

偏度系数与峰度系数的介绍

1 偏度系数

偏度系数 (skewness) 是描述数据分布不对称程度的统计量。它衡量了数据分布的偏斜程度，即数据分布是否对称，以及偏斜的方向。

1.1 定义

偏度系数是第三中心矩与标准差的立方的比值。简单来说，它衡量了数据分布的对称性。

1.2 计算公式

偏度系数 γ_1 可以用下面的公式计算：

$$\gamma_1 = \frac{E[(X - \mu)^3]}{\sigma^3} \quad (1)$$

其中：

- $E[(X - \mu)^3]$ 是第三中心矩（即每个数据点与均值之差的三次方的期望值）。
- μ 是数据的均值。
- σ 是数据的标准差。

1.3 解释

- 偏度系数为 0：数据分布是对称的（正态分布的偏度系数为 0）。
- 偏度系数为正：数据分布的右侧尾巴较长，分布向左倾斜（右偏）。
- 偏度系数为负：数据分布的左侧尾巴较长，分布向右倾斜（左偏）。

1.4 应用

偏度系数可以帮助我们理解数据的分布特征，尤其是在数据分析和建模中，知道数据的偏斜情况可以帮助选择合适的统计方法或转换。

2 峰度系数

峰度系数（kurtosis）是描述数据分布尖峭程度的统计量。它衡量数据分布的“尖峰”或“平坦”程度。峰度系数提供了关于数据集中度的信息，即数据分布在均值附近的集中程度。

2.1 定义

峰度系数表示数据分布的尾部和峰部的厚重程度，相对于正态分布的厚重程度进行比较。它通常通过计算第四中心矩来得到。

2.2 计算公式

峰度系数 γ_2 可以用以下公式计算：

$$\gamma_2 = \frac{E[(X - \mu)^4]}{\sigma^4} - 3 \quad (2)$$

其中：

- $E[(X - \mu)^4]$ 是第四中心矩（即每个数据点与均值之差的四次方的期望值）。
- μ 是数据的均值。

- σ 是数据的标准差。
- -3 是为了使得正态分布的峰度系数为 0（这个调整使得计算结果可以与正态分布的峰度进行直接比较）。

2.3 解释

- 峰度系数为 0：数据分布的尖锐程度与正态分布相同（通常称为“中性峰度”）。
- 峰度系数大于 0：数据分布比正态分布更尖锐，峰部较高且尾部较厚（称为“高峰度”或“超高峰度”）。
- 峰度系数小于 0：数据分布比正态分布更平坦，峰部较低且尾部较薄（称为“低峰度”或“次高峰度”）。

2.4 应用

峰度系数可以帮助分析数据的异常值和风险。高峰度的分布可能表示数据中存在较多的极端值或异常值，而低峰度的分布则表示数据较为平稳，极端值较少。

在金融领域，资产收益率的分布如果具有较高的峰度，意味着极端收益（如大涨或大跌）的概率较高，这对于风险管理非常重要。