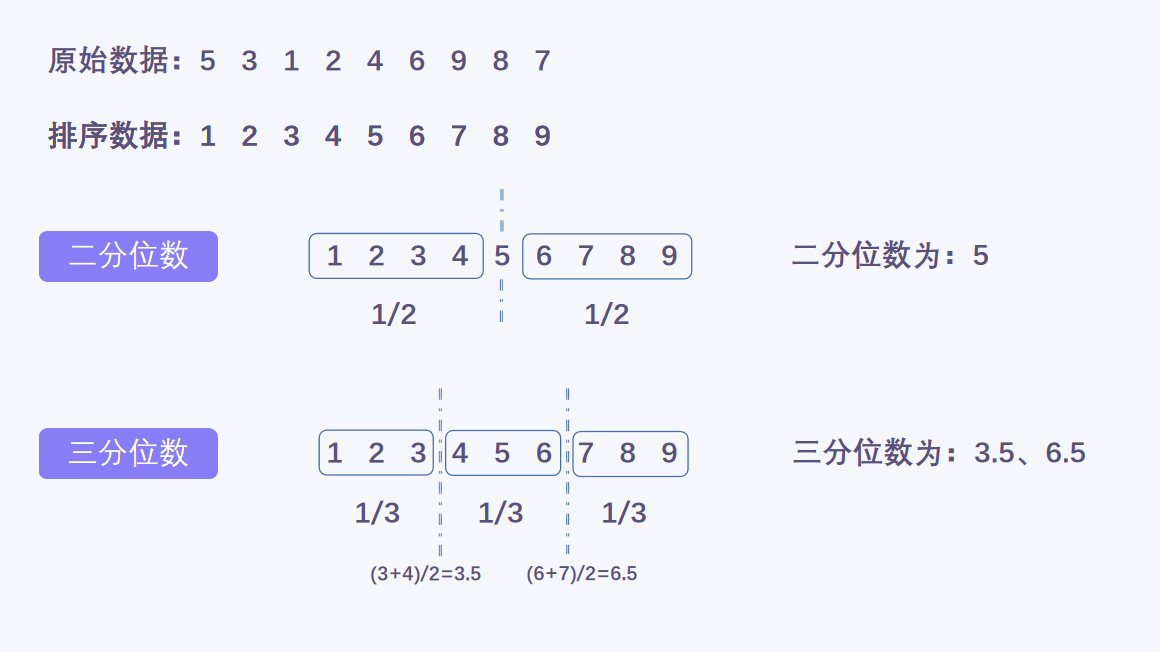
**分位数法**

**一、方法说明**

1. **概念含义**

分位数法，指通过对数据进行**排序**并且 **N 等划分**的方法，把数据划分为 N 个区域，产生得到 N-1 个 N 分位数。

2. **数学原理**



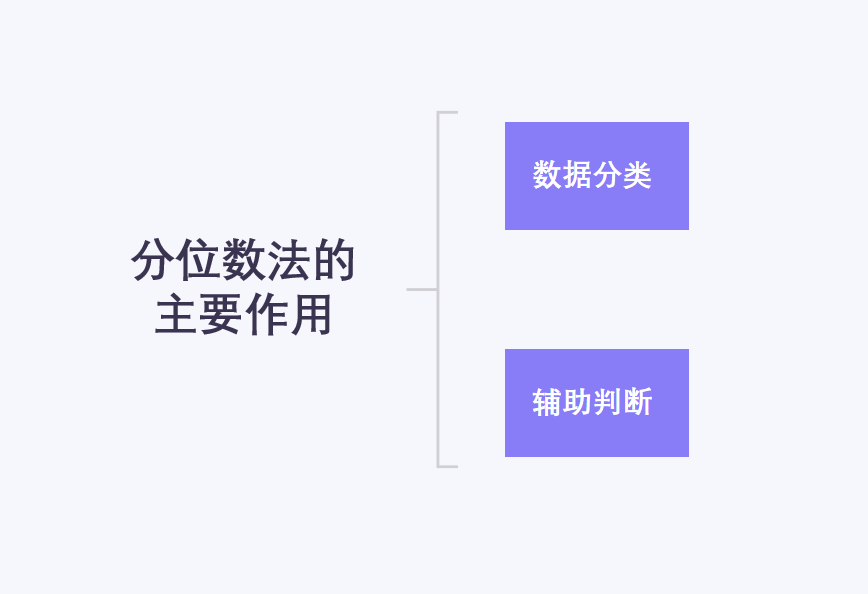
**中位数**是分位数中最常见的一种，由于它把排序数列分成了**相等的两部分**，因此也被称为**二分位数。**

|  |
| --- |
| **提示**：中位数是按顺序排序的一组数据中居于中间位置的数，代表一个样本、种群或概率分布中的一个数值，它可将数值集合划分为相等的上下两部分。 |

而其他分位数（例如三分位数、四分数等）的原理与中位数一致，将按从小到大排序的数值集合，划分为相等的 N 个部分，得到的 N-1 个分位数则分别是这 N 个部分的分界值。

|  |
| --- |
| **优势**：由于分位数是以它们在所有数值中所处的位置，来确定的全体数值的代表值，不受数值中极大值极小值的影响， 从而在一定在程度上提高了对分布数列的代表性。 |

3. **主要作用**



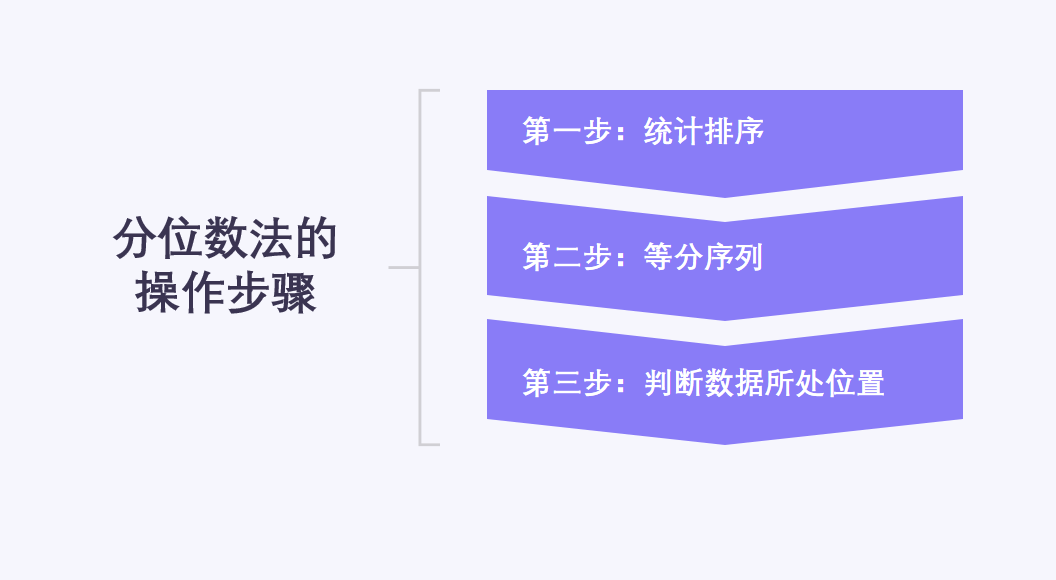
**1）数据分类**

由于排序的数列被划分为了相等的 N 个部分，因此可以根据这 N 个部分给数据进行分类，为不同部分的数据贴上不同的属性和“标签”。

**2）辅助判断**

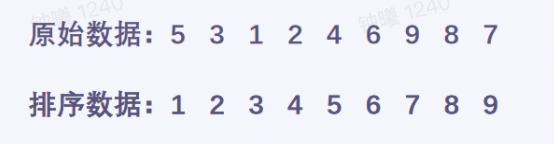
由于分位数即是相等的 N 个部分的分界值，它们是全体数值的代表性数值，因此可以将某一数据与它们进行比较，判断这一数据在全体数值中所处的位置。

4. **操作步骤**



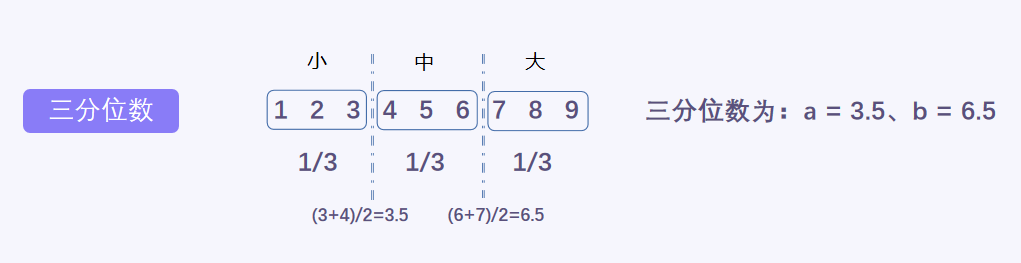
**1）步骤一：统计排序**

统计数列中的所有数据，将其从小到大进行排序。



**2）步骤二：等分序列**

以三分位数为例，将排序好的序列划分成比重相等的小中大三个区域，得到两个分位数 a，b（a 和 b 同时作为阈值，a为小中区域的分界值，b为中大两个区域的分界值）。



**3）步骤三：判断数据所处位置**

以三分位数为例，将某一数据与两个分位数 a=3.5，b=6.5 进行比较，从而判断其在整体数据中所处的位置。

* a）如，7.5 > b ；
* 7.5 在整体数据中处于“大”这个区域。
* b）如，b > 5 > a ；
* 5 在整体数据中处于“中”这个区域。
* c）如，0.5 < a ；
* 0.5 在整体数据中处于“小”这个区域。

**二、方法应用**

5. **交易经验**

**交易背景**

越来越多的北向资金购买 A 股，目前北向资金的成交额占比已维持在 10% 以上，影响力日渐加大。

而从历史数据来看，在北向资金单日净流入值为正的情况下，未来几天市场的平均涨跌幅以及上涨概率都有不错的表现。

在这种背景下，市场上逐渐诞生了一个交易经验，那就是北向资金大幅流入时，未来平均收益较高；北向资金大幅流出时，未来平均收益较低甚至为负。

因此，有一些声音认为，我们可以”抄北向资金的作业“，即跟随北向资金的动向进行买入卖出。

**交易想法**

|  |
| --- |
| **在北向资金大幅流入时，跟着买入**；**在北向资金大幅流出时，跟着卖出**。 |

6. **问题描述**

|  |
| --- |
| **大幅流入**和**大幅流出**的**判断标准**是什么？（我们**如何定义**”大幅流入”和“大幅流出”呢？） |

7. **量化思路**

**难点**

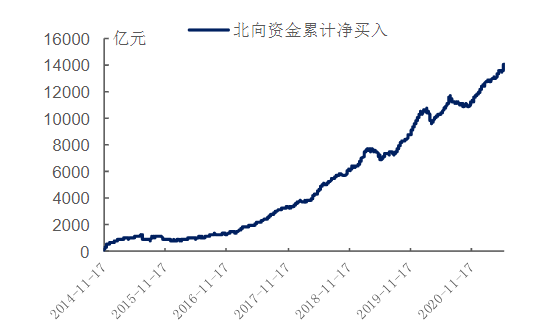
流入和流出我们可以很轻松地找到数值 0 作为它们的分界线，因为对于净流入值，正的代表流入负的代表流出，这个是毋庸置疑的。

但是对于”大幅流入”和“大幅流出”，我们**难以找到一个固定的具体数值来定义它的标准**。

**原因**

数据是处于不断地变化中的，**使用固定的数值来判断流入流出的幅度很容易失效**。

例如在 2014 ~ 2016 年期间，北向资金的累计净买入增长缓慢，而在 2016 年之后快速增长。那么 2014 年时的当日净流入值 10 亿放在 2018 年来看，就已经不可同日而语了，如果我们仍用同一个固定数值作为大幅流入的标准，这个标准也会失去参考性。



**思路**

|  |
| --- |
| 既然固定的数值容易失效，而且判断当天的数据情况需要以历史数据为参考依据，那么我们是否可以把思路转向判断当天数据在历史数据中的位置呢？ |

|  |
| --- |
| **从判断具体数值的大小，转变为判断其在历史数据中的位置**。以此定义大幅流入和大幅流出的标准**。** |

在统计学中，分位数常用来反映其在所有数据中所处的位置，且具有数据分类和辅助判断的作用。

而投资中的情绪和动作大体可以分为三种：**乐观买入**、**中性观望**和**悲观卖出**。其中的乐观买入和悲观卖出，可以对应大幅流入买入，大幅流出卖出的交易想法。

因此，可使用**分位数法**中的**三分位数法**，将排序后的指标历史数据划分成比重相等的**上中下**三个区域，以此来对应投资中的三种情绪，并且作为大幅流入还是大幅流出的判断标准。

那么该指标当日数据落在哪个区域，即代表它在历史数据中是处于哪个位置，对应属于哪种投资情绪，是属于大幅流入还是大幅流出。

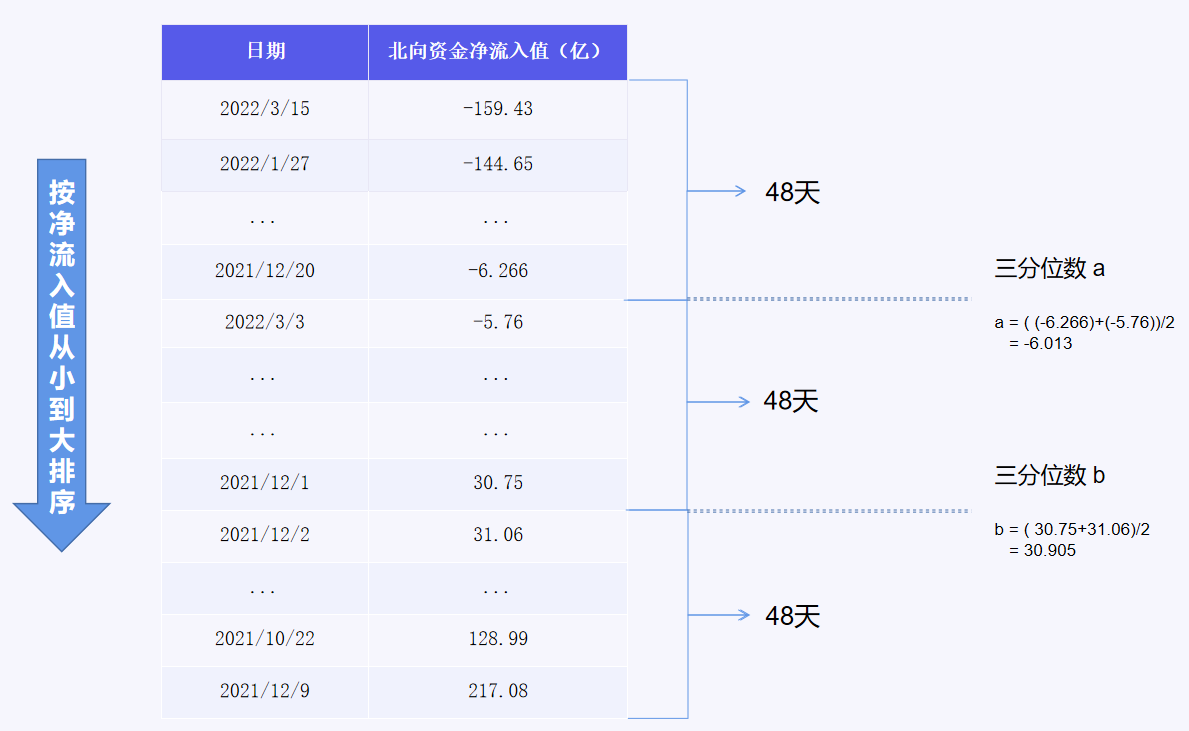
8. **场景应用**

|  |
| --- |
| **提示：**以下示例以今日交易日 2022-05-17 为例，假设北向资金的所有净流入历史数据为过往8个月，即截止到 2021-09-16。 |

收盘后，统计当日（2022-05-17）之前所有北向资金的净流入值数据，将其**从小到大进行排序。**



三等划分序列，将排序好的历史净流入值数据划分成比重相等的三个区域，得到两个分位数 a 和 b。（这两个分位数即可作为北向资金净流入值历史数据的代表性数值。）



将当日的净流入值与这两个分位数进行比较，即可知道当日数据在历史中所处的位置。



**1）当日净流入值 > b：**

* 例如当日净流入值为 **40** 亿，40 **> b**；
* 当日的净流入值在历史数据中处于小中大的**“大”**这个区域，即从历史上来看，它属于**净流入值数值较大**，**流入较大**的区域。
* 对应的投资情绪是**乐观情绪**，执行**下日买入**的操作。

**2）b > 当日净流入值 > a：**

* 例如当日净流入值为 **20** 亿，**b > 20 > a**；
* 当日的净流入值在历史数据中处于小中大的**“中”**这个区域，即从历史上来看，它属于**净流入值数值中等**，**流入流出均较小**的区域。
* 对应的投资情绪是**中性情绪**，**下日观望**。

**3）当日净流入值 < a：**

* 例如当日净流入值为 **-10** 亿，**-10 < a；**
* 当日的净流入值在历史数据中处于小中大的**“下”**这个区域，即从历史上来看，它属于**净流入值数值较小**，**流出较大**的区域。
* 对应的投资情绪是**悲观情绪**，执行**下日卖出**的操作。

**三、总结延伸**

9. **总结：**

|  |
| --- |
| 方法的核心是——**判断当前数据在历史数据中的位置。** |

我们可以抓住这个核心，从而对其进行变式，以灵活适应实际情形。

10. **延伸：**

除了上述的方法应用之外。

* 1）还可以考虑使用其他分位数，进行更细的划分，将数据划分成更多的区域。例如十二分位数：
* 将历史数据划分成 12 个区域，前 4 个区域代表的是乐观买入，中间 4 个区域代表的是中性观望，后 4 个区域代表的是悲观卖出。
* 那么即可为买入卖出设置不同的买入档位、卖出档位。
* 2）还可以考虑使用该方法对其他指标进行判断，作为买卖动作的参考依据。例如成交量、价格、分时量等。
* 3）还可以考虑使用该方法跟随其他大型资金的动向，判断当前数值在历史数据中的位置，作为是选择买入、观望，还是卖出的参考依据。例如公募基金、产业基金、设备基金等。

|  |
| --- |
| **备注：分位数法**可以反映某单一数据在整体数据中所处的位置，但由于它是通过得到几个代表性数值（分位数），进而得知这几个位置的数据水平。因此它并没有利用全部数据，无法反映整体数据的总体水平。如需反映数据的总体水平，还需结合平均数、方差、极差等数据。 |