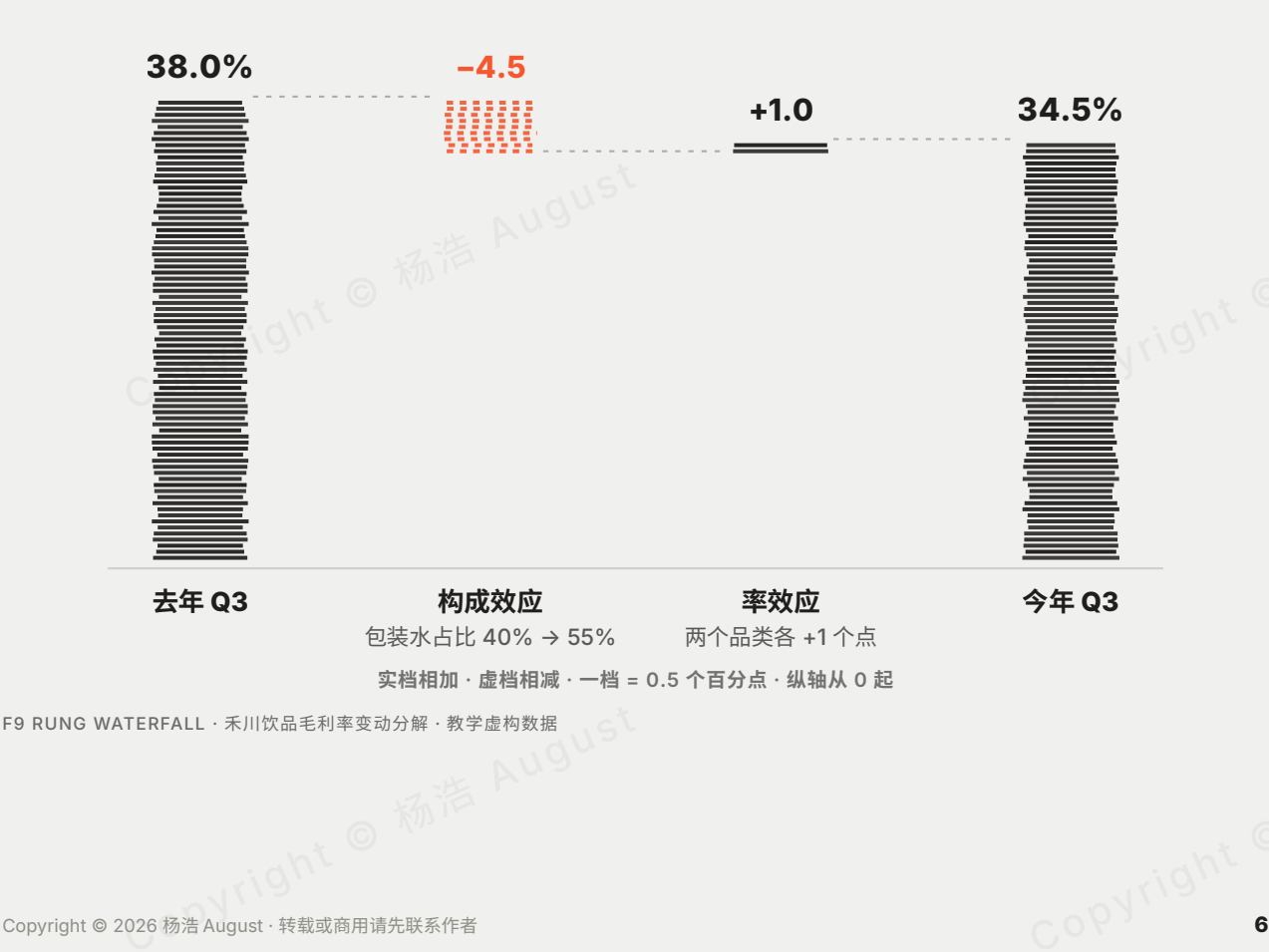


图 2

毛利率少了 3.5 个点：结构拖累 4.5，单品毛利反而贡献 1.0

一档 = 0.5 个百分点 · 纵轴从 0 起 · 实档 = 合计与加项 · 橙色虚档 = 减项



瀑布图里的 -4.5 和 +1.0，分两步算出来：

第一步，只换结构。占比用今年的，毛利率仍用去年的： $0.45 \times 50\% + 0.55 \times 20\% = 33.5\%$ 。比去年的 38.0% 低 4.5 个点。这 4.5 个点只来自“卖什么”的变化，叫**构成效应**。

第二步，再换毛利率。占比和毛利率都用今年的： $0.45 \times 51\% + 0.55 \times 21\% = 34.5\%$ 。比 33.5% 高 1.0 个点。这 1.0 个点来自各品类自己的毛利率变化，叫**率效应**。

$-4.5 + 1.0 = -3.5$ ，正好等于整体的变动。

结论翻转

单品毛利没有被侵蚀，两个品类的毛利率都在涨。整体下滑全部来自结构：包装水的收入占比从 40% 升到 55%。申请价格补贴，补的是一个不存在的窟窿。

该追问的是：低毛利的包装水放量，是主动抢份额，还是被渠道推着走？前者是一个有代价的战略选择，要算清楚值不值；后者是管理问题，要找原因。两种情况都用不上价格补贴。

分组和合并方向相反，不只出现在财务数据里。医学统计里有一组更有名的数：

图 3

分开看，开放手术两组都赢；合起来，经皮肾镜反而领先

每行 100 根 tick = 每 100 位患者 · 墨色 = 治愈 · 大结石更多交给了开放手术

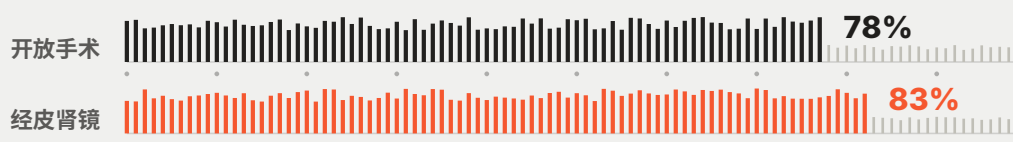
小结石



大结石



合计



一根 tick = 每 100 位患者中的 1 位 · 圆点每 10 根一个

L15 BALLOT TALLY · 肾结石治疗成功率 · CHARIG ET AL. (1986)

概念卡

辛普森悖论 (Simpson's paradox) 分组看是一个方向，合并看是另一个方向。Simpson (1951) 把它写清楚，Blyth (1972) 给它定名；更早的 Yule (1903) 已经记录过合并后关联消失的现象，所以文献里也叫尤尔-辛普森效应。

肾结石治疗是医学统计里被引用最多的例子。按结石大小分开统计，开放手术两组都胜出（小结石 93% 对 87%，大结石 73% 对 69%）；不分大小合起来算，经皮肾镜取石反倒领先（83% 对 78%）。原因在病例分配：医生倾向于把更难处理的大结石交给开放手术。结石大小同时影响了选哪种术式和治疗结果，统计上叫混杂因素 (Charig et al., 1986; Julious & Mullee, 1994)。

合并数和分组数哪个对？Pearl (2014) 的回答是：取决于数据背后的因果结构，光看表格判断不了。落到 BP 的场景：问定价和成本管控做得好不好，看分组；问公司整体赚钱能力，看合并；问要不要调结构，看分解。

实务卡

它就是你学过的连环替代法 这套拆法最早来自人口学，用来把两个地区死亡率之差拆成“人口结构不同”和“各年龄段死亡率不同”两部分 (Kitagawa, 1955)。在管理会计里，它对应销售组合差异；在 FP&A 里，常叫价格—数量—结构分析 (price-volume-mix, PVM)；中级和注会《财务管理》教材里，它就是因素分析法中的连环替代法。

连环替代的结果跟替代顺序有关：先换结构还是先换效率，两个效应的大小会略有不同。本例两个品类的毛利率涨幅相同，顺序不影响结果；一般情况下会影响，练习 1 专门演示。想避开顺序问题，可以用对称分解，两个效应都用两期的平均值作权重 (Kitagawa, 1955)：

$$\text{构成效应} = \sum (\text{今年占比} - \text{去年占比}) \times (\text{两期毛利率的平均}) \quad \text{率效应} = \sum (\text{今年毛利率} - \text{去年毛利率}) \times (\text{两期占比的平均})$$

AI 提醒

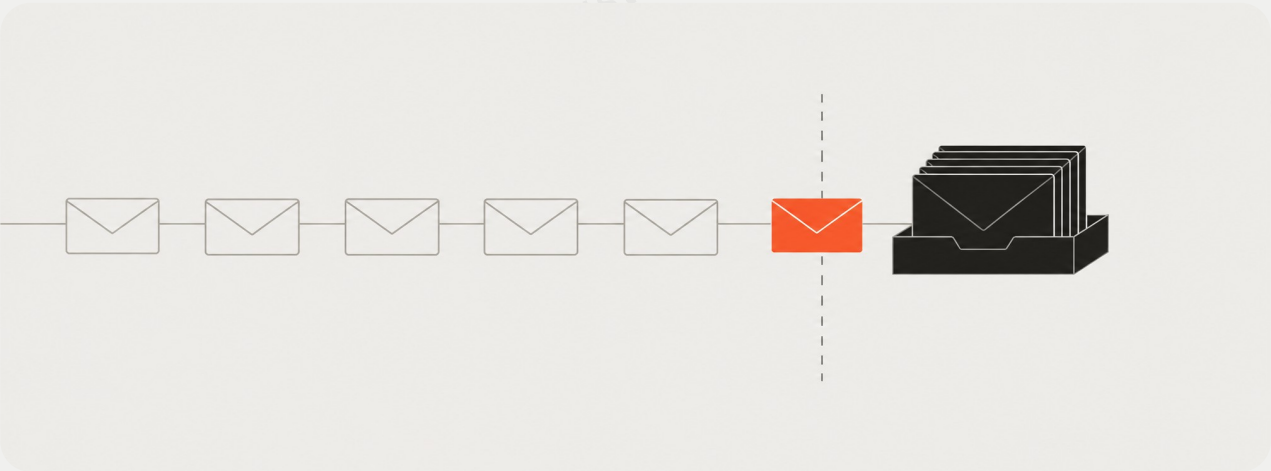
让 AI “算一下各区域的平均毛利率”，它可能按收入加权，也可能对各区域的毛利率做简单平均，然后照着算出来的数写一段很流畅的归因。它通常不会告诉你选的是哪一种。要它先写出合并方式和权重，再给结论。

带走一句

怎么合的？合计和分项方向相反时，先拆结构，再下结论。

误判二

误判二：只算了回来的钱



对应结论二：“平均回款周期 38 天，较 Q2 的 41 天缩短 3 天。”

表 2 票据回款（截至 9 月 30 日）

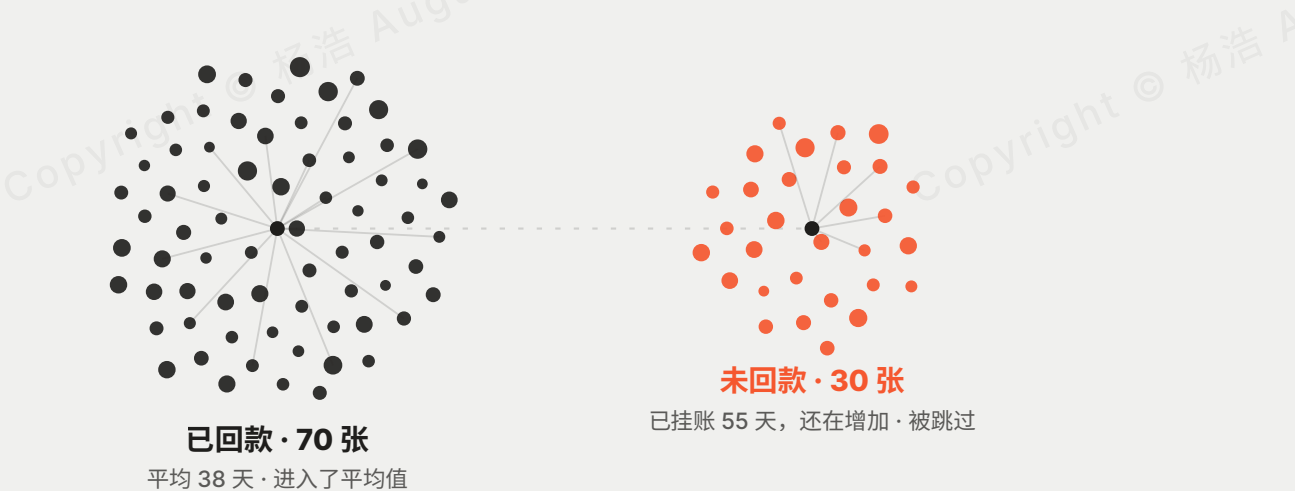
队列	开票张数	已回款	已回款平均天数	未回款	未回款平均已挂账天数
Q2 开票	100	100	41	0	—
Q3 开票	100	70	38	30	55（仍在增加）

38 天只来自已回款的 70 张票。另外 30 张还没回款，平均已经挂账 55 天，而且每过一天就多一天。

图 4

100 张票，平均回款周期只算了其中 70 张

一点 = 一张 Q3 开出的票 · 墨色 = 已回款，进入了平均值 · 橙色 = 截至 9 月 30 日仍未回款，被跳过



一点 = 一张票 · 70 + 30 = 100 · 补回 30 张，平均至少 43.1 天

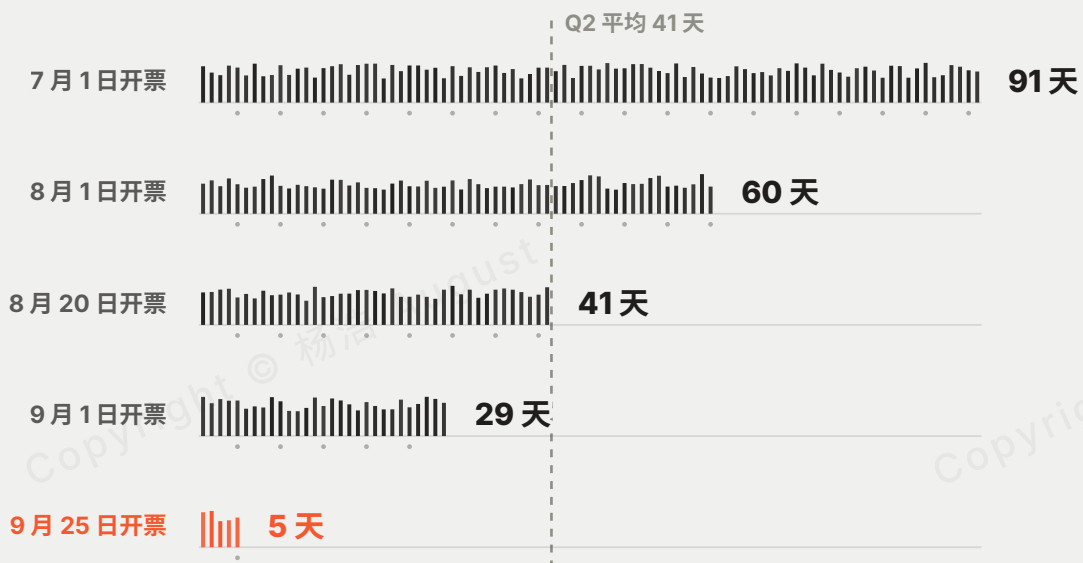
9 月 25 日开的票

一张 9 月 25 日开的票，要出现在“已回款”那一列，必须在 5 天之内回款。开票日离截止日越近，进这一列的门槛越高。8 月 20 日以后开的票，回款天数在这张表里不可能超过 41 天；回得慢的，到 9 月 30 日还待在“未回款”里。所以“已回款”这一列，天然偏向回得快的票。

图 5

开票越晚，能看到的回款天数越短

一 tick = 一天 · 每行是截至 9 月 30 日最多能观察到的天数 · 虚线 = Q2 平均回款 41 天



一 tick = 一天 · 圆点每 5 天一个 · 8 月 20 日以后开的票，表里不可能超过 41 天

F5 TICK ROWS · 观察截止日造成的右删失 · 教学虚构数据

把分母补回来

把 30 张未回款的票按今天的挂账天数算进去，并假设它们明天全部回款：

$$(70 \times 38 + 30 \times 55) \div 100 = 43.1 \text{ 天}$$

这只是下限，实际只会更长，而它已经比 Q2 的 41 天长了。

Q2 的 41 天，是 100 张票全部回款之后的完整数字；Q3 的 38 天，来自一个还没长完的队列。拿完整的数跟没长完的数比，“缩短 3 天”是口径制造出来的。

正确的比法

两期用同一套规则。常用的有两种。一是固定观察窗口：比如“开票后 45 天内回款的比例”，Q2、Q3 都只统计开票满 45 天的票，两期在同样的条件下比。二是等队列长完再比：Q3 的票全部回款之前，只报告已回款比例和未回款票据的账龄分布，不报告“平均回款周期”。

概念卡

右删失与 Kaplan-Meier 法 观察期结束时事件（这里是回款）还没发生，这种数据叫右删失（right censoring）。直接丢掉删失的记录，结果系统性偏短；把它们当作“今天回款”，得到的是下限。Kaplan-Meier 乘积极限法的做法是：每张未回款的票，在它被观察到的那些天里都留在分母中，算作“还没回款”，到观察截止才退出计算（Kaplan & Meier, 1958）。这个方法最早用于医学里的生存时间估计，同样适用于回款、客户流失、员工离职这类“等事件发生”的数据。

概念卡

幸存者偏差与样本选择偏差 只看留下来的，结论会系统性偏好。这是样本选择偏差的一种：一个对象不能进入样本，本身就和研究的结果有关（Heckman, 1979）。金融研究里的经典证据：只用仍在运作的基金做业绩研究，会高估基金业绩，因为业绩差的基金早已清盘、退出了样本。Elton 等（1996）的估计是每年约 0.9 个百分点；Brown 等（1992）指出，这种偏差甚至会让人误以为基金业绩有持续性。

实务卡

分母在开头定下，谁都不许中途离场 临床试验有一条原则叫意向性治疗（intention-to-treat）：所有被随机分组的对象，都按最初的分组纳入分析，不管后来有没有退出、有没有按方案接受治疗（Hollis & Campbell, 1999）。BP 可以借这个思路：分析开始前先写下分母是谁，比如“Q3 开出的全部 100 张票”，之后无论记录是否完成，都留在分母里。

AI 提醒

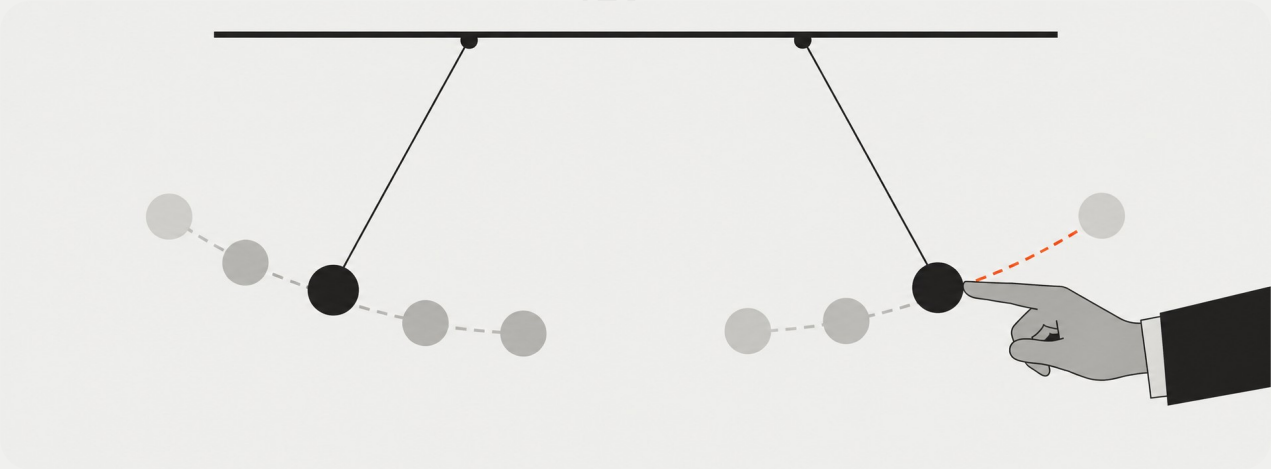
回款日期为空的行，和开票日期相减会得到缺失值。pandas 的 `mean()` 默认跳过缺失值（`skipna=True`），Excel 的 `AVERAGE` 默认忽略空白单元格（pandas development team, 2026; Microsoft, n.d.）。不报错，也不提示，30 张票就这样离开了分母。要 AI 报告三个数：起始条数、进入计算的条数、两者的差额和原因。

带走一句

算了谁，漏了谁？看到平均天数、平均 ROI、满意度，先问分母从哪批对象开始，中途有谁离开了。

误判三

误判三：把自然回落当成管控效果



对应结论三：“超支率由 36% 降至 14%，降幅 22 个百分点。建议 Q4 推广至全国。”

先想清楚机制：实力加运气

一个区域某个季度的差旅超支，可以拆成两部分：结构性的需求（区域大、客户分散、拜访频繁），加上一次性的因素（一场展会、一批集中报销）。挑出 Q2 超支最严重的 20 个区域，挑中的往往是一次性因素最集中的那一刻。下一个季度运气回到平常，数字自己就会往回落，管不管都会落一些。

表 3 差旅专项管控（超支率 = (实际差旅费 - 预算) ÷ 预算)

Q2 超支最严重的 20 个区域，因管控人力有限分两批：按区域编号单双号，10 个在 Q3 实施，10 个排期到 Q4。

组别	区域数	Q2 超支率	Q3 超支率	变动
管控组（Q3 实施）	10	36%	14%	-22 个点
待管控组（Q4 实施）	10	35%	21%	-14 个点

AI 报告只引用了管控组。待管控组一样严重，Q3 没做任何管控，也从 35% 降到了 21%。

双重差分：一次减法

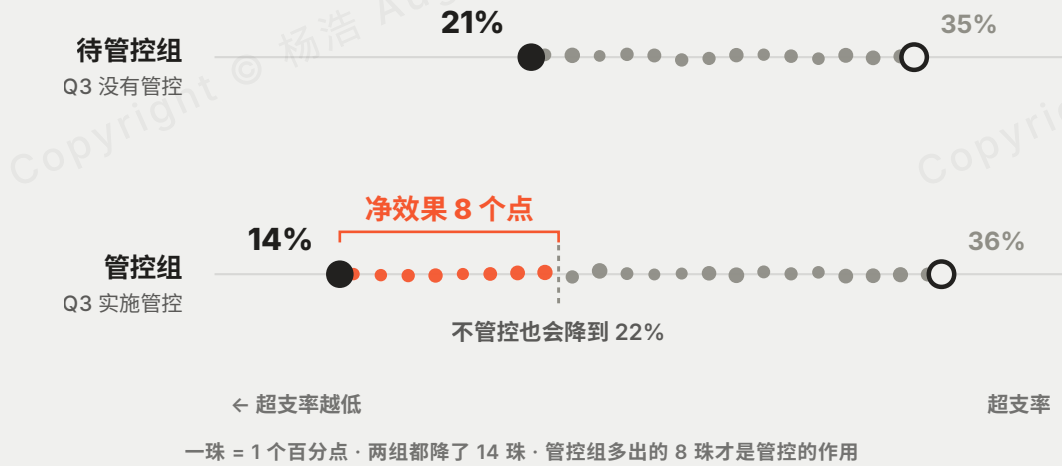
管控的净效果 = 管控组的变化 - 待管控组的变化 = (-22) - (-14) = -8 个点

22 个点里，约 14 个点不管控也会发生。待管控组同时吸收了两种干扰：一是回归均值，二是 Q3 本身可能是差旅淡季这类对两组都起作用的共同因素。

图 6

降了 22 个点，管控做出来的只有 8 个点

一珠 = 1 个百分点 · 空心 = Q2 · 实心 = Q3 · 灰珠 = 不管控也会降的部分 · 橙珠 = 管控的净效果



F12 DUMBBELL QUEUE · 差旅超支率双重差分 · 教学虚构数据

这个算法有一个前提：两组在管控之前足够相像，而且受同样的外部因素影响，方法论里叫平行趋势假设。本例两组都来自最严重的 20 个区域，按编号单双号分，Q2 超支率 36% 对 35%，前提基本成立。如果换成“最差的 10 个先管，次差的 10 个后管”，两组起点就不一样，这个比较也就不可靠了。

回到决策

管控的真实效果是 8 个点。拿去跟管控成本比的，应该是这 8 个点。8 个点可能仍然值得推广，但结论要改成：算清成本收益之后再决定；推广时继续分批，保留对照。

概念卡

回归均值 (regression toward the mean) “回归”这个词就出自 Galton (1886) 的身高研究：子女身高偏离人群均值的幅度，平均只有父母双方平均身高偏离幅度的三分之二左右，父母特别高的，子女往往没那么高。

管理场景里最传神的例子来自 Tversky 和 Kahneman (1974)。飞行教官发现，表扬一次漂亮的降落，下一次往往变差；严厉批评一次糟糕的降落，下一次往往变好，于是认定批评比表扬管用。这个结论站不住：极端表现之后，不管表扬还是批评，下一次都会往平均水平回落。专挑最差的部门帮扶，再把帮扶后的回升当成效，是同一个结构。

注意区分：金融里的“均值回归” (mean reversion) 指价格向长期均值收敛的过程，是另一个概念。

概念卡

反事实与双重差分 一项措施的效果，定义是“做了之后的结果”减去“同一批对象如果没做会怎样” (Rubin, 1974)。后一项永远观察不到，叫反事实 (counterfactual)，只能找最像的一组来代替，

这就是对照组。用两组前后变化之差估计效果，叫双重差分（difference-in-differences）。Card 和 Krueger（1994）用它比较新泽西州和宾夕法尼亚州的快餐店，研究最低工资上调对就业的影响，这个方法由此成为经济学的常用工具。

研究方法论把“前后对比为什么会骗人”归纳成内部效度威胁（threats to internal validity），其中包括回归假象、选择、流失、同期发生的其他事件等（Shadish et al., 2002）。误判二的漏票对应选择和流失，误判三的自然回落对应回归假象。

实务卡

先设计评估，再上线方案 分批上线本身就是一个免费的对照组。决定谁先谁后，用抽签或编号，不要按“谁最差谁先上”。上线之前，先写下要比的指标和比较的时间窗口。

AI 提醒

AI 默认只做前后对比。你不要求，它不会去找对照组，也不会提醒你样本是按“最差”挑出来的。

带走一句

跟谁比？看到“实施后改善了 X”，先问同期没做的那组变了多少。

工具

工具的默认设置与 AI 复核指令

三个误判里，有两个是被工具的默认设置悄悄带进来的。下表列出常见的几处。

工具和写法	默认做了什么	对应的误判	怎么查
pandas 的 <code>mean()</code> 、 <code>sum()</code> 等聚合函数	默认 <code>skipna=True</code> ，跳过缺失值	误判二	先用 <code>isna().sum()</code> 数出缺失条数，再决定怎么处理
Excel 的 <code>AVERAGE</code>	忽略空白单元格，但计入值为 0 的单元格	误判二	用 <code>COUNTBLANK</code> 数空白；空白和 0 含义不同，分开处理
对比率列求平均： <code>groupby(...)[“毛利率”].mean()</code> 、透视表的“平均值”	每一行一票一权，做简单平均	误判一	比率指标用“分子合计 ÷ 分母合计”重算
AI 生成的“改善”“见效”类结论	只比较同一批对象的前后	误判三	问有没有同期未处理的对照组

pandas 与 Excel 的默认行为见官方文档（pandas development team, 2026; Microsoft, n.d.）。

给 AI 的复核指令

下面这段可以直接粘贴到对话里，或写进你常用的分析模板。它的作用是让 AI 在下结论之前，先把三个决定摊开给你看。判断这些口径适不适合眼前的决策，仍然是你的工作。

- 你在为财务 BP 准备经营分析。给出任何结论之前，先逐项披露下列口径。
无法确定的写“未确定”，不要猜。
- 【误判一：怎么合的】
 - 每个比率指标的合并方式：加权平均（写明权重）还是简单平均。
 - 合并数与分组数方向不一致时，把变动拆成构成效应与率效应，分别报告。
- 【误判二：算了谁】
 - 分析对象的起始名单与条数、最终进入计算的条数，以及差额的原因（缺失值、未完成、已剔除）。
 - 缺失值与空白的处理方式；不得静默跳过。
 - 被比较的两期是否处于同样的观察成熟度；不同时，改用固定观察窗口。
- 【误判三：跟谁比】
 - 任何“改善”“见效”“拉动”类结论，写明对照组是谁；
没有对照组时，标注“仅前后对比，不能归因”。
 - 分析对象若因极端表现入选（最差或最好的 N 个），提示回归均值的可能。

最后单列一段：结论的可信度，以及需要人工核实的地方。

练习

练习与解析

先自己做，再看下一节的解析。练习 1-4 各对应一个误判，练习 5 把三个误判合起来用。

练习 1 误判一 某 SaaS 公司有两类客户。大客户收入占比 70% → 50%，毛利率 62% → 64%；中小客户收入占比 30% → 50%，毛利率 42% → 45%。整体毛利率变动多少？拆成构成效应和率效应，并试试换一个替代顺序。

练习 2 误判二 创新项目复盘会上的汇报：“近两年结项的 12 个创新项目，平均 ROI 46%。”你要先补问哪个数？

练习 3 误判一 + 误判二 AI 助手给出下面两行代码。它们各替你做了什么口径决定？

```
df.groupby("区域")["毛利率"].mean()  
(df["回款日期"] - df["开票日期"]).dt.days.mean()
```

练习 4 误判三 客诉率最高的 15 家门店上月参加了服务培训，本月客诉率下降 30%。区域经理建议全面推广。你会怎样设计评估？

练习 5 综合 以第 5 章的复核指令为底稿，为你所在的业务写一份 AI 复核指令，并补一条行业专属口径（如零售的同店口径、SaaS 的 ARR 口径）。

迁移任务 回到工作中，挑一份最近的经营分析，用三句话过一遍：怎么合的？算了谁？跟谁比？记下哪一句出了问题。

解析

练习 1 整体毛利率 56.0% → 54.5%，跌 1.5 个点。先换结构、后换毛利率：构成效应 -4.0，率效应 +2.5。先换毛利率、后换结构：率效应 +2.3，构成效应 -3.8。两种顺序结果不同，这就是连环替代法“替代顺序影响结果”的性质。对称分解给出构成效应 -3.9、率效应 +2.4，不依赖顺序。无论哪种算法，结论都一样：两类客户的毛利率都在涨，整体下滑来自中小客户占比上升。

练习 2 同期立项的项目一共有多少个，包括中途终止的和仍在进行的。终止项目的 ROI 多数为负，它们不在“结项”名单里。正确口径是按立项队列统计：终止项目计入分母，在途项目单列，或按删失处理。

练习 3 第一行对每一行的毛利率做简单平均，没有按收入加权，大单和小单一票一权（误判一）；区域毛利率应为毛利合计 ÷ 收入合计。第二行里，未回款票据的回款日期为空，相减得到缺失值，被 `mean()` 默认跳过（误判二）；应先报告缺失条数，再改用固定观察窗口。

练习 4 门店是按“最差”入选的，回归均值会制造一部分下降。方案：从客诉率最高的 30 家里抽签选 15 家先培训，另 15 家下月培训；比较两组同期变化之差（双重差分），同时排查新品上市、节假日等对两组都起作用的因素。

练习 5 自查要点：三个误判各有至少一条可执行的指令；措辞是“披露”“报告”，不是“注意”；规定了无法确定时怎么处理；行业口径写清了分子、分母和时间窗口。

延伸

延伸阅读

两个视频里没讲的现象，和误判一同属“合并”的问题。

延伸 A 按平均需求做计划

禾川饮品 Q4 有一款新品，产能上限 10 万箱，每箱毛利 20 元。需求有三种情景，概率相等：6 万、10 万、14 万箱，平均正好 10 万箱。

按平均需求做计划： $20 \times \min(10, 10) = 200$ 万元。逐个情景算再取平均： $(20 \times 6 + 20 \times 10 + 20 \times 10) \div 3 = 173.3$ 万元。

需求 14 万箱的那个情景里，多出来的 4 万箱卖不出去。产能把好运气截掉了，坏运气却照单全收。按平均做计划，高估 26.7 万元，约 15.4%。

图 7

按平均需求做计划，毛利高估了 26.7 万元

三种需求情景概率相等 · 一档 = 5,000 箱 · 墨色 = 卖出 · 橙色虚档 = 超过产能、卖不出的需求



三种情景毛利平均 173.3 万元 · 按平均需求 10 万箱算是 200 万元

F7 STACKED RUNGS · 平均值缺陷 · 教学虚构数据

数学上这是琴生不等式 (Jensen, 1906)：当结果是输入的凹函数时，函数的期望不超过期望的函数。Savage (2009) 把它叫作平均值缺陷的强形式 (the flaw of averages, strong form)。与之相关的另一种情形，是两列相关的数各自取平均再相乘： $E[XY] = E[X] \cdot E[Y] + \text{Cov}(X, Y)$ ，协方差为负时，平均值之积大于积的平均，协方差为正时相反。实务里的对策是做情景测算或蒙特卡洛模拟：用分布做计划，不用单点。

延伸 B 调一家店，两级都进步

A 类门店 4 家，月销售额 500、400、300、120 万元；B 类 3 家，90、80、70 万元。把 120 万那家降为 B 类：

	调整前平均	调整后平均
A 类	330.0	400.0
B 类	80.0	90.0
7 家整体	222.9	222.9

两级的平均都涨了，整体一分没动。条件只有一个：被挪动的对象，低于原组平均、高于新组平均。这个现象由 Feinstein 等（1985）命名为威尔·罗杰斯现象。原始场景是肺癌的分期迁移：研究比较了 1953–1964 年和 1977 年两批患者，后一批做了大量新式影像检查，查出了原本看不到的微小转移，许多人因此被改判为更晚的分期。他们比原晚期组的病情轻、比原早期组的病情重，于是各分期的生存率同时提高，而每位患者的实际预后都没有变化。BP 遇到“组织调整后各单元指标普遍改善”时，先看总盘子。

书单（中译本）

卡尼曼, D. (2012). 思考，快与慢（胡晓姣, 李爱民, 何梦莹, 译）. 中信出版社. 第 17 章讲回归均值。

珀尔, J., & 麦肯齐, D. (2019). 为什么：关于因果关系的新科学（江生, 于华, 译）. 中信出版集团. 讲辛普森悖论与因果结构。

哈夫, D. (2009). 统计数字会撒谎（廖颖林, 译）. 中国城市出版社. 统计陷阱类读物的经典。

附录

术语表与参考文献

术语表

中文	英文	位置
自动化偏差	automation bias	第 1 章
辛普森悖论（尤尔-辛普森效应）	Simpson's paradox (Yule–Simpson effect)	误判一
混杂因素	confounder	误判一
构成效应 / 率效应	composition (mix) effect / rate effect	误判一
因素分析法（连环替代法）；Kitagawa 分解	factor analysis by chain substitution; Kitagawa decomposition	误判一
价格—数量—结构分析	price–volume–mix (PVM) analysis	误判一
右删失；Kaplan-Meier 法（乘积极限法）	right censoring; Kaplan–Meier estimator	误判二
幸存者偏差	survivorship bias	误判二
样本选择偏差	sample selection bias	误判二
意向性治疗原则	intention-to-treat (ITT) principle	误判二
回归均值	regression toward the mean	误判三
反事实	counterfactual	误判三
双重差分法；平行趋势假设	difference-in-differences; parallel trends assumption	误判三
内部效度威胁	threats to internal validity	误判三
平均值缺陷；琴生不等式	the flaw of averages; Jensen's inequality	延伸 A
威尔·罗杰斯现象；分期迁移	Will Rogers phenomenon; stage migration	延伸 B

参考文献

APA 第 7 版。案例公司与数据为教学虚构。

Blyth, C. R. (1972). On Simpson's paradox and the sure-thing principle. *Journal of the American Statistical Association*, 67(338), 364–366. <https://doi.org/10.1080/01621459.1972.10482387>

Brown, S. J., Goetzmann, W., Ibbotson, R. G., & Ross, S. A. (1992). Survivorship bias in performance studies. *The Review of Financial Studies*, 5(4), 553–580. <https://doi.org/10.1093/rfs/5.4.553>

Card, D., & Krueger, A. B. (1994). Minimum wages and employment: A case study of the fast-food industry in New Jersey and Pennsylvania. *The American Economic Review*, 84(4), 772–793.

Charig, C. R., Webb, D. R., Payne, S. R., & Wickham, J. E. (1986). Comparison of treatment of renal calculi by open surgery, percutaneous nephrolithotomy, and extracorporeal shockwave lithotripsy. *British Medical Journal (Clinical Research Ed.)*, 292(6524), 879–882. <https://doi.org/10.1136/bmj.292.6524.879>

Elton, E. J., Gruber, M. J., & Blake, C. R. (1996). Survivorship bias and mutual fund performance. *The Review of Financial Studies*, 9(4), 1097–1120. <https://doi.org/10.1093/rfs/9.4.1097>

- Feinstein, A. R., Sosin, D. M., & Wells, C. K. (1985). The Will Rogers phenomenon: Stage migration and new diagnostic techniques as a source of misleading statistics for survival in cancer. *The New England Journal of Medicine*, 312(25), 1604–1608. <https://doi.org/10.1056/NEJM198506203122504>
- Galton, F. (1886). Regression towards mediocrity in hereditary stature. *The Journal of the Anthropological Institute of Great Britain and Ireland*, 15, 246–263. <https://doi.org/10.2307/2841583>
- Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, 47(1), 153–161. <https://doi.org/10.2307/1912352>
- Hollis, S., & Campbell, F. (1999). What is meant by intention to treat analysis? Survey of published randomised controlled trials. *BMJ*, 319(7211), 670–674. <https://doi.org/10.1136/bmj.319.7211.670>
- Jensen, J. L. W. V. (1906). Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta Mathematica*, 30, 175–193. <https://doi.org/10.1007/BF02418571>
- Julious, S. A., & Mullee, M. A. (1994). Confounding and Simpson's paradox. *BMJ*, 309(6967), 1480–1481. <https://doi.org/10.1136/bmj.309.6967.1480>
- Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282), 457–481. <https://doi.org/10.1080/01621459.1958.10501452>
- Kitagawa, E. M. (1955). Components of a difference between two rates. *Journal of the American Statistical Association*, 50(272), 1168–1194. <https://doi.org/10.1080/01621459.1955.10501299>
- Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(4), 989–998. <https://doi.org/10.1037/a0015729>
- Microsoft. (n.d.). AVERAGE function. Microsoft Support. Retrieved October 12, 2026, from <https://support.microsoft.com/en-us/office/average-function-047bac88-d466-426c-a32b-8f33eb960cf6>
- pandas development team. (2026). `pandas.DataFrame.mean` (Version 3.0) [Software documentation]. Retrieved October 12, 2026, from <https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.mean.html>
- Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors*, 52(3), 381–410. <https://doi.org/10.1177/0018720810376055>
- Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, 39(2), 230–253. <https://doi.org/10.1518/001872097778543886>
- Pearl, J. (2014). Comment: Understanding Simpson's paradox. *The American Statistician*, 68(1), 8–13. <https://doi.org/10.1080/00031305.2014.876829>
- Richland, L. E., Kornell, N., & Kao, L. S. (2009). The pretesting effect: Do unsuccessful retrieval attempts enhance learning? *Journal of Experimental Psychology: Applied*, 15(3), 243–257. <https://doi.org/10.1037/a0016496>
- Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5), 688–701. <https://doi.org/10.1037/h0037350>
- Savage, S. L. (2009). *The flaw of averages: Why we underestimate risk in the face of uncertainty*. John Wiley & Sons.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.
- Simpson, E. H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 13(2), 238–241. <https://doi.org/10.1111/j.2517-6161.1951.tb00088.x>
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124–1131. <https://doi.org/10.1126/science.185.4157.1124>
- Yule, G. U. (1903). Notes on the theory of association of attributes in statistics. *Biometrika*, 2(2), 121–134. <https://doi.org/10.1093/biomet/2.2.121>

谢谢阅读。

关于作者

杨浩 August

一个喜欢研究 AI 的财务 BP

邮箱	1215868565@qq.com
----	-------------------

GitHub	github.com/August06exe
--------	------------------------

电话	131 2260 5260
----	---------------

Copyright © 2026 杨浩 (August) . 保留所有权利。

本讲义及配套视频的文字、图表与插图均为作者原创。欢迎学习交流；转载、改编或商用，请先联系作者。