

波动的语法

给投资者的统计学入门

王建硕

费曼式写法 · 只用加减乘除和开根号

导读

给投资者的统计学入门

先别打开这本书的任何一章，先回想一个画面。

你打开券商 App，看到持仓的数字在跳。昨天收盘是两万零四十，现在是两万零一百一十二——又变了，一万九千九百六十七。红的绿的，一秒一变。你的心跳跟着它走，手指悬在屏幕上，脑子里只有一个问题：**这一下，是正常的，还是出事了？**

现在给你一个反直觉的答案：**大概率，这两下跳动根本不配拥有这个问题。**就像你不会问「今天为什么比昨天低一度」——一度之内的气温起伏是天气的正常呼吸，而绝大多数日子的股价波动，也是呼吸。可惜没人教过我们怎么区分呼吸和窒息，于是我们把每一次呼吸都当窒息处理：查新闻、找原因、做操作。焦虑是真的，损耗是真的，唯独那个「原因」多半是虚构的。

这本书想给你的，就是区分二者的能力。它叫统计学，但请忘掉你对这门课的全部记忆——这里不考试、不背公式、不碰微积分，全书只用加减乘除和开根号。统计学在我们手里的用途只有一个：**给意外定价。**先量出一只资产平常的呼吸幅度（它会变成一个叫 σ 的数），再拿今天的跳动去除以它——商不够大，就锁屏去干别的。

你会跟着我手算一遍所有东西：从一只真实高波动个股的日线数据里算出均值、方差、标准差；用 68/95/99.7 三个数给任何波动「定级」；学会日波动率乘 $\sqrt{252}$ 变年化的根号法则，以及为什么偏偏是根号；然后看到同一个 -2% 在三种资产身上分别是呼吸、月度事件和历史课本。后半本是三件更硬的功夫：为什么分散投资被称为「唯一的免费午餐」（我们会手算出来）；为什么市场比教科书爱走极端（肥尾）；以及为什么冠军基金明年大概率平庸、手感火热多半是错觉（回归均值）。

我是一个工程师，建仓两万美元的那个晚上亲手把这些全算了一遍。我数学不算好，统计学早就还给了老师——这本书就是写给我自己这种人的：统计学没学好，但刚开始投资，而且足够聪明，知道「看不懂数字在说什么」迟早要付钱。

学完整本，你预测不了市场——没有任何诚实的书能让你预测市场。但你会拥有一件更稀缺的东西：**知道什么不需要解释。**在一个每条推送都想让你一惊一乍的时代，这是一种防御力。

现在，从数字为什么一直在跳开始。

第一章 · 数字为什么一直在跳

建仓那晚

我是一个工程师。几个月前，我把两万美元放进了一个指数基金组合——不是什么大动作，就是普通人开始投资的那种开始。钱转进去的那个晚上，我做了每个刚开户的人都会做的事：打开券商 App，盯着屏幕看。

数字在跳。20041。19987。20012。红的绿的，一秒一变。

然后我发现自己一直在问同一个问题：**这一下跳动，是正常呼吸，还是真出事了？**

跌 0.3% 的时候我问自己要不要慌。跌 1.2% 的时候我又问了一遍。第二天新闻推送「市场震荡」，我再问一遍。这个问题我答不上来——不是因为我不够聪明，而是因为我缺一门语言。跳动的数字是一种语言，而我不会读。

那晚我关掉 App，把数据导出来，拿加减乘除和开根号，把整件事算了一遍。这本书就是那晚算出来的东西。

统计学不是数学课

很多人对统计学的记忆是大学课堂：希腊字母、查表、考试前背公式，考完试还给老师。如果你是这种人，我们是同类。

但统计学本质上不是数学。它是一门语言，一门**给意外定价**的语言。

想想你已经在生活里熟练使用这门语言的时刻：今天 18 度，昨天 19 度，你不会问「为什么降温了」——一度之内的气温起伏是天气的正常呼吸，不值得一个解释。但如果你家孩子平时体温 36.5 度，今天早上 38.2 度，你立刻就会问为什么、要不要去医院。同样是数字变了，一个配拥有「为什么」，一个不配。

大白话：统计学干的事，就是给每个数字配一把体温计。先量出这个东西平时的「正常体温」和「正常波动范围」，然后看今天这一下有没有超出去。没超出去，就是呼吸；超出去太多，才轮到「为什么」登场。

所以学统计学不是为了预测市场。这点必须在一开始就说死：**这本书教不了你明天涨还是跌，没有任何诚实的书能教你这个**。它教你的是另一件、而且可能更有用的事——判断什么不需要解释。投资者的绝大部分焦虑，都花在给噪音找原因上。今天跌了 0.8%，去找新闻，找到一条，心满意足——其实那 0.8% 很可能根本不需要任何新闻，就像今天比昨天低一度不需要任何新闻一样。

全书只用一把尺

这门语言的语法出奇地简单，核心就一个动作：**拿这次的跳动，去除以它平时的呼吸幅度**。

得到的商，就是这本书反复使用的度量衡，叫「几个 σ 」（读作「几个西格玛」）。它回答的问题是：这次波动相当于多少个「日常呼吸幅度」？

- 不到 1 个 σ ：纯呼吸，连看都不用看。
- 两三个 σ ：值得抬头看一眼，但也仅仅是抬头。
- 五六个 σ 以上：稀有事，这时候「为什么」才真正登场。

你会在第七章看到这把尺的威力：同样是单日跌 2%，放在一只高波动个股身上是 0.4 个 σ ——纯粹的呼吸，问原因等于问今天为什么比昨天低一度；放在全市场指数基金身上是 2.2 个 σ ——值得抬头看一眼；放在债券基金身上是 7 个 σ ——历史事件。**同一个数字，在三种资产身上是三种事**。不懂 σ 的人看的是 2%，懂 σ 的人看的是它在谁的身上。

这本书怎么带你走

全书只用加减乘除和开根号，没有微积分，没有希腊字母轰炸—— σ 大概是唯一的希腊字母，而且你会先把它当一把尺用熟，再知道它的学名。

每一章的结构都一样：先用家常比喻搭梯子——体温、天气、身高——再落到真实的证券数据上，带你亲手算一遍。第二、三章从一串真实的日收益率出发，一步步算出均值、方差、标准差，讲清方差为什么要平方、标准差为什么要开回来；第四章给你钟形曲线和 68/95/99.7 法则；第五、六章把 σ 磨成速查卡，并讲清日波动率怎么年化；第七章是贯穿全书的案例总验收；第八、九章手推「分散投资是免费午餐」的数学出处；第十、十一、十二章是三剂解毒药——肥尾、样本量、回归均值，专治「学了前半本就想速成」的冲动。

期权定价、随机微积分、任何需要微积分的推导，这本书明确不讲。不是它们不重要，是你判断「这次跳动配不配拥有为什么」用不上它们。

读完这本书，你再打开券商 App，看到数字在跳，你会做一个以前不会做的动作：心算一下「这有几个 σ 」，然后大多数时候，心安理得地锁屏去干别的。这个动作值多少钱？对大多数投资者来说，值他们亏掉的很大一部分钱。

第二章 · 先给「平常的一天」画像

先把「跳动」变成可以算的东西

App 里跳的是价格，但价格本身没法直接比较——900 块的股票跳 9 块，和 90 块的股票跳 9 块，完全是两回事。所以统计学家不看价格，看日收益率：今天收盘价比昨天收盘涨了百分之几。

大白话

大白话：日收益率 = (今天收盘 - 昨天收盘) ÷ 昨天收盘。昨天 100 块、今天 103.2 块，日收益率就是 +3.2%。它把不同价格的股票拉到同一把尺上——从此我们只说百分比，不说块。

下面要开始手算了。案例是全书要反复用的一只真实的高波动个股——为了保护你的注意力而不是保护它，我们给它起个外号叫「**疾风**」。我从它的日线数据里截了连续十个交易日，日收益率如下（单位：%）：

+3.2 -4.1 +6.8 -2.5 +0.9 -5.6 +7.4 -1.8 +2.7 -4.0

看一眼这串数字。注意你的直觉已经开始工作了：上蹿下跳，大的有六七个点，小的不到一个点。「上蹿下跳」这个词，就是第三章要把它变成数字的东西。但这一章先回答一个更朴素的问题：**它平常的一天，长什么样？**

均值：平常到底有多平

你早就会算均值——加起来，除以个数。全班平均身高、家里过去一个月的平均气温，都是这么算的。对疾风这十天：

$3.2 - 4.1 + 6.8 - 2.5 + 0.9 - 5.6 + 7.4 - 1.8 + 2.7 - 4.0 = 3.0$

十天加起来一共涨了 3.0%，除以 10，**均值 = +0.30%**。

先别嫌这个数小——它小得非常有信息量。把它翻译成成人话：这只一天动不动跳四五个点的股票，**平常的一天几乎是不涨不跌的**。+0.3% 的均值，在一天 ±4% 的上下翻飞面前，约

等于零。

这是新手最反直觉的一课，值得停下来细嚼：你打开 App 看到的那场惊心动魄的拉锯战，多空双方杀得天昏地暗，一天的均值却安静得像没有发生。想象一个酒鬼走钢丝——他在钢丝上方两米、下方两米之间疯狂摇摆，但一小时后你再看，他还在原来那个位置附近。**摇摆的幅度和移动的距离，是两件完全独立的事。**均值管的是「移动了多远」，而第三章的主角管的是「摇摆得多凶」。

均值是个沉默的家伙

看均值还有什么用？有，但主要是反着用。

- **它能帮你戳破「感觉」。**这十天里疾风有五天是跌的，有两三天跌得还挺狠，你的直觉会说「这票一直在跌吧」。均值说不：十天净涨 3.0%。感觉会被大跌的日子绑架，均值不会。
- **它对「平常的一天」几乎无话可说。**这是均值最大的局限。两个人的平均体温都是 36.6 度，一个一直稳定在 36.6，一个在 35 度和 38 度之间来回跳——均值一模一样，人完全是两种人。投资里这个区别叫风险，而均值对它完全失明。

大白话

大白话：均值告诉你这趟车最终往哪开、开多快，但一个字都没提路上颠不颠。对投资来说，颠不颠往往比开多快更重要——因为绝大多数人不是被方向甩下车的，是被颠簸甩下车的。

给「平常」画像，才谈得上「意外」

回头想想第一章那把体温计。要判断今天 38.2 度算不算发烧，你得先知道两件事：这孩子平常多少度（均值），以及他平常的体温在多大范围内浮动（下一章的主角）。只有均值、没有波动范围，体温计只有刻度没有「正常区间」，38.2 度你还是不知道要不要紧。

所以这一章我们完成了画像的第一半：疾风平常的一天 $\approx$ 不涨不跌 (+0.3%)。第二半——平常的一天「浮动多大」——就是这串数字里 +6.8、-5.6、+7.4 这些大跳要告诉我们的事。怎么把「上蹿下跳」变成一个干净的数字，为什么不能直接平均（正负会抵消）、为什么要平方、平方完又为什么要开回来——这是第三章的全部内容，也是全书的第一个硬功夫。

本章带走的就一句：日收益率的均值管方向，不管颠簸；而十个交易日的疾风告诉你，方向可以很平静，颠簸可以很狂野，两件事互不拖欠。

第三章 · 方差：把「离平常多远」加起来

「上蹿下跳」要变成一个数字

第二章结尾留了个活：疾风平常一天均值 +0.3%，但它真实的每一天都在 +7% 到 -6% 之间翻飞。怎么把「翻飞」本身变成一个干净的数字？

最自然的想法：看看每天离均值有多远。这叫离差，就是当天的收益率减去均值（0.3%）：

+2.9 -4.4 +6.5 -2.8 +0.6 -5.9 +7.1 -2.1 +2.4 -4.3

涨的日子是正的，跌的日子是负的。接下来顺手把它们加起来求个平均，就得到「平均波动」了，对吧？试一下：

$2.9 - 4.4 + 6.5 - 2.8 + 0.6 - 5.9 + 7.1 - 2.1 + 2.4 - 4.3 = 0.0$

零。这不是巧合，这是均值的定义决定的——均值就是所有数字的「平衡点」，正的超额和负的超额永远恰好抵消。**直接平均离差，永远得零，对任何数据都一样。**此路不通。

为什么要平方

问题出在正负号上。涨 4% 和跌 4%，对「波动多大」这个问题是同等的贡献，可它们的离差一正一负，加在一起就互相消灭了。我们需要一个办法，把负号抹掉。

抹负号的办法有两个：取绝对值，或者平方。两个都是诚实的办法，统计学选了平方，理由很实在：

- **平方让大波动「罪加一等」。**一天 -5.9%，平方后贡献 34.8；一天 +0.6%，平方后只贡献 0.36。大日子的权重不是线性放大，是平方放大。这恰好符合风险的真实形状：一天亏 6% 对你的伤害，远大于两天各亏 3%——就像从三楼摔一次和从一楼摔三次，不是一回事。
- **平方在数学上好相处。**后面第六章的 $\sqrt{t}$ 法则、第九章的「分散是免费午餐」，推导能走通，靠的都是平方的一个漂亮性质（和勾股定理是亲戚）。绝对值在这些推导里会处处卡壳。

于是动手，把十个离差挨个平方（单位是「%的平方」，先别管它别扭）：

8.41 19.36 42.25 7.84 0.36 34.81 50.41 4.41 5.76 18.49

加起来：192.10。除以天数 10，得 **19.21**。这个数就是方差：每一天离均值多远，平方之后再平均——「平均的平方距离」。

大白话

大白话：方差 = 每天偏离平常的程度，先平方（抹掉负号、重罚大日子），再取平均。它衡量的正是均值看不见的那半张画像：颠簸。

为什么要开回来

方差有个别扭之处：单位。收益率的单位是 %，平方之后方差的单位是「%²」——百分之百平方，谁也没长过这个直觉。说「疾风的方差是 19.21 %²」，就像说两个地方的「平方距离」是 25 平方公里——数字本身没错，但你没法拿它和任何真实距离直接比。

解决办法就是当初怎么上去的，现在怎么下来：**开平方，把单位开回来**。

$$\sqrt{19.21} \approx 4.4$$

这个数叫标准差（记作 σ ，读西格玛），单位重新变回了 %。现在一切顺了：疾风的日收益率均值是 +0.3%，日标准差是 4.4%——两个数单位相同，可以直接对着看：**它平常一天不动，而平常的「一步」是四个多点**。方向是零，步幅是四。这就是这只股票的完整画像。

顺带交代一句：我用了一整年的日线数据（约 250 个交易日）算过同一个东西，得到日 $\sigma = 4.5\%$ ——我们这十天手算的 4.4% 已经很接近了。全书后面说到疾风，一律用全年数据的 4.5%。

（一个诚实的脚注：讲究的教科书算方差时除以 $n-1$ 而不是 n ，我们这里是 9 而不是 10，算出来标准差是 4.6% 而不是 4.4%。差别很小，原因是第十一章的主角——样本量。现在忘掉它，不耽误任何事。）

你已经会算波动率了

回头看这三步：**减均值，平方，平均，开根号**。就这四下，没有任何更高级的东西。你刚刚手算出来的 4.5%，在券商研报的措辞里叫「该标的年化前日波动率约为……」——对，华

尔街管标准差叫波动率，是同一把尺的两个名字。整份研报里你最需要看懂的那个数，你现在会算了。

但一把尺光有刻度还不够，还得知道刻度之间的概率：离均值 1 个 σ 以内，大概多久发生一次？2 个 σ 以外呢？这需要一个形状——一座钟。第四章见。

本章带走的：方差 = 平方距离的平均（平方为了抹负号、重罚大日子）；标准差 = 方差开根号（开回来为了单位回到 %、能和收益率直接比）。 σ 从此是本书的第一度量衡。

第四章 · 钟形曲线与 68/95/99.7

一座钟是怎么长出来的

先看一个和投资毫无关系的东西：成年男性的身高。

中国成年男性的平均身高大约 170 厘米。大多数人都挤在 165 到 175 之间；150 以下或 190 以上的，你在地铁里一天也遇不到几个。为什么身高会长成这样？因为身高不是「一个原因」决定的，是几百个独立的小原因——几百个基因各加一点减一点、小时候的营养、睡眠——叠加出来的总账。

而「许多个独立的小随机因素相加」，在数学里有一个近乎魔术的性质：**不管每个小因素自己长什么形状，它们加起来的总账，都会趋向同一座钟形曲线**——中间高、两边低、左右对称。这个性质有个正式名字，叫中心极限定理，但你只需要记住那个画面：游乐场里的弹珠钉板，每颗弹珠在每一排钉子处随机往左或往右掉，成千上万颗弹珠落到底部，堆出来的轮廓永远是一座钟。

大白话

大白话：单个因素是乱的，大量独立的小乱加在一起，反而整齐得可以预测。乱到一定程度，就成了秩序——整座统计学大厦都盖在这块地基上。

股票一天的收益率，恰好也是这种东西：一天里有成千上万笔买卖，每一笔都是一个小小的推力，方向各异、互不相干，加总起来就是当天的涨跌幅。所以日收益率的分布大致是钟形的——这条曲线叫正态分布，名字的意思是「正常情况下的分布」。（「大致」两个字是伏笔，第十章会回来收。）

68 / 95 / 99.7：值得背下来的三个数

钟形曲线最实用的性质是：**它每个刻度上的概率是固定的**，不随股票、不随年份改变。以均值为中心、以 σ 为步长量出去：

- 落在 $\pm 1\sigma$ 以内的概率约 **68%**——大约三天里有两天；

- 落在 $\pm 2\sigma$ 以内的概率约 **95%**——二十天里大约有十九天；
- 落在 $\pm 3\sigma$ 以内的概率约 **99.7%**——三百天里大约只有一天会出圈。

这就是为什么 σ 是一把尺而不只是一个数：一旦你知道了一只资产的 σ ，这三个百分比你就可以直接往上套，就像知道了一件衣服的尺码就知道它大概合身。

拿疾风套一遍（日 $\sigma = 4.5\%$ ，均值 ≈ 0 ）：

- $\pm 1\sigma = \pm 4.5\%$ ：一年约 250 个交易日里，应该有一百七十来天涨跌幅在 4.5% 以内；
- $\pm 2\sigma = \pm 9\%$ ：大约每二十天（一个月）才有一天波动超过 9%；
- $\pm 3\sigma = \pm 13.5\%$ ：大约一年都未必轮到一天。

我拿疾风那一年的真实数据对过账：比如全年有 63% 的交易日波动超过 2%——钟形曲线对此的预测是约 66%（2% 相当于 0.44σ ，曲线说 0.44σ 以内的概率约 34%，以外就是 66%）。63% 对 66%，大致合身。一只以狂野著称的股票，日常的呼吸节奏居然被一座抽象的钟描得八九不离十——这正是这三个数值得背下来的原因。

σ 的另一面：它和均值的关系

留意一个之前埋下的对照：疾风的均值只有 $+0.3\%$ ， σ 却有 4.5%——均值连 σ 的零头都不到。这意味着在「一天」这个尺度上，信号完全被噪音淹没：某一天是涨是跌，几乎纯看运气，趋势要在几百天的积累之后才探出头来。**时间尺度越短，市场越像抛硬币。**记住这个感觉，第六章讲年化的时候它会变成一条公式。

也要把丑话说在前面：钟形曲线是个极其有用的近似，但它只是近似。它说 5σ 大约三百万个交易日才有一次（折合约一万四千年），而真实市场几年就来一次。差距在哪、怎么办，是第十章的事。在那之前，先用好它——因为下一章你会发现，就这么三个百分比，已经够造一件随身工具了。

本章带走的：大量独立小推力相加必成钟形；钟形曲线上 $\pm 1\sigma/2\sigma/3\sigma$ 的概率恒为 68%/95%/99.7%；知道 σ 就知道每个幅度多久该出现一次。

第五章 · 几个 σ 才配拥有「为什么」

一个动作：除一下

前四章给了你全部零件，这一章把它们拧成一件随身工具。动作只有一个：

拿这次的涨跌幅，除以这只资产的日 σ 。

得到的商就是「几个 σ 」。疾风昨天跌了 2%： $2 \div 4.5 \approx 0.4$ 个 σ 。做完了，就这么简单。没有计算器也行，心算就够——这正是这件工具的设计目标：打开 App 的三秒钟内给出判断。

除完之后怎么读？把第四章那三个百分比翻译成行动指南，就是下面这张卡：

σ 速查卡

不到 1σ → 纯呼吸。钟形曲线说一年里三分之一的日子都在 $\pm 1\sigma$ 之外，这种幅度天天有。不配有「为什么」，连新闻都不用点开。 $1\sigma \sim 2\sigma$ → 正常偏大一点。一个月总会来几次。喝了口咖啡，放下。 $2\sigma \sim 3\sigma$ → 值得抬头看一眼。二十天到三百天才有一次。看看是不是真有消息，但默认答案仍然是「没有」。 3σ 以上 → 稀有。一年未必一次。这时候「为什么」通常真有货：政策、财报、黑天鹅。 5σ 以上 → 历史事件。别查卡了，去看新闻头条——但要注意第十章：它来得比钟形曲线承诺的勤。

这张卡可以背下来。它的全部依据就是上一章的 68/95/99.7——你不是在背玄学，是在背三个有出处的概率。

「问原因」是有成本的

为什么值得给波动设一个「起问点」？因为给噪音找原因，不是无害的消遣，是有真实成本的动作。

成本一：**你一定会找到**。财经媒体的生产机制决定了，每天收盘后必须给当天的涨跌配一个原因——「受加息预期影响」「因避险情绪升温」。如果那天波动只有 0.5σ ，这个原因就是

对噪音的文学加工。问「今天为什么比昨天低一度」，气象小编也能给你写出三百字，但写得出来不代表存在一个值得知道的原因。

成本二：**找到原因你就会行动**。而行动——卖出、买入、加仓、止损——是要交手续费、交价差、交税的，更重要的是，对不到 2σ 的波动采取行动，统计上约等于对着呼吸开枪。绝大多数散户的收益损耗，不来自选错股，来自对噪音过度反应。

大白话

大白话：速查卡省的不是「搞懂市场」的力气——那力气本来就不存在。它省的是你每天两次的焦虑、每月几次的无效操作、每年一笔可观的摩擦成本。知道什么不需要解释，本身就是一种收益。

用卡之前，先确认尺是对的

这张卡只有一个使用前提：**分母要用对**。 σ 是「这只资产自己的」日常呼吸幅度，不是别人的。用疾风的 4.5% 去量债券基金，或用债券基金的 0.3% 去量疾风，判断会荒谬到可笑——而第七章你会看到，这个「用错尺」的错误，正是绝大多数人看行情软件时的默认状态：屏幕上所有资产的涨跌幅都用同一种红色绿色显示，仿佛 2% 对它们是同一件事。

还有一个前置疑问可能已经在挠你了：App 和研报里印的波动率通常是「年化」的（比如疾风的年化波动率是 71.8%），而速查卡用的是日 σ 。年化数和日数怎么换算？为什么不是乘 365 也不是乘 252，而是乘一个奇怪的开根号？这个换算本身藏着一个漂亮的直觉，值得单独一章——下一章就是它。

本章带走的：涨跌幅 $\div$ 日 σ = 几个 σ ；不到两三个 σ 的波动是噪音，问原因等于问今天为什么比昨天低一度。这张卡可以背，但尺必须是这只资产自己的。

第六章 · $\sqrt{t}$ 法则

研报上的 71.8% 是什么意思

疾风的资料里写着：年化波动率 71.8%。我们手算过它的日 σ 是 4.5%。这两个数说的是同一件事的不同时间尺度，中间隔着一个换算。问题是：怎么换？

直觉的第一个答案是直接乘：一年约 252 个交易日， $4.5\% \times 252 \approx 1134\%$ 。这个数字显然是荒谬的——波动率怎么会超过 1000%？第二个答案听起来合理些： $4.5\% \times 252$ 太离谱，那就……别乘？可什么都不乘，日尺度和年尺度又怎么对上？

正确答案是乘 $\sqrt{252} \approx 15.9$ ： $4.5\% \times 15.9 \approx 71.4\% \approx 71.8\%$ ，和研报的数对上了。开根号，不多不少。为什么是开根号？这是本章唯一的问题，而它有一个不需要任何证明的直觉答案。

醉汉走了 252 步

回到第二章那个醉汉。他站在路灯下，每一步随机往任意方向迈一米，互不相关。问：走了 252 步之后，他离路灯多远？

不会是 252 米。因为步子是随机的，向东的一步大概率会被向西的一步抵消一部分。也不会是 0——抵消不可能那么干净。真实的答案大约是 $\sqrt{252} \approx 16$ 米。**随机行走的位移，随步数的平方根增长。**

这就是整条法则的全部直觉：股价一年的旅程，就是 252 个随机交易日的叠加。日子和日子之间大致互不相关（今天涨不预示明天涨），所以它们像醉汉的步子一样大量互相抵消。252 天的累积波动不是 252 份，而是 $\sqrt{252}$ 份——约 16 份。

大白话

大白话：如果每天的波动是「同向叠加」的，年就波动就是日的 252 倍；如果每天的波动完全互相抵消，年波动就是 0。真实情况是随机——介于两者之间，而数学证明这个「介于」恰好落在开根号上： $\sqrt{252} \approx 16$ 倍。既不叠加，也不清零，抵消到只剩平方根。

平方相加，斜边开根号

想再深一层的人可以记住这个画面（第三章埋过的伏笔）：方差——也就是平方后的波动——是可以直接相加的。互不相关的日子，其方差像勾股定理里的两条直角边：一年的方差 = $252 \times$ 一天的方差。你平时看到的 σ 是斜边，不是直角边，所以最后要开一次根号： $\sigma_{\text{年}} = \sqrt{(252 \times \sigma_{\text{日}}^2)} = \sqrt{252} \times \sigma_{\text{日}}$ 。方差按时间乘，标准差按时间开根号。这句话可以背。

这条法则的三个日常用法

- **正着用：日 σ 换年化。** 全市场指数基金的日 $\sigma \approx 0.9\%$ ，年化 = $0.9\% \times 15.9 \approx 14.3\%$ ；债券基金日 $\sigma \approx 0.3\%$ ，年化 $\approx 4.8\%$ 。研报上的年化数从此可以反查。
- **反着用：年化换日 σ 。** 看到「某基金年化波动率 20%」，除以 15.9，日 $\sigma \approx 1.26\%$ ——再代入第五章的速查卡，你就知道它跌 2.5% 才算「抬头看一眼」(2σ)，跌 1% 纯属呼吸。对任何资产，你三秒钟就能配好它的体温计。
- **任意尺度：**周化乘 $\sqrt{52} \approx 7.2$ ，月化乘 $\sqrt{12} \approx 3.5$ ，季度化乘 $\sqrt{4} = 2$ 。永远是开根号。

根号里藏着长期主义的数学

最后留意一个漂亮的对照。均值——那个微小的方向——随时间是**直接乘**的：日收益 0.3%，252 天大致累积 252 份（这就是复利前身的线性部分）。而波动——那个巨大的噪音——随时间只按**平方根**涨。时间拉长，方向按 252 倍长大，噪音只按 16 倍长大。

这正是第四章那句「时间尺度越短，市场越像抛硬币」的公式版：一天里，疾风的均值（0.3%）只有 σ （4.5%）的十五分之一，信号被噪音彻底淹没；拉长到一年，累积的方向约 75%，累积的噪音约 71%——信噪比改善了 $\sqrt{252} \approx 16$ 倍，从 1:15 扳到大致打平。噪音还在，但信号终于探出了头。**长期持有在数学上的全部含义，就是等到按 t 增长的东西跑赢按 $\sqrt{t}$ 增长的东西。**

一句诚实的提醒： $\sqrt{t}$ 法则的前提是「日子之间互不相关」。大部分时候这够真；但恐慌的日子里市场会「记性变好」——暴跌之后往往跟着暴跌。这个例外有多大，第十章讲肥尾时再算。

本章带走的：日 σ 乘 $\sqrt{252} \approx 15.9$ 得年化波动率；开根号的原因是随机的日子互相抵消到只剩平方根（方差按时间乘，标准差按时间开根号）；方向按 t 长、噪音按 $\sqrt{t}$ 长，这是长期主义的数学内核。

第七章 · 同一个 -2%，三种资产三种事

2024 年某个普通交易日的三块屏幕

想象同一天收盘后，三个人各自打开 App，看到的都是「-2.0%」：

- 第一个人持有疾风——我们那只高波动个股，日 $\sigma = 4.5\%$ ，年化波动率 71.8%；
- 第二个人持有全市场指数基金（就是我那两万美元组合的主力），日 $\sigma = 0.9\%$ ，年化约 14.3%；
- 第三个人持有债券基金，日 $\sigma = 0.3\%$ ，年化约 4.8%。

同一条红字新闻，同一个人文标题「市场下挫 2%」。现在，用第五章的动作，一人一除：

疾风： $2 \div 4.5 \approx 0.4\sigma$ 指数基金： $2 \div 0.9 \approx 2.2\sigma$ 债券基金： $2 \div 0.3 \approx 7\sigma$

三个商，三种事。逐个看。

疾风的 0.4σ ：连呼吸都算不上用力

对疾风来说，单日 -2% 是 0.4 个 σ 。第四章对过账：这只股票一年里 63% 的交易日波动都超过 2%。换句话说，-2% 在它身上一周大约出现一两次，比它呼吸重的日子还有的是。

持有疾风的那个人如果为这一天焦虑、去搜新闻、找到了一条「分析师下调评级」的解释并若有所思——他被骗了一个晚上。 0.4σ 的波动不配拥有任何解释，就像今天比昨天低一度不配拥有一篇气象分析。**对高波动资产的每一次小波动找原因，是把自己的注意力按市价卖给了噪音。**

指数基金的 2.2σ ：抬头看一眼，然后多半坐下

对全市场指数基金来说，-2% 是 2.2 个 σ 。钟形曲线说， 2σ 以外的事大约二十天有一次， 2.2σ 大约一个月一次。这不是呼吸了，值得抬头看一眼。

「抬头看一眼」具体是什么动作？查一下是不是真有宏观消息——加息、就业数据、地缘事件。有，知道了，然后**坐下，什么都不做**。因为你买的是整个市场的长期均值回归能力（第六、十二章），一次 2σ 级别的回撤在长期曲线里连个凹痕都算不上。没有消息，也坐下——可能只是市场在消化它自己的情绪。

这就是速查卡上「 $2\sigma\sim3\sigma$ 」档的设计意图：它给你的是知情权，不是行动令。一个月一次的事，不配改变一个以十年为单位的计划。

债券基金的 7σ ：出事了，而且是大事

对债券基金来说， -2% 是 7 个 σ 。 7σ 是什么概念？按钟形曲线，这事大约四千亿个交易日才轮到一次——从恐龙灭绝那天开始天天开盘，到今天也轮不上一次。也就是说：如果这一天真的发生了，结论几乎只有一个——**不是运气，是世界变了**。利率政策剧变、信用体系受冲击、或者持有的债券出了信用事件。这时候不是「抬头看一眼」，是立刻搞清楚发生了什么，因为你对这只基金的全部预期（日 $\sigma\ 0.3\%$ 的安稳）已经被证伪了。

真实世界里这样的事确实发生过：2022 年英国养老金危机期间，以「稳」著称的英国国债几天内走出七八个 σ 的波动，逼得央行进场救市。安稳资产的极端波动，从来都不是噪音，是警报。

行情软件为什么不告诉你这些

回头看那三块屏幕：App 用完全相同的红色、相同的字号、相同的推送语气，向你报告了三件性质完全不同的事——一次呼吸、一次月度事件、一次历史事件。**行情软件对所有资产用同一把尺显示，而波动的语法要求每只资产用自己的尺衡量**。这是工具的错，但现在你没法怪它了，因为你会自己除了。

大白话

大白话：以后看到任何「暴跌 $X\%$ 」的标题，先问一句：这是谁身上的 $X\%$ ？分母不同，新闻不同。同一个 -2% ，是疾风的星期二、指数基金的月度事件、债券基金的历史课本。

顺带说一句我自己：我那两万美元组合的主力是指数基金（日 $\sigma\ 0.9\%$ ），配了一部分债券基金（ 0.3% ）。为什么是这两个、它们放在一起会发生什么——「放在一起」这四个字是统计

学的另一件大礼物，需要两章才讲得完。接下来两章，我们手算它。

本章带走的： -2% 不是一个事件，是三个事件—— 0.4σ 的呼吸、 2.2σ 的抬头、 7σ 的警报。波动的意义 = 幅度 $\div$ 这只资产自己的 σ ，除此没有第二把尺。

第八章·协方差：两件事是不是一起动

分散的老话，缺一个数字

「鸡蛋不要放在一个篮子里」——这句话你听过一百遍。但老话没有回答关键问题：为什么分开放就更稳？稳多少？放在什么样的篮子里才算数？要回答这些，得先有一个衡量「两件事是不是一起动」的数字。这就是本章的主角，而第九章会用它推出投资里最重要的一条公式。

先用家常例子热身。冰淇淋店和火锅店，一个夏天火一个冬天火，两家店的流水几乎总是反向走；两条街对门的两家奶茶店，则大概率同涨同跌——商圈人流多的时候两家都忙。如果你选一家投资，两家各自都很颠簸；但如果让你各买一半，「冰淇淋+火锅」的组合明显比「奶茶+奶茶」睡得安稳——前者互相错开，后者一起挨打。**「错开」这件事，可以算出来。**

协方差：每天的「同向程度」加起来

回到第三章的老朋友：离差——每天离自己的均值多远。协方差的想法朴素得可爱：**把 A 今天的离差和 B 今天的离差乘在一起，然后对所有天取平均。**

大白话

大白话：两只资产同一天都比自己的平常强，乘积是正的；同一天都比平常差，负乘负，乘积还是正的；一个强一个差，乘积是负的。把这些乘积平均下来——正得多，说明它俩爱同向动（协方差为正）；负得多，说明爱对着干（协方差为负）；正负乱炖抵消完，说明各走各的（协方差接近零）。

动手算。A 是只小疾风式的股票，B 是只债券型基金，八个交易日的日收益率（%）：

A: +2.0 -3.0 +4.0 -1.0 +1.0 -2.0 +3.0 0.0 (均值 0.5%) B: +0.3 +0.5
-0.2 +0.4 -0.3 +0.1 +0.6 -0.4 (均值 0.125%，手算按 0.1% 取整)

第一步，各自的离差（减去自己的均值）：

A: +1.5 -3.5 +3.5 -1.5 +0.5 -2.5 +2.5 -0.5 B: +0.2 +0.4 -0.3 +0.3
-0.4 0.0 +0.5 -0.5

第二步，逐天相乘：

+0.30 -1.40 -1.05 -0.45 -0.20 0.00 +1.25 +0.25

注意第二天：A 大跌 3.5 个离差，B 反而比平常强 0.4，乘积 -1.40 ——「股票挨打、债券接住了」的一天，在协方差里是一笔重重的负贡献。八个乘积加起来 $= -1.30$ ，除以 8：**协方差 $= -0.16$** 。负数，说明这八天里它俩总体上在对着干——股票差的日子债券好。

相关系数：把答案压进 -1 到 $+1$

协方差有个毛病：大小没谱。 -0.16 算大算小？它和资产的 σ 纠缠在一起，没法横向比。解决办法和第七章一样：**除以各自的尺**。把协方差除以两只资产各自的 σ ，就得到相关系数（记作 ρ ，读「柔」）：

$$\rho = \text{协方差} \div (\sigma_A \times \sigma_B) = -0.16 \div (2.29 \times 0.36) = -0.16 \div 0.82 \approx -0.2$$

（ σ_A 和 σ_B 就是第三章的手艺：A 的方差 $= 42 \div 8 = 5.25$ ， $\sigma_A = \sqrt{5.25} \approx 2.29$ ；B 的方差 $= 1.04 \div 8 = 0.13$ ， $\sigma_B \approx 0.36$ 。）

相关系数永远落在 -1 到 $+1$ 之间，读法固定：

- **$\rho = +1$** ：完全同步，像同一家公司的两个分身（两家对门奶茶店）；
- **$\rho \approx 0$** ：各走各的，互不相关；
- **$\rho = -1$** ：完全镜像，你涨我跌严丝合缝（完美的冰淇淋+火锅）。

我们算出的 -0.2 ，是「轻度对着干」——这大致就是股票和债券在真实市场里的关系：多数日子各走各的，恐慌的日子里还常常反向（钱从股票逃进债券）。而同属股票的组合，比如疾风和另一只科技股， ρ 往往在 $+0.6$ 到 $+0.9$ ——同向，但不同步。

为什么「不同步」是宝

到这里你只是多会算了一个数，还没看到它的威力。威力在下一章：把 ρ 不高的两样东西各放一半，组合的波动不是两个波动的平均，而是**比平均还小**——小多少，由我们刚算的协方差决定。那句「分散是投资里唯一的免费午餐」，说的就是这个差额，而它完全出自一条我们能手推的公式。

本章带走的：协方差 = 两只资产每天离差的乘积再平均，正=同向、负=反向、零=无关；相关系数 = 协方差 $\div$ 两个 σ ，永远落在 -1 到 $+1$ ，是可以横向比较的「同步率」。

第九章 · 免费的午餐

第八章那个 -0.16 ，值多少钱

上一章我们算出：股票 A 和债券基金 B 的协方差是 -0.16 ，相关系数 -0.2 ——它俩轻度对着干。这一章回答那个最实在的问题：各买一半，组合的波动到底是多少？

你可能脱口而出：A 的 σ 是 2.29，B 的是 0.36，一半一半，组合的 σ 就是平均数 $(2.29 + 0.36) \div 2 \approx 1.33$ 吧？

不对。真实答案比这个小。小出来的那一块，就是「免费午餐」。我们把它手推出来。

一步代数：平方打开

组合每天的收益率就是两半相加： $\text{组合} = 0.5A + 0.5B$ 。问组合的方差，按第三章的定义走：先算每天的离差。组合的离差也很简单——就是两半各自离差的相加： $(0.5a + 0.5b)$ ，这里 a 、 b 是 A、B 当天的离差。

方差 = 离差的平方再平均。把这个小括号平方打开（初中代数： $(x+y)^2 = x^2 + y^2 + 2xy$ ）：

$$(0.5a + 0.5b)^2 = 0.25a^2 + 0.25b^2 + 0.5ab$$

现在对每一天都这样算，再取平均。注意三件事： $0.25a^2$ 的平均 = A 的方差 $\times 0.25$ ； $0.25b^2$ 的平均 = B 的方差 $\times 0.25$ ；而 ab 的平均——逐天离差相乘再平均——**这正是第八章定义的协方差**。于是：

$$\text{组合方差} = 0.25 \times \sigma_A^2 + 0.25 \times \sigma_B^2 + 0.5 \times \text{协方差}$$

这就是著名的两资产组合方差公式的一半一半版。推导只用了「平方打开」和「平均可以拆开算」，没有更多了。代入我们的数字：

$$\begin{aligned} \text{组合方差} &= 0.25 \times 5.25 + 0.25 \times 0.13 + 0.5 \times (-0.16) &= 1.31 + 0.03 - 0.08 \\ &= 1.26 &\text{组合 } \sigma &= \sqrt{1.26} \approx 1.12 \end{aligned}$$

1.12，不是 1.33。组合的波动比「两个波动的平均」小了 16%。这 16% 没有付出任何代价——不是选股选出来的，不是择时择出来的，纯粹是「把对着干的东西放在一起」这个动作本身白送的。这就是诺贝尔奖得主马科维茨说的那句话的数学出处：**分散化是投资里唯一的免费午餐**。午餐就是公式里第三项：它可以是负的。

大白话

大白话：前两项是「你自己带来的颠簸」，永远为正；第三项是「两个资产互相拉扯的账」，协方差越负，它减掉得越多。分散的全部秘密，就是去搜罗第三项为负（或接近零）的资产。

ρ 从 +1 滑到 -1：一顿午餐的完整菜单

把数字放大到年度尺度看得更清楚。假设股票部分年 $\sigma = 20\%$ ，债券部分年 $\sigma = 10\%$ ，各放一半。波动的加权平均恒为 15%，而真实组合 σ 随 ρ 变化：

$\rho = +1.0$ （完全同步）→ 组合 $\sigma = 15.0\%$ 没有午餐 $\rho = +0.5$ （同向不同步）→ 组合 $\sigma = 13.2\%$ 省 1.8 $\rho = 0.0$ （各走各的）→ 组合 $\sigma = 11.2\%$ 省 3.8 $\rho = -0.5$ （轻度对着干）→ 组合 $\sigma = 8.7\%$ 省 6.3 $\rho = -1.0$ （完全镜像）→ 组合 $\sigma = 5.0\%$ 省 10

盯住两端看。 $\rho = +1$ 时，公式变成 $(0.5 \times 20 + 0.5 \times 10)^2 = 15^2$ ——第三项满额为正，组合波动等于加权平均，**一分午餐都没有**。这解释了「我买了十只科技股为什么还是分散失败」：十只 $\rho = 0.9$ 的股票，约等于一只大科技股。 ρ 越往 -1 走，第三项减得越狠；理论上 $\rho = -1$ 且配比合适，波动可以减到接近零——当然现实中不存在完美镜像的资产，真实世界里股票和债券的 ρ 在 -0.2 到 +0.3 之间浮动，对应的免费午餐大致就是上面中间三行的水平。

三句诚实的边界

- **午餐管的是颠簸，不是收益。**分散降低的是 σ ，不提高均值。它的正确读法是「同样的预期收益，更小的颠簸」或「同样的颠簸，更高的预期收益」——取决于你怎么搭配。
- **ρ 不是刻在石头上的。**它是历史数据算出来的平均关系，会变。尤其记住：恐慌来临时，很多平时不相关的资产会突然一起跌——「危机时所有相关性都奔向 1」。第十章还会回来收拾这个例外。

- **免费不等于万能。**分散抹掉的是「单个资产特有的颠簸」，抹不掉整个市场一起的颠簸。全市场指数本身已经是几百只股票的免费午餐套餐，但它照常有 14% 的年化波动——那是所有公司共享的部分，谁也分散不掉。

我那两万美元的组合——主力全市场指数基金、配债券基金——此刻你可以用新眼光看了：这不是拍脑袋的「稳健配置」，是 $\rho \approx -0.2$ 的第三项在替我天天省波动。

本章带走的：组合方差 = $0.25\sigma_A^2 + 0.25\sigma_B^2 + 0.5 \times \text{协方差}$ ；第三项为负时，组合波动小于各自波动的加权平均——差额就是免费午餐； $\rho = +1$ 没有午餐，买十只同涨同跌的股票不算分散。

第十章·肥尾：市场比钟形曲线更爱走极端

一万四千年一次的事，几年就来一次

第四章卖了个关子，现在收。钟形曲线说： $\pm 3\sigma$ 以内覆盖 99.7% 的日子， 5σ 以外的事一万四千年一遇。按这个说法，市场应该是一部相当温吞的机器。现在看三个真实的日子：

- **1987 年 10 月 19 日**，「黑色星期一」，标普 500 一天跌 20.5%。按当时指数约 1% 的日 σ 算，这是一个 **20σ** 事件。钟形曲线给它的概率，小数点后面要跟八十多个零——从宇宙大爆炸到今天每天开盘，也轮不上一次。它发生了。
- **2008 年 10 月**，金融危机最深处，标普 500 一个月内跌出好几次 7σ 、 9σ 级别的日子。
- **2020 年 3 月**，疫情恐慌，一个月内出现多个 5σ 以上的交易日，其中 3 月 16 日一天 -12% ，约 10σ 。

按钟形曲线，这些事加起来应该「永远不发生」。真实历史里，它们几年就来一回。**真实市场的分布曲线，两边比正态分布更厚**——极端值比承诺的多得多。这个现象的学名叫**肥尾**：钟还是那座钟，但两只尾巴比正态的肥。

尾巴为什么肥：日子之间不是不相干的

回想钟形曲线的两块地基：每天的波动由许多小推力合成（这块没问题），以及每天的日子互相独立（第六章 $\sqrt{t}$ 法则也用了它）。问题就出在第二块——市场在恐慌日子里是有记忆的。

恐慌会自我喂养：大跌触发止损盘，止损盘造成更大跌，更大跌触发更多杠杆强平；新闻把恐惧广播给所有人，所有人同时想卖出。于是一段日子里，波动不再是互相独立的醉汉步子，而是一串同向的冲锋。统计上这叫「波动聚集」——**大波动的日子爱扎堆来**。第九章埋的那句话也是同一回事：危机时，平时不相关的资产会突然一起跌，因为驱动它们的不再是各自的基本面，而是同一个东西——恐惧。

大白话：正态分布假设每天独立洗牌，真实市场的牌偶尔是连着的。平时你看不出区别——63% 对 66% 那种小出入；但尾巴上区别是致命的：正态说万年一遇，市场几年一遇。

那 σ 速查卡还有用吗

有，但要换个拿法。它依然是最好的「日常体温计」： $\pm 2\sigma$ 以内的世界——覆盖一年 95% 的日子——钟形曲线描得相当准，速查卡在那个区域完全可靠。它失灵的地方在远尾： 3σ 以外的「稀有度标价」，真实市场比卡上写的便宜得多—— 3σ 不是一年一遇，是一年几次； 5σ 不是万年一遇，是十年里总要来几回。

所以正确的姿势是：**用速查卡过滤噪音，但别拿它当保险**。具体来说，三条操作纪律：

- **判断「配不配拥有为什么」照旧用卡。** 0.4σ 是呼吸，这个结论不受肥尾影响。
- **给杠杆定价时，把卡扔了。** 如果你的仓位结构（杠杆、融资、满仓单一资产）是「 5σ 我就爆仓」，钟形曲线说你安全一万四千年，真实市场说你十年内大概率死一次。设计仓位时按「极端事几年必来」设防，而不是按「极端事万年一遇」裸奔。
- **看到极端日，先想到肥尾再想到末日。** 10σ 的日子来了，它多半不是宇宙的奇迹，而是肥尾的日常发作——市场过几天还在。1987 年那个 20σ 的黑色星期一之后，市场两年内创了新高。极端波动吓人的是幅度，不是后果。

本书立场的完整版

到这儿可以把第一章的立场说完整了。统计学是给意外定价的语言——但任何语言的标价都有一个置信区间。 σ 这把尺，在 $\pm 2\sigma$ 以内是钢尺，在远尾是皮筋。知道尺在哪里准、在哪里软，恰恰是「会用尺」的一部分。

接下来两章换个战场：同样的统计学，除了给波动定价，还能给「证据」定价。一个广告词 3 次点击说明什么？一只基金去年的冠军说明什么？答案是第十一章和第十二章。

本章带走的：真实市场是肥尾的—— 5σ 正态说万年一遇、实际几年一遇，因为恐慌日子彼此不独立、波动爱扎堆；速查卡照常在 $\pm 2\sigma$ 内用，但涉及杠杆和「绝不可能」四个字时，按肥尾而不是按钟形设防。

第十一章 · 多少次才算数

同一枚硬币，两个判决

怀疑一枚硬币有偏心，你抛了 10 次，7 次正面。70%——硬币有问题吗？

先别回答，换个场景：你抛了 1000 次，700 次正面。还是 70%。硬币有问题吗？

直觉这次分得清：前者说不定是运气，后者铁定有鬼。把直觉翻译成我们已经会的语言。抛硬币的「正面率」本身有一把 σ 尺： $\sigma \text{率} = \sqrt{(p \times (1-p) \div n)}$ ， p 是真实正面率（公正硬币是 50%）， n 是次数。对 10 次： $\sqrt{(0.25 \div 10)} \approx 15.8\%$ 。观测到的偏差是 $70\% - 50\% = 20\%$ ，只有 **1.3 σ** ——按速查卡，呼吸。对 1000 次： $\sqrt{(0.25 \div 1000)} \approx 1.6\%$ ，同样的 20% 偏差是 **13 σ** ——铁案。

同一枚硬币、同一个 70%，判决相反，唯一的变量是 n 。这就是本章的全部要点：**统计结论不看百分比，看百分比除以 $\sqrt{n}$ 之后还剩几个 σ 。**

为什么偏偏又是 $\sqrt{n}$

注意这个公式的形状——误差按 $\sqrt{n}$ 缩小。眼熟吗？第六章的醉汉回来了： n 次观测的误差就像 n 步随机行走，互相抵消到只剩平方根。所以平均 100 天的收益率，其不确定性不是一天的 $1/100$ ，是 $1/10$ ；想把精度再提一倍，数据要翻四倍。**信息是昂贵的，按平方根收费。**

大白话

大白话：样本量给「证据」定价的方式，和 σ 给「波动」定价是同一把逻辑：先算这个偏差相当于几个 σ ，再决定信不信。 n 小， σ 就大，什么结论都立不住； n 大， σ 就小，一点点偏差都是铁证。

三次点击的广告词

你在给自己的小店投广告。新文案上线第一天：展示 20 次，点击 3 次，点击率 15%。旧文案的历史点击率是 8%。换不换？

15% 几乎是 8% 的两倍，看起来很香。算 σ ： $\sqrt{(0.08 \times 0.92 \div 20)} \approx 6.1\%$ 。偏差 = $15\% - 8\% = 7\%$ ，刚过一个 σ 多一点 (1.1σ)。翻译成人话：就算新旧文案**毫无差别**，纯粹靠运气，20 次展示里点出 3 次这种事也是家常便饭。结论：继续跑，别换，也别庆祝。三次点击说明不了一个广告词，就像三天的天气说明不了气候。

要多少才算数？按 2σ 的老规矩反推：想让 7% 的偏差达到 2σ ， σ 得压到 3.5%， n 得从 20 涨到约 60——而且这只是「勉强信」的门槛。想稳稳区分 8% 和 15%，得几百次展示。数字不性感，但它是真的。

两天的 A/B 实验，和天天偷看的人

放大到正经的 A/B 实验。两个版本各跑了 200 次展示：A 版 24 次点击 (12%)，B 版 16 次 (8%)。4 个百分点的差距，翻倍的一半，群里已经开始讨论上线了。算两版之差的 σ （两版各有不确定性，公式里两项相加）： $\sqrt{(0.10 \times 0.90 \times (1/200 + 1/200))} = 3\%$ 。4% $\div$ 3% $\approx 1.3\sigma$ 。**两天的「胜出」，大概率是噪音**。同样的差距维持到各 10000 次展示时， σ 差缩到 0.42%，4% 就是 9σ ——同一个 4%，从噪音变成了结论。实验有没有说服力，看的不是差距多大，是差距有几个 σ ，而 σ 只随 $\sqrt{n}$ 缩小。

现在讲那个最阴的坑：**偷看**。实验设好了跑两周，但你每天忍不住打开看一次数据——第一天 B 领先 2σ ，心痒；第三天回落了；第七天 A 又领先 2σ ……只要某一刻撞上「 2σ 显著」就宣布胜利停掉实验，会发生什么？

想想醉汉。给他足够长的时间，他总会晃过任何一条线——哪怕他整体上哪儿也没去。数据也一样：就算两版真实效果完全相同，领先幅度这条曲线也会在噪音里晃来晃去，你每偷看一次，就是多买一张「恰好晃过 2σ 线」的彩票。统计上的后果很实在：规矩的「 2σ 、5% 假阳性」标准，在天天偷看的用法下会膨胀到三成左右——你以为自己在用 5% 的严格标准，实际在用 30% 的宽松标准。这是「系统性地把随机波动误当真效果」的传送带。

解药古老而简单：**先定样本量，再开始；没跑完，结论不存在**。中途的数据只许看、不许判。

这把尺，量基金也一样

把广告换成基金：某基金经理连续 4 个月跑赢指数 6 个百分点。厉害吗？月度收益率差的 σ 大约 4 个百分点（指数日 σ 0.9% 乘 $\sqrt{21} \approx 4.1\%$ ），4 个月平均的 σ 就是 $4.1 \div \sqrt{4} \approx 2$ 个百分点；而实际跑赢折合每月 1.5 个百分点——只有 0.7σ ，呼吸。要分辨一个基金经理是真有本事还是运气，按 $\sqrt{n}$ 的收费标准，往往需要十年以上的数据。下一章我们会看到，市场连这个分辨过程都常常等不及——它直接把去年的运气加冕为今年的冠军。

本章带走的：率的 $\sigma = \sqrt{(p(1-p)/n)}$ ，误差按 $\sqrt{n}$ 缩小、按平方收费；结论 = 差距 $\div \sigma$ ，差距相同、 n 不同、判决相反；天天偷看会把 5% 的假阳性膨胀成三成——没跑完的实验，结论不存在。

第十二章 · 回归均值

去年冠军的第二年

每年年初，财经媒体都有一个固定节目：盘点去年的冠军基金。去年的冠军今年表现如何？这个盘点的结果也几乎是固定的：大部分冠军，第二年泯然众人，不少直接掉进倒数。这不只是中国股市的现象——美国标普公司的长期统计（SPIVA 报告）年复一年地显示：前一年排名前四分之一的基金，下一年还能留在前四分之一的，只有两成上下，跟抛硬币差不多。

为什么？是冠军们集体堕落了么？都不是。是统计学里一条安静而无情的规律在收税：回归均值——极端的表现，下一次大概率往平常的方向回落。

表现 = 实力 + 运气

用一个梯子把它搭起来。班级里考第一名的学生，下一次考试还是第一名吗？多半不是，但他多半还是前几名。因为一次成绩 = 真实水平 + 当天发挥。能考第一的人，通常是「水平本来就高」**而且**「那天发挥超常」——两个条件叠加。下次考试，水平还在，但发挥是个随机数，大概率回归平常，于是名次回落。注意：**不需要任何「骄傲自满」的故事，光是「发挥是随机的」这一条，就必然推出冠军回落。**

基金一模一样：一年收益 = 经理的真实水平 + 当年的运气。能在几千只基金里拿年度冠军，几乎必然是两个都占了——有点水平，外加重仓的方向恰好撞上当年的风口。第二年，水平部分还在，运气部分重新掷骰子——按第一章的话：运气那部分不配有「为什么」，它就是噪音，而噪音的均值是零。于是冠军的平均下场，就是往均值回归。

大白话

大白话：凡是「实力+运气」共同决定的事，极端值都掺着运气，而运气不续费。下一次，实力留下，运气重掷——所以极端之后大概率平庸。这不是世事无常的感慨，是「运气均值是零」的推论。

「手感火热」的错觉

同一个规律的另一个分身：篮球里的「手感火热」。球员连进三球，解说员说他「手烫」，队友开始给他喂球。研究者把 NBA 几十年的投篮数据翻了个底朝天，结论很冷：连进三球之后，下一球的命中率并不比球员的平均命中率更高（Gilovich、Tversky 等人 1985 年的经典研究）。手感火热，是我们大脑在随机序列里强行读出的图案。

人脑天生讨厌「这段没有原因」。连进三球、连涨五天、连赢四把——随机序列里这种串本来就很常见（抛硬币 20 次，出现至少一次三连正面的概率接近八成，四连也接近一半），但我们每次都给它配一个故事：他状态来了、趋势形成了、时代变了。用这本书的语言说：**我们把 1σ 以内的连续噪音，反复误认为信号**。这与第五章的速查卡是同一条纪律的两种表述。

什么时候回归，什么时候别等回归

回归均值有个致命的误用方向，必须说清：**它只对「运气成分大」的序列成立**。

- 基金年度排名、球员手感、短期股价——运气占比巨大，极端表现大概率回归。
- 但「这家公司收入连涨十年」不是运气，是实力序列，别等它「回归」而做空它。
- 反过来，「这只股票跌了 80%，该反弹回归了吧」——股价没有必须回归的「均值」，公司可以归零。回归均值回归的是「运气成分」，不是「价格本身」。

区分的方法你已经有了：问一句「这个结果里运气占几成」。运气占比越高（样本越小、 σ 越大——第十一章），回归越确定；实力占比越高，越别把回归当闹钟。

回到那个跳动的数字

绕了一圈，回到第一章建仓那晚的屏幕。现在再看它，你手里的东西不一样了：

- 数字在跳，你先算它几个 σ ——不到两三个，锁屏，它不配拥有「为什么」（第五章）；
- 你知道每只资产有自己的尺，同一个 -2% 是三种事（第七章）；
- 你知道日 σ 乘 $\sqrt{252}$ 得年化，方向和噪音按不同的速度长大（第六章）；
- 你知道组合里对着干的资产在替你省波动（第九章）；
- 你也知道这把尺在远尾是皮筋，极端事件几年必来，仓位要按这个设防（第十章）；

- 你还知道运气均值是零，所以冠军会平庸、手感是错觉、短期的自己不说明问题（本章）。

这就是「波动的语法」的全部语法。它没让你更会预测市场——没人会。它让你把市场里 95% 的跳动归进「不需要解释」的抽屉，把注意力和操作留给真正剩下的 5%。在一个按你的焦虑收费的行业里，这个抽屉是付费墙外面为数不多的免费资产。

愿你的 App 安静，愿你的组合长久。

第十三章 · 波动从哪儿来

一个被晾了十二章的问题

回头看这本书，我们一直在用 σ ，却从没问过它从哪儿来。疾风的日 σ 是 4.5%，指数基金是 0.9%，债券基金是 0.3%——这些数字我是怎么知道的？拿历史收益率算的（第二、三章）。但这只回答了「 σ 是多少」，没回答「 σ 是谁定的」。为什么偏偏是疾风一天晃 4.5%，而不是 0.5%？**一只股票的波动性，到底是什么东西决定的？** 这个问题值得专门一章，因为答案能反过来告诉你：什么时候你手里的 σ 会悄悄失效。

先把答案的骨架说出来，再一节一节拆开：一只股票的波动 $\approx$ **这家公司价值的不确定性** $\times$ **放大器** $\times$ **交易环节的放大器**。三个因子，对应三个层次。

第一层：生意本身有多难预测

股价归根结底是对一家公司未来全部盈利的定价。所以第一个问题是：这家公司的未来盈利，好猜吗？

卖酱油的公司，今年卖一亿瓶，明年大概率还是一亿瓶上下——需求稳定、价格稳定、成本稳定，它的未来像一条直线，误差很小。而一家创新药公司，未来三年的收入几乎全押在一款还没获批的药上：批了，收入翻十倍；没批，归零。同样叫「公司」，前者未来的可能范围是一条窄缝，后者是一把张开的扇子。股价是给未来定价的，**未来的扇子张得越开，每天的定价就晃得越厉害**。这就是为什么公用事业、消费股的 σ 天然小，而生物科技、小盘成长股的 σ 天然大——不是交易员神经质，是生意本身的不确定性不同。

第二层：杠杆，把生意的晃动放大成股权的晃动

同一家公司、同一份生意，借钱和不借钱，股票的波动可以天差地别。关键在于：股东拿的是「剩下的」。

假设一家公司值 100 亿，全是股东的钱，没有负债。生意的价值波动 10%，股权就波动 10%。现在换成另一个结构：公司值 100 亿，其中 80 亿是借的债，股东只出了 20 亿。

生意的价值还是波动 10%——也就是 10 亿——但债主的钱是合同写死的，雷打不动还 80 亿，于是这 10 亿的晃动**全部**砸在那 20 亿股权上：股权波动 50%。生意一点没变，股票的 σ 放大了五倍。

这个放大器有个正式名字，叫**财务杠杆**——用借来的钱做本金的生意。它解释了为什么银行股的波动天生比它们的生意看起来更刺激：银行的资产负债率动辄九成，股权是薄薄一层垫子，垫在巨额债务下面，吸收全部晃动。租房时房东和银行的关系，搬到股票上就是债主和股东的关系——房价跌 10%，先亏完的永远是首付。

大白话

大白话：公司的价值晃动先还债，剩下的归股东。债越多，股东那层垫子越薄，同样的晃动砸上去，股价的振幅就越大。 σ 里有一大半是资本结构算出来的，不全是生意本身的。

第三层：交易环节——谁拿着它，好不好卖

就算价值的不确定性完全一样，交易本身还会再放大一次。两个机制：

- **流动性。**流动性就是「想卖的时候，能不能立刻找到人按公道价接走」。大盘股的买单卖单排着队，一笔大单砸下去，价格只挪一点；小盘股一天成交不了几笔，同样一笔卖单砸下去，得把价格压下一个台阶才有人接。清淡的盘口像浅水池，扔块石头水花四溅；深厚的盘口是深潭，石头进去悄无声息。所以小市值、冷门股的 σ 天然偏高。
- **信息的分歧。**有些公司被几百个分析师盯着，每个季度每个数字都被反复核对，大家心里对它的估值收敛在一个窄区间；有些公司没人研究、披露含糊，十个人心里有十个价。分歧越大，每条新消息引起的「重新定价」就越剧烈。新股上市头几天波动吓人，部分原因就是市场还没就「它值多少钱」吵出结果。

把三层合起来：再算一次疾风和债券

现在回到第七章那三块屏幕，这次不问「几个 σ 」，问「 σ 从哪来」：

- **疾风（日 σ 4.5%）：**单一业务、前景是张开的扇子，可能还带些杠杆，盘子小、散户多、消息一出就重新定价——三层全是放大项， σ 不大才怪。

- **指数基金 (日 σ 0.9%)**：一只指数打包几百家公司。各家公司的特殊消息——这家药没批、那家签了单——方向随机、互相抵消（第九章的免费午餐），剩下的只有所有人都躲不开的宏观消息。个体的不确定性被分散掉大半， σ 自然小一个数量级。
- **债券基金 (日 σ 0.3%)**：债券的未来现金流是合同写死的数字——扇子几乎没张开。不确定的只有利率变化和「会不会赖账」两件事，所以它天生是全社会最安静的资产——直到利率政策剧变那天，安静被打破，才有了那个 7σ 。

σ 的排名不是抽签抽出来的，是三个层次的因子一层层乘出来的。

最重要的推论： σ 是会变的

这个答案带着一个必须带走的警告：**决定 σ 的因子会变，所以 σ 本身会变**。公司加了杠杆、业务从卖酱油换成赌一款药、市场从风平浪静切换到恐慌模式——任何一个因子动了，旧的 σ 就过期了。真实市场里这甚至是常态：波动喜欢扎堆来（第十章的肥尾有一半就是这么来的），恐慌期所有股票的 σ 一起膨胀，平静期又一起收缩。

所以你拿过去 250 天算出来的那个 σ ，准确的读法不是「这只股票的脾气是 4.5%」，而是「**以它最近的生意、杠杆和盘口，在过去这段市场环境下，脾气是 4.5%**」。第五章教你的速查卡依然好用，但心里要留一个抽屉：当你知道这家公司发生了什么结构性变化时，别等数据慢慢反映出来——尺子可以先于刻度更新。

本章带走的：一只股票的波动性由三层决定——生意未来的可预测性、杠杆的放大倍数、交易环节的流动性与信息分歧。指数的 σ 小是因为分散抵消了个股消息，债券的 σ 小是因为现金流写死在合同里。而这三层因子都会变，所以 σ 不是常数，是「当前结构下」的脾气。