第十一章 数据描述

一、平均值

设 n 个数 $x_1,x_2,\cdots,x_n$，称 $\bar{x}=\dfrac{x_1+x_2+\cdots+x_n}{n}$ 为这 n 个数的平均值.

二、方差

设一组样本数据 $x_1,x_2,\cdots,x_n$，其平均数为 $\bar{x}$，则称

$$s^2=\frac{1}{n}[(x_1-\bar{x})^2+(x_2-\bar{x})^2+\cdots+(x_n-\bar{x})^2]=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

为这个样本的方差.

三、标准差

因为方差与原始数据的单位不同，且平方后可能夸大了离差的程度，将方差的算术平方根称为这组数据的标准差，即 $s=\sqrt{\dfrac{1}{n}\sum\limits_{i=1}^{n}(x_i-\bar{x})^2}$.

四、方差和标准差的意义

方差的实质是各数据与平均数的差的平方的平均数. 标准差是方差的一个派生概念，它的优点是单位和样本的数据单位保持一致，给计算和研究带来方便.

方差和标准差用来比较平均数相同的两组数据波动的大小，也用它描述数据的离散程度. 方差或标准差越大，说明数据的波动越大，越不稳定；方差或标准差越小，数据波动越小、越整齐、越稳定.

利用方差比较数据波动大小的方法和步骤：先求平均数，再求方差，然后判断得出结论.

五、其他概念

1. 众数

出现次数最多的数称为众数.

2. 中位数

将数据由小到大排列，若有奇数个数据，则正中间的数为中位数；若有偶数个数据，则中间两个数的平均数为中位数.

六、直方图

1. 定义

把数据分为若干个小组，每组的组距保持一致，并在直角坐标系的横轴上标出每组的位

置(以组距作为底),计算每组所包含的数据个数(频数),以该组的"频率/组距"为高作矩形,这样得出若干个矩形构成的图叫做直方图.

2. 定义所包含的要点

(1) 组距的确定:一般是人为确定,不能太大也不能太小.

(2) 组数的确定:组数=极差/组距.

(3) 每组频率的确定:频率=频数/数据容量.

(4) 每组所确定的矩形的面积=组距$\times\dfrac{\text{频率}}{\text{组距}}$=频率.

(5) 频率直方图下的总面积等于1(各个矩形面积之和等于1)

(6) 分组时要遵循"不重不漏"的原则:"不重"是指某一个数据只能分在其中的某一组,不能在其他组中出现;"不漏"是指组别能够穷尽,即在所分的全部组别中每项数据都能分在其中的某一组,不能遗漏.

【注意】为了解决上述问题,分组时采用左闭右开的区间表示:[ ).

例如某数据分组时,其中的两组分别为[227,290)和[290,353),这样290这个数据就只存在于第二个区间中了,避免了290同时属于两个区间的情况发生.

3. 在直方图中,众数是最高矩形底边中点的横坐标;中位数左边和右边的直方图的面积相等;平均数是直方图的重心,它等于每个小矩形的面积乘以小矩形底边中点横坐标之和.

七、饼图

饼图是一个划分为几个扇形的圆形统计图表,用于描述量、频率或百分比之间的相对关系.在饼图中,每个扇区的弧长(以及圆心角和面积)大小为其所表示的数量的比例.这些扇区合在一起刚好是一个完全的圆形.顾名思义,这些扇区拼成了一个切开的饼形图案.其所用公式为:某部分所占的百分比等于对应扇形所占整个圆周的比例.

第十一章 基础例题

【题型一】平均数的计算

【例1】假设三个相异正整数中的最大数是54,则三个数的最小平均值是( ).

(A) 17 (B) 19 (C) 21 (D) 23 (E) 18

【解析】最大是54,取最小的正数,是1,2.所以平均数是19=(1+2+54)/3,选B.

【例2】在一次法律知识竞赛中,甲机关20人参加,平均80分,乙机关30人参加,平均70分,则两个机关参加竞赛的人总平均分是( ).

(A) 76 (B) 75 (C) 74 (D) 73 (E) 77

【解析】$(20\times80+30\times70)\div(20+30)=74$,选C.

【题型二】方差于标准差的计算

【例3】给出两组数据:甲组:20,21,23,24,26;乙组:100,101,103,104,106.甲组,乙组的方差分别为s_1^2,s_2^2.则下列正确的是( ).

(A) $s_1^2 > s_2^2$　(B) $s_1^2 < s_2^2$　(C) $s_1^2 = s_2^2$　(D) $s_1^2 \neq s_2^2$　(E)无法确定

【解析】甲组：$\bar{x}_1 = 20 + (0+1+3+4+6)/5 = 22.8$，

乙组：$\bar{x}_2 = 100 + (0+1+3+4+6)/5 = 102.8$，

得：$s_1^2 = \frac{1}{5}[(2.8)^2 + (1.8)^2 + (0.2)^2 + (1.2)^2 + (3.2)^2]$

$s_2^2 = \frac{1}{5}[(2.8)^2 + (1.8)^2 + (0.2)^2 + (1.2)^2 + (3.2)^2]$，可得 $s_1^2 = s_2^2$，故选 C.

【评注】本题说明将每个数都加上相同一个数，方差不变.

【题型3】饼图　开心提示此题型为重点题型

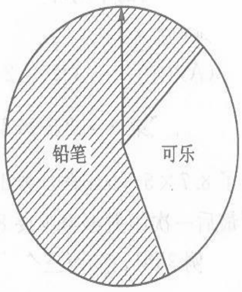
【例 4】某商场设立了一个可以自由转动的转盘(如图)，并规定：顾客购物 10 元以上能获得一次转动转盘的机会，当转盘停止时，指针落在哪一区域就可以获得相应的奖品，下表是活动进行中的一组统计数据：

转动转盘的次数 n	100	150	200	500	800	1000
落在“铅笔”的次数 m	68	111	136	345	546	701
落在“铅笔”的频率 $\frac{m}{n}$	0.68	0.74	0.68	0.69	0.6825	0.701

(1)请估计，当 n 很大时，频率将会接近多少？

(2)转动该转盘一次，获得铅笔的概率约是多少？

(3)在该转盘中，标有“铅笔”区域的扇形的圆心角大约是多少？(精确到 1°)



【解析】(1)从表中可以看出，当 n 很大时，频率会接近 0.69.

(2)用频率近似表示概率，故概率约为 0.69.

(3)所求扇形的圆心角为 $0.69 \times 360° \approx 248°$.

【评注】(1)试验的次数越多，所得的频率越能反映概率的大小；(2)频数分布表、扇形图、条形图、直方图都能较好地反映频数、频率的分布情况，可以利用所提供的信息估计概率.

【题型4】数表

【例 5】某农贸市场出售西红柿，当价格上涨时，供给量相应增加，而需求量相应减少，具体调查结果如下表：

表 1　市场供给量

单价/元·kg^{-1}	2	2.4	2.8	3.2	3.6	4
供给量/1000kg	50	60	70	75	80	90

表 2　市场需求量

单价/元·kg^{-1}	4	3.4	2.9	2.6	2.3	2
需求量/1000kg	50	60	65	70	75	80

根据以上提供的信息，市场供需平衡点(即供给量和需求量相等时的单价)应在区间(　　).

(A) (2.3,2.6)　　(B) (2.4,2.6)　　(C) (2.6,2.8)

(D) (2.8,2.9)　　(E) (2.7,2.9)

【解析】从表中可以看出，当价格从 2 涨到 4 时，市场供给量从 50 增加到 90；而当价格从 4 降到 2 时，市场需求量从 50 增加到 80. 因此市场供需平衡点(即供给量和需求量相等时的单价)应在区间(2.6,2.8)，选 C.