



第二节 离散程度的测度

离散程度反映的是数据之间的差异程度。集中趋势的测度值是对数据水平的一个概括性的度量，它对一组数据的代表程度，取决于该组数据的离散水平。数据的离散程度越大，集中趋势的测度值对该组数据的代表性就越差，离散程度越小，其代表性就越好。

测度数据离散程度的有方差、标准差、离散系数。



第二节 离散程度的测度

1、方差

方差是数据组中各数值与其均值离差平方的平均数，它能较好地反映出数据的离散程度，是实际中应用最广泛的离散程度测度值。方差越小，说明数据值与均值的平均距离越小，均值的代表性越好。



第二节 离散程度的测度

对于总体数据，常用的方差计算公式有两种：

$$\sigma^2 = \frac{\sum_{i=1}^N (X_i - \bar{X})^2}{N}$$

$$s^2 = \frac{\sum_{i=1}^N (X_i - \bar{X})^2}{N - 1}$$

前者的分母是总体规模N，后者的分母是总体规模N减去1。



第二节 离散程度的测度

对于样本数据，常用的方差计算公式为：

$$s^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}$$

样本方差 s^2 的分母是样本规模 n 减去1。根据前例中5名营业员元旦当天的销售额样本数据，计算日销售额的样本方差。

$$\begin{aligned} s^2 &= \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1} \\ &= \frac{(520 - 570)^2 + (600 - 570)^2 + (480 - 570)^2 + (750 - 570)^2 + (500 - 570)^2}{5 - 1} \\ &= 12200 \end{aligned}$$



第二节 离散程度的测度

2、标准差

我们经常使用标准差来测度数据的离散程度，**标准差**即方差的平方根。对于样本数据，常用的标准差计算公式为：

$$s = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}}$$



第二节 离散程度的测度

同样地，利用5名营业员元旦当天的销售额样本数据，计算日销售额的标准差。

$$s = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}} = \sqrt{12200} = 110.45(\text{元})$$

标准差与方差只适用于数值型数据。此外，与均值一样，它们对极端值也很敏感。



第二节 离散程度的测度

3、离散系数

离散系数也称变异系数或标准差系数，即标准差与均值的比值，主要用于不同类别数据离散程度的比较，记为CV。离散系数计算公式为：

$$CV = \frac{s}{\bar{X}}$$



第二节 离散程度的测度

根据5名营业员元旦当天的销售额案例数据，计算日销售额的离散系数。

$$CV = \frac{110.45}{570} = 0.19$$

标准差的大小不仅与数据的测度单位有关，也与观测值的均值大小有关，不能直接用标准差比较不同变量的离散程度。离散系数消除了测度单位和观测值水平不同的影响，因而可以直接用来比较变量的离散程度。



第二节 离散程度的测度

【例-单选题】集中趋势的测度值对一组数据的代表程度，取决于该组数据的离散水平。数据的离散程度越大，集中趋势的测度值对该组数据的代表性（ ）。

- A. 越好
- B. 越差
- C. 始终不变
- D. 在一定区间内反复变化



第二节 离散程度的测度

答案：B

解析：离散程度反映的是数据之间的差异程度。集中趋势的测度值是对数据水平的一个概括性的度量，它对一组数据的代表程度，取决于该组数据的离散水平。数据的离散程度越大，集中趋势的测度值对该组数据的代表性就越差，离散程度越小，其代表性就越好。



第二节 离散程度的测度

【例-单选题】离散系数比标准差更适用于比较两组数据的离散程度，这是因为离散系数（ ）。

- A. 不受极端值的影响
- B. 不受数据差异程度的影响
- C. 不受观测值水平或计量单位的影响
- D. 计量更简单

答案：C

解析：离散系数消除了测度单位和观测值水平不同的影响，因而可以直接用来比较变量的离散程度。



第三节 分布形态的测度

1、偏态系数

偏度是指数据分布的偏斜方向和程度，描述的是数据分布对称程度。测度数据分布偏度的统计量称为偏态系数，其计算公式为：

$$SK = \frac{n}{(n-1)(n-2)} \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{s} \right)^3$$



第三节 分布形态的测度

如果偏态系数等于0，说明数据的分布是对称的；

如果偏态系数为正值，说明分布为右偏的，取值在0和0.5之间说明轻度右偏，取值在0.5和1之间说明中度右偏，取值大于1说明严重右偏；

如果偏态系数为负值，说明分布为左偏，取值在0和-0.5之间说明轻度左偏，取值在-0.5和-1之间说明中度左偏，取值小于-1说明严重左偏。

偏态系数的绝对值越大，说明数据分布的偏斜程度越大。



第三节 分布形态的测度

2、标准分数（也称为Z分数）

在统计上，均值和标准差不同时，不同变量的数值是不能比较的。来自不同分布的变量值不可比，但是每个数值在变量分布中相对于均值的相对位置是可比的，因此可以通过计算标准分数来比较不同变量的取值。标准分数可以给出数值距离均值的相对位置，计算公式为：

$$z_i = \frac{X_i - \bar{X}}{S} \quad \text{标准分数} = \frac{\text{数值} - \text{均值}}{\text{标准差}}$$



第三节 分布形态的测度

【例】在考核A中员工得分的均值为80分，标准差为20分，员工得分80分转化后的标准分数为0

$$A\text{标准分数} = \frac{\text{数值} - \text{均值}}{\text{标准差}} = \frac{80 - 80}{20} = 0$$

在考核B中员工得分的均值为60分，标准差为5分。员工得分80分转化后的标准分数为4

$$B\text{标准分数} = \frac{\text{数值} - \text{均值}}{\text{标准差}} = \frac{80 - 60}{5} = 4$$



第三节 分布形态的测度

同样是80的得分，在考核A中属于中等水平，在考核B中可能名列前茅。



第三节 分布形态的测度

标准分数是统计上常用的一种标准化方法，转变后的标准分数并没有改变数值在原分布中的位置，也没有改变数据原分布的偏度，但是标准分数的平均数为0，标准差为1。



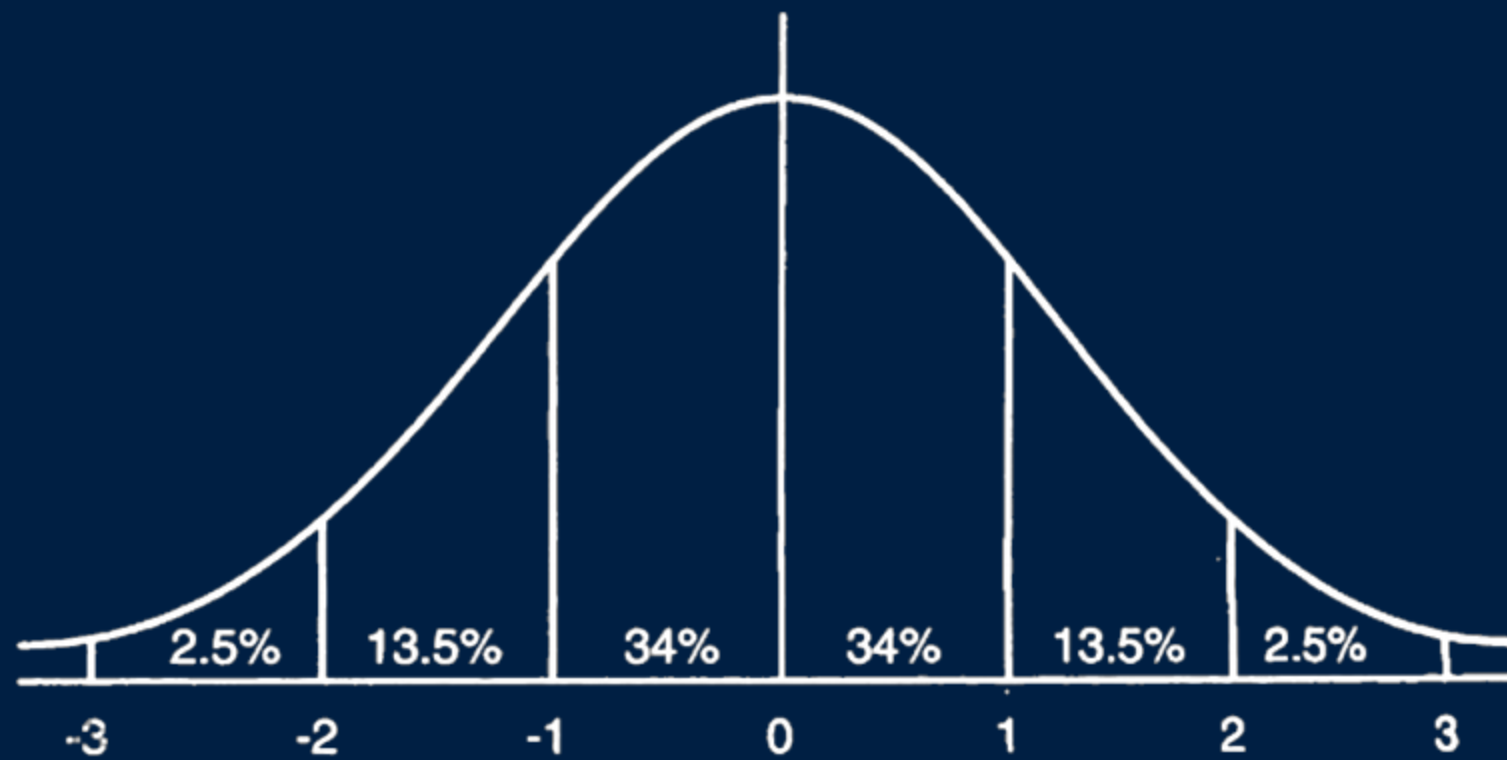
第三节 分布形态的测度

在实际应用中，当数据服从对称的钟形分布时，可以运用经验法则来判断与均值的距离在特定倍数标准差之内的数据项所占比例。经验法则表明：

- 1、约有68%的数据与平均数的距离在1个标准差之内，标准分数在【-1，+1】范围内；
- 2、约有95%的数据与平均数的距离在2个标准差之内；标准分数在【-2，+2】范围内；
- 3、约有99%的数据与平均数的距离在3个标准差之内；标准分数在【-3，+3】范围内；



第三节 分布形态的测度





第三节 分布形态的测度

【例-单选题】根据经验法则，服从对称钟形分布的标准分数在【-1， 1】范围内的概率是（ ）。

- A. 95%
- B. 50%
- C. 68%
- D. 99%

答案：C

解析：对于服从对称的钟形分布的标准分数，68%的标准分数在【-1， 1】范围内。