第三章
库存管理的统计学思维

统计学是通过搜索、整理、分析、描述数据等手段，认识客观现象规律性的方法论科学。它是应用数学的一个分支，主要通过利用概率论建立数学模型，收集所观察系统的数据，进行量化的分析、总结并进行推断和预测，来为相关决策提供依据和参考。

统计学中的数据分析能够用适当的统计分析方法，对收集的大量数据进行提炼、分析，从数据中发掘出能够完善供应链管理的规律，以帮助人们做出判断，并采取适当的决策和行动。

第一节 统计学是让数据说话

在库存管理上，涉及不少数据处理分析，因此非常适合使用统计学相关的知识帮助我们制定合理的策略。

一、你知道数据有哪些吗

进行库存管理的时候，我们会接触到各种各样的数据，比如最近三个月进库 1000 吨，第一季度平均在库是 250 吨，或者上周销售量是 160 件，又或者某电商网站，来自亚洲区域的点击有 5700 次，而来自欧洲区域的点击有 2300 次。不过并非只有 1000 吨、250 吨、160 件才叫数据，实际上，来自亚洲区域和来自欧洲区域的点击次数都可视为数据。

数据的类型因为采用的计量尺度不同，分为以下三种。

（一）分类数据

该类型数据只能从特定集合中取值，表示一系列可能的分类。它是对事物进行分类的结果，数据表现为类别，是用文字来表达的。分类数据是由分类尺度计量形成的。如针对某电商网站的点击情况，就可以根据区域的不同将“来自亚洲区域”和“来自欧洲区域”划分成亚洲和欧洲等分类。

（二）顺序数据

该类型数据只能归于某一有序类别的非数字型数据，属于定性数据或品质数据。它是由顺序尺度计量形成的。如职称上的高级、中级、初级等分类。

（三）数值型数据

该类型数据指按数字尺度测量的观测值，属于定量数据或数量数据。数值型数据是使用自然计量单位或度量衡单位对事物进行测量的结果，其结果为具体的数值，如身高的 1 米、1.2 米，或书本页数的 100 页、300 页。

二、连续和离散数据分布

数值型数据可以分为连续和离散两个类别。

离散数据是其数值只能用自然数、整数、计数单位等描述的数据，其数据的取值是不连续的，每个数值之间都有明确的间隔。

王家卫导演的电影《花样年华》在海内外获奖无数，同时，男女主角的精彩演出更是获得了大家的赞赏。

《花样年华》讲的是一个含蓄的爱情故事，大多数的电影片段都靠男女主角的语言、肢体表情来展现和推进，因此两人独处的画面频频出现。这样的镜头之中，对于人数，就只有 2 这个数值，不会存在 1.5 或者 1.75 这样的数值，因为人数只有 2 个，不会有 0.5 个梁朝伟或者 0.75 个张曼玉的存在。这个数据就是离散数据了。

连续数据是指在一定区间内可以取任意值。相邻的数值可进行无限分割。

《花样年华》这部典型的王家卫作品是边拍边构思的，并没有一个完整的剧本。王家卫最终把拍摄的所有片段剪辑成一个简单而充满内涵的故事。不过当中也剪掉了一些精彩片段，这些精彩片段没有被安排上映。其中男女主角在房间里跳舞这一段被剪掉了，甚是可惜。跳舞过程中，两人的距离渐渐拉近，近到 0.5 米，不过更加精准的话，可能是 0.53 米，甚至精确为 0.535125 米，这个距离值能够取到多少位小数，取决于量度的精度。这样的数值无限分割下去，就是连续数据了。当然，《花样年华》这个故事的结局很是遗憾，两人的距离纵使再近，都只是一个连续数据，而无法为零。

其他典型的连续数据，如身高，既可以是 183 厘米，也可以是 183.15 厘米，甚至精确为 183.151712 厘米。

而数据在统计图中的形状，则称为它的分布。因此，离散数据在统计图中的形状为离散分布，连续数据在统计图中的形状就是连续分布。

离散分布和连续分布分别如图 3-1、图 3-2 所示。

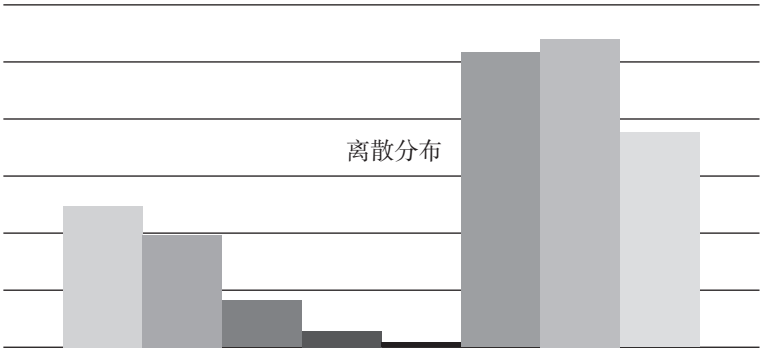


图 3-1

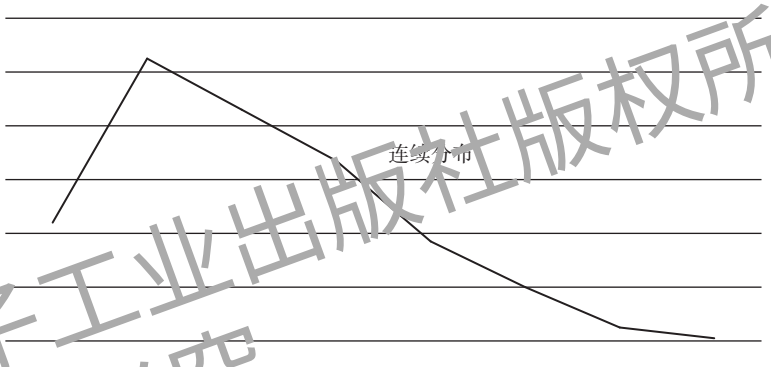


图 3-2

很多计算安全库存的公式，都是基于正态分布演化而来的，而正态分布是连续分布的一种。但并非所有分布都属于正态分布，这说明安全库存的计算存在不适应性，如果需求的情况不服从正态分布的话，计算就毫无意义。

区分连续数据和离散数据其实很简单。离散可以度量“可数事物”的多少，而连续则可以度量“不可数事物”的多少。

三、为什么我总是拖了全市平均工资的后腿

平均数在库存管理、计划管理中起着重要的作用，比如在不少公式的应用背景中都涉及正态分布（也叫高斯分布）。这个分布的重点就是关注平均水平，而把意外当作极端问题。

我们熟知的平均数实际上为算术平均数，除此之外还有几何平均数和调和

平均数等。算术平均数就是把所有数字加起来，然后除以数字个数。在统计学中，这个算出来的数称为均值，符号为 μ，它是一个希腊字母（读作“缪”）。

公式记为：

$$\mu=\frac{\sum x}{n}$$

$\sum x$ 表示的是所有数字的集合，n 指的是数字个数。

均值的计算在 Excel 中可以通过 AVERAGE 函数快速地得出结果。如图 3-3 所示为某企业过去 12 个月的月销售量，根据相关函数就可以计算出平均每月的销售量了。

图 3-3

在库存管理中，时间过于久远的历史数据不能直接用作参考数据，应对其根据时间远近做出合理的权重分配，越接近现在的数据越值得信赖，权重越高，从而利用全部数据并算出加权平均数。

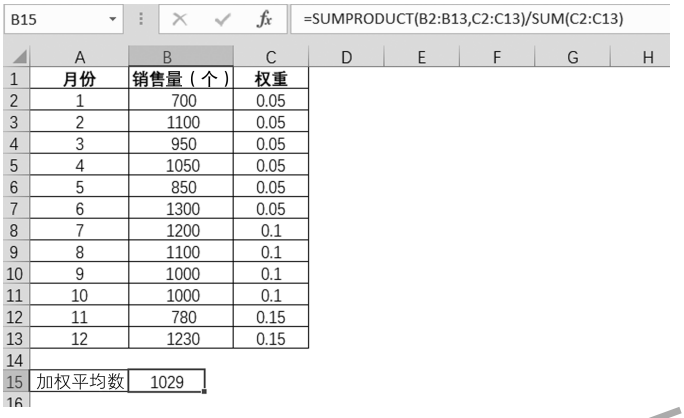
其公式为：

$$\mu=\frac{\sum f(x)}{\sum f}$$

其中 $\sum f(x)$ 表示各数字乘以对应的权重后再求和，而 $\sum f$ 表示权重之和。

某企业过去 12 个月的月销售数字，根据发生时间的远近应赋予不同的权重，前六个月权重均为 0.05，接下来的四个月权重均为 0.1，而最近的两个月权重则均为 0.15，权重合计为 1。使用 Excel 函数 SUMPRODUCT 和 SUM 可以快

捷地求出加权平均数，如图 3-4 所示。



	A	B	C	D	E	F	G	H
1	月份	销售量（个）	权重					
2	1	700	0.05					
3	2	1100	0.05					
4	3	950	0.05					
5	4	1050	0.05					
6	5	850	0.05					
7	6	1300	0.05					
8	7	1200	0.1					
9	8	1100	0.1					
10	9	1000	0.1					
11	10	1000	0.1					
12	11	780	0.15					
13	12	1230	0.15					
14								
15	加权平均数	1029						
16								

图 3-4

均值是一组常规样本大概率上最有代表性的统计量，比如哪一支篮球队选手的身高会更高一些，哪一个仓库的库存的放置时间更长一些，可以通过了解均值来了解样本的整体情况。

不过，均值却是个“坏家伙”，它可骗了不少人。

2021 年全国各省、自治区和直辖市税前平均工资如表 3-1 所示。该表公布的时候，不少人的朋友圈都是满满的吐槽，都是这么一句话：自己拖后腿了。

表 3-1

序　号	省 / 自治区 / 直辖市	2021 年税前平均工资（元）
1	上海	8513.16
2	北京	9758.13
3	天津	5896.18
4	重庆	6412.68
5	河北	5815.43
6	河南	5755.05
7	湖北	6166.08
8	湖南	6279.40
9	西藏	7195.38
10	海南	5658.60
11	青海	6024.34

续表

序　　号	省 / 自治区 / 直辖市	2021 年税前平均工资（元）
12	贵州	5835.58
13	江苏	6687.55
14	浙江	6570.13
15	广东	7015.52
16	广西	5641.82
17	山东	5782.01
18	山西	6201.25
19	四川	6421.14
20	新疆	5959.73
21	福建	6215.30
22	内蒙古	6231.76
23	陕西	5784.34
24	云南	5633.71
25	宁夏	5787.05
26	甘肃	5746.01
27	江西	6138.07
28	安徽	6220.21
29	黑龙江	5123.98
30	辽宁	5131.95
31	吉林	5178.62

一心要追上平均工资的员工，莫不抱着“给我涨工资”的心态，提出诉求。

然而老板们以一脸冷漠的表情反驳道：“身在福中不知福啊！我国还有 6 亿人月入不足 1000 元，你比那 6 亿人好多了！”

然而，事实上真的有那么多高工资的人吗？还是尚有 6 亿人月入不足 1000 元？

均值只是大概率上反映了整体情况，那是因为其样本有着特殊情况，并不能反映样本数据的真实特征。当一些特殊的数据，即非常高的工资收入纳入计算样本中时，最终得出的均值就无法描述整体的真实情况了。

引入中位数、众数作为参考值，连同均值，才能更加清晰地了解整体的真

实情况。

中位数是样本升序排列后最中间的数。根据数据个数的不同，计算方法分为两种。

当数据个数为奇数时，中位数即最中间的数，如果有 n 个数，则中间数的位置为 $(n+1)/2$，比如最近 5 天的销售数据为 10、20、30、40、50，那么中位数就是按大小排序后最中间的数。由于是 5 天的数据，最中间的数的位置就是 (5+1)/2=3，排序第 3 位的数就是 30 了。

当数据个数为偶数时，中位数为最中间两个数的均值，中间位置的算法是 $(n+1)/2$。比如最近 6 天的销售数据为 10、20、30、40、50、60，由于是 6 个数据，其中间位置就是 (6+1)/2=3.5，即排序后的第 3 位和第 4 位的中间，也就是这两个数的均值，因此就是 (30+40)/2=35 了。

均值容易受到特殊数据的影响，而中位数则不会，因此中位数能相对客观地反映出整体的情况。

刚才的最近 5 天销售数据，通过计算得出中位数和均值，都是 30，如图 3-5 所示。

10	20	30	40	50

中位数	30
均值	30

图 3-5

如果再加入前一期的销售数据 500，这个数据非常大，和最近 5 天的数据有着明显的差异，由此而得出的中位数和均值也相应地变化，如图 3-6 所示。

500	10	20	30	40	50

中位数	25
均值	108

图 3-6

均值因为这个特殊数据的加入，产生了较大的变化，并且得出的结果和这 6 个数据都相差甚远，而中位数则变化不大。中位数的计算和每个数据的位置有关系，即使有极端的特殊数据加入，也是排在两端的位置上，而不会插在