二、离散程度的测度

二、离散程度的测度

考点1 方差★★★★

考点2 标准差★★★★

考点3 离散系数★★★★

二、离散程度的测度

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。

集中趋势与离散程度的关系

（1）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

（2）数据的离散程度越小，集中趋势的测度值对该组数据的代表性就越好

二、离散程度的测度

考点1 方差★★★★

含义

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。

方差与均值的关系

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。

公式

对于样本数据，常用的方差公式为：$S^2=\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n-1}$

二、离散程度的测度

【例题】

某售货小组有5名营业员，元旦一天的销售额分别为520元、600元、480元、750元和500元，求该日销售额的样本方差

二、离散程度的测度

【计算过程】

（1）计算均值（平均数）（520+600+480+750+500）÷5=570元

（2）计算各数值与均值的差

520-570=-50；600-570=30；480-570=90；750-570=180；500-570=-70

（3）计算差平方和

$(-50)^2+(30)^2+(90)^2+(180)^2+(-70)^2$

二、离散程度的测度

（4）计算方差

$$\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n-1}=\frac{(-50)^2+(30)^2+(90)^2+(180)^2+(-70)^2}{5-1}=12200$$

二、离散程度的测度

考点2 标准差★★★★

含义

即方差的的平方根。

公式

对于样本数据，常用的方差公式为：$S=\sqrt{\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n-1}}$

示例

上例题算出方差后，开平方根即可

样本标准差=√12200=110.45

二、离散程度的测度

地位

标准差与方差是应用最广泛的统计离散程度的测度方法。

适用范围

但是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。

特点

标准差不仅能度量数值与均值的平均距离，还与原始数值具有相同的计量单位。

二、离散程度的测度

考点3 离散系数★★★★

含义

也称为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。

公式 $CV = \frac{S}{\bar{X}}$

示例 $CV = \frac{110.45}{570} = 0.19$

仍利用本节前面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

二、离散程度的测度

特点

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，因此可以直接用来比较变量的离散程度。

【习题演练】

【单选题】标准差系数是一组数据的标准差与其相应的（　　）之比。

A. 均值

B. 极值

C. 众数

D. 几何平均数

【习题演练】

答案：A

解析：离散系数也称为变异系数或标准差系数，即标准差与均值的比值。均值为算数平均数，不是几何平均数。

【习题演练】

【单选题】某学校学生的平均年龄为20岁，标准差为3岁；该校教师的平均年龄为38岁，标准差为3岁。比较该校学生年龄和教师年龄的离散程度，则（　）。

A. 学生年龄和教师年龄的离散程度相同

B. 教师年龄的离散程度大一些

C. 教师年龄的离散程度是学生年龄离散程度的1.9倍

D. 学生年龄的离散程度大一些

【习题演练】

答案：D

解析：平均值不同的情况下，用离散系数比较离散程度。

学生年龄的离散系数=3/20=0.15

教师年龄的离散系数=3/38=0.0789

离散系数大的说明数据的离散程度也就大，离散系数小的说明数据的离散程度也就小。

【习题演练】

【单选题】下列离散程度的测度值中，能消除测度单位和观测值水平不同的影响是（　　）。

A. 标准差

B. 离散系数

C. 方差

D. 均值

答案：B

二、离散程度的测度

总结

<table>
<tr><th></th><th>指标
（测度值）</th><th>是否受极
端值影响</th><th>数值型数据
（定量数据）</th><th>分类
数据</th><th>顺序
数据</th></tr>
<tr><td rowspan="3">测度数
据的集
中趋势</td><td>均值</td><td>受影响</td><td>适用</td><td>不适用</td><td>不适用</td></tr>
<tr><td>中位数</td><td>不受影响</td><td>适用</td><td>不适用</td><td>适用</td></tr>
<tr><td>众数</td><td>不受影响</td><td>不适用</td><td>适用</td><td>适用</td></tr>
<tr><td rowspan="3">测度数
据的离
散趋势</td><td>方差</td><td rowspan="3">受影响</td><td rowspan="3">适用</td><td rowspan="3">不适用</td><td rowspan="3">不适用</td></tr>
<tr><td>标准差</td></tr>
<tr><td>离散系数</td></tr>
</table>

【习题演练】

【单选题】下列指标中，用于描述数据集中趋势并且易受极端值影响的是（　　）。

A. 平均数

B. 中位数

C. 方差

D. 标准差

【习题演练】

答案：A

解析：均值、方差、标准差都容易受极端值影响，但是用于描述数据集中趋势的指标是均值（即平均数），方差和标准差是描述数据离散趋势的指标。

【习题演练】

【多选题】数值型数据离散程度的测度指标有（　　）。

A. 中位数

B. 均值

C. 离散系数

D. 标准差

E. 方差

答案：CDE

【习题演练】

【单选题】下列统计量中，适用于描述分类数据集中趋势的是（　　）。

A.均值

B.中位数

C.变异系数

D.众数

答案：D

【习题演练】

【多选题】下列统计量中，容易受极端值影响的有（　　）。

A. 均值

B. 方差

C. 标准差

D. 众数

E. 中位数

答案：ABC

三、分布形态的测度

三、分布形态的测度

考点1 偏态系数★★★★

考点2 标准分数★★★★

三、分布形态的测度

考点1 偏态系数★★★★

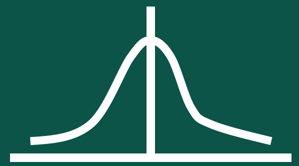
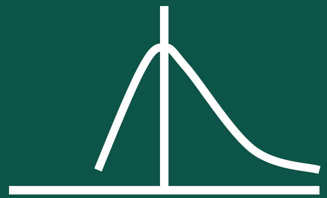
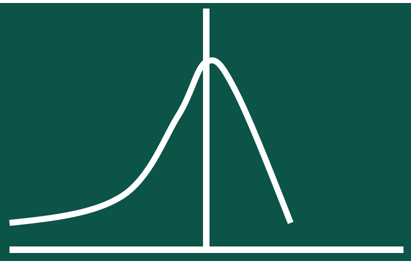
定义

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。

公式

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

三、分布形态的测度

<table>
<tr><th colspan="2">分布形态</th><th>说明</th><th>图形</th></tr>
<tr><td colspan="2">对称分布
（正态分布）</td><td>偏态系数等于0</td><td></td></tr>
<tr><td rowspan="2">偏态分布</td><td>右偏分布</td><td>偏态系数为正值
0-0.5轻度右偏
0.5-1中度右偏
大于1严重右偏</td><td></td></tr>
<tr><td>左偏分布</td><td>偏态系数为负值
0- -0.5轻度左偏
-0.5- -1中度左偏
小于-1严重左偏</td><td></td></tr>
<tr><td colspan="4">【补充】偏态系数的绝对值越大，数据分布偏斜程度越大。</td></tr>
</table>

【习题演练】

【单选题】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。当偏态系数=0.96时，说明（　）。

A.数据的分布是对称的

B.数据分布中度右偏

C.数据分布严重右偏

D.数据分布中度左偏

答案：B

【习题演练】

【单选题】测度数据分布偏度的统计量称为偏态系数，下列偏态系数说明数据分布的偏斜程度最大的是（　）。

A. 偏态系数为0

B. 偏态系数为0.5

C. 偏态系数为1

D. 偏态系数为-2

【习题演练】

答案：D

解析：偏态系数的绝对值越大，说明数据分布的偏斜程度越大。

【习题演练】

【多选题】某企业员工年收入数据分布的偏态系数3.0，则该组数据分布形态为（   ）。

A.右偏

B.左偏

C.轻度偏斜

D.严重偏斜

E.中度偏斜

答案：AD

三、分布形态的测度

考点2 标准分数★★★★

标准分数含义	标准分数也称为Z分数，是统计上常用的一种标准化方法。标准分数用于某一个数值在一组数据中相对位置的度量。计算方法是用数值减去均值所得的差除以标准差。
标准分数计算公式	$Z_i = \frac{X_i - \bar{X}}{s}$

三、分布形态的测度

【例1】甲、乙、丙、丁四人在某次语文考试中分别获得72分、60分、48分和90分，而全体学生的语文平均成绩为60分，标准差为12分，则这4个相应的标准分数分别为：

$$Z_{甲}=\frac{72-60}{12}=1 \qquad Z_{乙}=\frac{60-60}{12}=0$$

$$Z_{丙}=\frac{48-60}{12}=-1 \qquad Z_{丁}=\frac{90-60}{12}=2.5$$

三、分布形态的测度

标准分数经验法则

经验法则表明

对于服从对称的钟形分布的标准分数：

- 约有68%的标准分数在[-1，+1]范围内
- 约有95%的标准分数在[-2，+2]范围内
- 约有99%的标准分数在[-3，+3]范围内

【习题演练】

【单选题】根据经验法则，服从对称钟形分布的标准分数在[-2，+2]范围内的概率是（　　）。

A. 50%

B. 95%

C. 68%

D. 99%

【习题演练】

答案：B

解析：对于服从对称的钟形分布的标准分数：

约有68%的标准分数在[-1，+1]范围内

约有95%的标准分数在[-2，+2]范围内

约有99%的标准分数在[-3，+3]范围内