数学系统精讲

数据描述

MBA大师 — 数学董璞

12 数据描述

平均值与方差：
- 【1】基本概念，计算公式
- 【2】部分的平均值与总量的平均值
- 【3】不同数据方差/平均值比较大小

附加需求能力：能从图表中理解数据（条件有时候会以表格的方式给出）

12.1 基本概念、公式

【1】平均值

算术平均值：设x_1，x_2，……，x_n为n个数，称$\frac{x_1+x_2+\cdots+x_n}{n}$为这$n$个数的算术平均值，记为：$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$

【2】方差

在一组数据x_1，x_2，……，x_n中，各数据与它们的平均数$\bar{x}$的差的平方的平均值称为这组数据的方差，通常用s^2表示

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

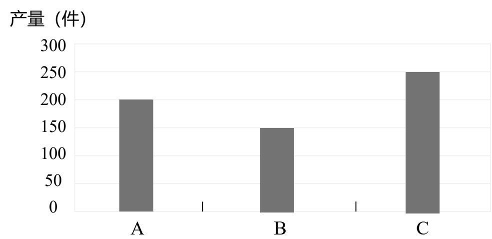
方差的算术平方根称为这组数据的标准差

方差用来反映数据波动的大小，方差大波动大，方差小波动小

12.1 基本概念、公式

【3】直方图

把数据分成若干个小组，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，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形叫做直方图。



12.1 基本概念、公式

【2006.1.4】如果x_1, x_2, x_3三个数的算术平均值为5，则x_1+2, x_2-3, x_3+6与8的算术平均值为（ C ）

(A) $3\frac{1}{4}$ (B) 6 (C) 7 (D) $9\frac{1}{5}$ (E) $7\frac{1}{2}$

给出：x_1, x_2, x_3三个数的算术平均值为5

$\frac{x_1+x_2+x_3}{3}=5$，$x_1+x_2+x_3=15$

要求：x_1+2, x_2-3, x_3+6与8的算术平均值

$$\frac{(x_1+2)+(x_2-3)+(x_3+6)+8}{4}=\frac{x_1+x_2+x_3+2-3+6+8}{4}=7$$

12.1 基本概念、公式 平移数据、此消彼长

【2018.2】为了解某公司员工的年龄结构，按男、女人数的比例进行了随机抽样，结果如下：根据表中数据估计，该公司男员工的平均年龄与全体员工的平均年龄分别是（单位：岁）（ A ）

A.32，30 B.32，29.5 C.32，27 D.30，27 E.29.5，27

方法2：

	24								40
男员工年龄（岁）	$\boxed{23}$	26	28	30	32	34	36	38	$\boxed{41}$
女员工年龄（岁）	23	25	27	27	29	31			

男工：等差数列：$9\times32\div9=32$ 女工：等差数列：$6\times27=162$ 全体员工$=\frac{288+162}{15}=30$

方法1：

$$男员工年龄=\frac{23+26+28+30+32+34+36+38+41}{9}=\frac{288}{9}=32$$

$$全体员工年龄=\frac{288+23+25+27+27+29+31}{15}=30$$

12.1 基本概念、公式

【2016.21】设两组数据S_1:3，4，5，6，7和S_2:4，5，6，7，a，则能确定a的值.（ A ）

（1）S_1与S_2的均值相等. （2）S_1与S_2的方差相等.

条件（1）$\frac{3+4+5+6+7}{5}=\frac{4+5+6+7+a}{5}$，$a=3$，条件（1）充分

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

条件（2）${S_1}^2=\frac{1}{5}[(3-5)^2+(4-5)^2+(5-5)^2+(6-5)^2+(7-5)^2]=2$

$$S_2^2=\frac{1}{5}(4^2+5^2+6^2+7^2+a^2)-\left[\frac{1}{5}(4+5+6+7+a)\right]^2$$

$a=3$或8，不唯一确定，条件（2）不充分

12 数据描述

【1】基本概念，计算公式

【2】部分的平均数与总量的平均值

如果总体数量分为甲、乙两部分，甲的平均价格为a，乙的平均价格为b，总平均价格为c

考点1）总体的平均数，一定在两个部分平均数范围之间。

具体值在中间的什么位置，取决于甲乙数量的比例。

甲平均价格 < 总平均价格 < 乙平均价格（$a<b$时）

乙平均价格 < 总平均价格 < 甲平均价格（$a>b$时）

考点2）总额 $=c\times$总量$=c\times$(甲数量+乙数量)$=a\times$甲数量$+b\times$乙数量

$(a-c)\times$甲数量$=(c-b)\times$乙数量

【3】不同数据方差/平均值比较大小

12.2 部分均值与总体均值

【2003.1.2】车间共有40人，某技术操作考核的平均成绩为80分，其中男工平均成绩为83分，女工平均成绩为78分，该车间有女工（ D ）

（A）16人 （B）18人 （C）20人 （D）24人 （E）28人

(总平均分 − 女生平均分) × 女工数量 = (男工平均分 − 总平均分) × 男工数量

$$\frac{总平均分-女生平均分}{男生平均分-总平均分}=\frac{男工数量}{女工数量}=\frac{80-78}{83-80}=\frac{2}{3}$$

共有5份，男工2份，女工3份

共40人，即8人/份，女工3份，24人

12.2 部分均值与总体均值

【2011.10.19】甲、乙两组射手打靶，两组射手的平均成绩是150环.（ C ）

（1）甲组的人数比乙组人数多20%.

（2）乙组的平均成绩是171.6环，比甲组的平均成绩高30%.

(乙组平均成绩 − 总平均成绩) × 乙组人数 = (总平均成绩 − 甲组平均成绩) × 甲组人数

乙组的平均成绩是171.6环，甲组平均成绩是132环

设总平均成绩为x

$$\frac{171.6-x}{x-132}=\frac{甲组人数}{乙组人数}=1.2，x=150$$

12.2 部分均值与总体均值

【2013.10.2】某学校高一年级男生人数占该年级学生人数的40%。在一次考试中，男、女生的平均分数分别为75和80，则这次考试高一年级学生的平均分数为（ D ）

（A）76 （B）77 （C）77.5 （D）78 （E）79

(女生平均分 − 总平均分) × 女生数量 = (总平均分 − 男生平均分) × 男生数量

$$\frac{男生数量}{女生数量}=\frac{女生平均分-总平均分}{总平均分-男生平均分}=\frac{40\%}{60\%}=\frac{80-x}{x-75}，x=78$$

【权重】$75\times0.4+80\times0.6=78$

【特值法】设男生有4人，女生有6人，则有：$\frac{4\times75+6\times80}{10}=78$

12.2 部分均值与总体均值

【2014.1.1】某部门在一次联欢活动中共设了26个奖，奖品均价为280元，其中一等奖单价为400元，其他奖品均价为270元，一等奖的个数为（ E ）

（A）6 （B）5 （C）4 （D）3 （E）2

解法一：设一等奖个数为x，二等奖个数为y

$$\begin{cases}x+y=26\\400x+270y=26\times280\end{cases}\Rightarrow\begin{cases}x=2\\y=24\end{cases}$$

解法二：(总平均价 − 其他奖平均价) × 其他奖数量 = (一等奖平均价 − 总平均价) × 一等奖数量

总奖品均价280元，一等奖单价400元，其他奖品均价270元

$$\frac{400-280}{280-270}=\frac{12}{1}=\frac{其它奖数量}{一等奖数量}$$

故共13份，26个奖，每份2个，一等奖占1份，有2个

12.2 部分均值与总体均值

【2015.5】在某次考试中，甲、乙、丙三个班的平均成绩分别为80、81和81.5，三个班的学生分数之和为6952，三个班共有学生（ B ）

(A) 85名　(B) 86名　(C) 87名　(D) 88名　(E) 90名

三个班平均分一定在80~81.5之间，设总人数为a

$$80a < 6952 < 81.5a \Rightarrow \begin{cases} a < 87 \\ a > 85 \end{cases}$$

12 数据描述

【1】基本概念，计算公式

【2】部分的平均数与总量的平均值

【3】不同数据方差/平均值比较大小

方法1：直接套公式计算

方法2：看比例/看权重/做对比（数字敏感）

12.3 方差/均值比较大小

【2017.14】甲、乙、丙三人每轮各投篮10次，投了三轮，投中数如下表：

记σ_1，σ_2，σ_3分别为甲、乙、丙投中数的方差，则（ B ）

A. $\sigma_1 > \sigma_2 > \sigma_3$　B. $\sigma_1 > \sigma_3 > \sigma_2$　C. $\sigma_2 > \sigma_1 > \sigma_3$　D. $\sigma_2 > \sigma_3 > \sigma_1$　E. $\sigma_3 > \sigma_2 > \sigma_1$

	第一轮	第二轮	第三轮
甲	2	5	8
乙	5	2	5
丙	8	4	9

$\bar{x}_1 = \frac{2+5+8}{3} = 5$，$\bar{x}_2 = \frac{5+2+5}{3} = 4$，$\bar{x}_3 = \frac{8+4+9}{3} = 7$

$$\sigma_1 = \frac{1}{3}[(2-5)^2 + (5-5)^2 + (8-5)^2] = 6$$

$$\sigma_2 = \frac{1}{3}[(5-4)^2 + (2-4)^2 + (5-4)^2] = 2$$

$$\sigma_3 = \frac{1}{3}[(8-7)^2 + (4-7)^2 + (9-7)^2] = \frac{14}{3} \qquad \sigma_1 > \sigma_3 > \sigma_2$$

12.3 方差/均值比较大小

【2017.14】甲、乙、丙三人每轮各投篮10次，投了三轮，投中数如下表：

记σ_1，σ_2，σ_3分别为甲、乙、丙投中数的方差，则（ B ）

A. $\sigma_1 > \sigma_2 > \sigma_3$　B. $\sigma_1 > \sigma_3 > \sigma_2$　C. $\sigma_2 > \sigma_1 > \sigma_3$　D. $\sigma_2 > \sigma_3 > \sigma_1$　E. $\sigma_3 > \sigma_2 > \sigma_1$

	第一轮	第二轮	第三轮
甲	2	5	8
乙	5	2	5
丙	8	4	9

甲极差：$8-2=6$

已极差：$5-2=3$

丙极差：$9-4=5$

甲极差>丙极差>乙极差

【方差】表明数据偏离平均数的程度/波动大小。

【极差】最大值与最小值之间的差距。

标志分布范围和离散度。

仅取决于两个极端值的水平，不反应其间的变量分布情况，同时易受极端值的影响。

注意：极差和方差是不同的两个概念

极差大不等同于方差大。

如以下数据：S_1：4，7，7，7，7，7，10

S_2：6，6，6，8，10，10，10

S_1极差6，方差18/7；S_2极差4，方差24/7

12.3 方差/均值比较大小

【2012.1.6】甲、乙、丙三个地区的公务员参加一次测评，其人数和考分情况如下表，三个地区按平均分由高到低的排名顺序为（ E ）

A. 乙、丙、甲　B. 乙、甲、丙　C. 甲、丙、乙　D. 丙、甲、乙　E. 丙、乙、甲

地区 \ 人数 \ 分数	6	7	8	9
甲	10	10	10	10
乙	15	15	10	20
丙	10	10	15	15

甲地区：$\frac{6\times10+7\times10+8\times10+9\times10}{40}=7.5$　丙地区：$\frac{6\times10+7\times10+8\times15+9\times15}{50}=7.7$

乙地区：$\frac{6\times15+7\times15+8\times10+9\times20}{60}=7.6$

12.3 方差/均值比较大小

【2012.1.6】甲、乙、丙三个地区的公务员参加一次测评，其人数和考分情况如下表，三个地区按平均分由高到低的排名顺序为（ E ）【技巧】看比例看权重

A. 乙、丙、甲　B. 乙、甲、丙　C. 甲、丙、乙　D. 丙、甲、乙　E. 丙、乙、甲

地区 \ 人数 \ 分数	6	7	8	9
甲	10	10	10	10
乙	15	15	10 10+5	20 20-5
丙	10	10	15 10+5	15 10+5
	15	15	22.5	22.5

甲地区平均分：$\frac{6\times10+7\times10+8\times10+9\times10}{40}=7.5$

分子分母同时同比例改变，均值不变

甲丙相比：如果丙全是10，那么平均分相等，多了5个8分和5个9分，拉高平均分，故丙>甲

甲乙相比：如果把5个9分降低为5个8分，那么与乙平均分相同，降低后相同，故乙>甲

乙丙相比：将丙按比例扩大为15　15　22.5　22.5（平均分不变），此时与乙平均分相同，多了12.5个8分，2.5个9分，故丙>乙

THANK YOU FOR WATCHING