

高考全国卷中统计小命题的命题动向分析

■湖南省郴州市教育科学研究院 汪昌华

高考对统计的考查主要围绕抽样、统计图表、样本估计总体和统计推断四个方面。本文聚焦近几年高考的热点问题,给出一些复习备考的建议和对高考命题的一些看法,以期帮助大家高效地备战高考,实现科学备考的目标。

考向一、分层抽样与样本均值、方差、百分位数等融合

例 1 某高中有学生 500 人,其中男生 300 人,女生 200 人,希望获得全体学生的身高信息,按照分层抽样的方法抽取了容量为 50 的样本。经计算得到男生身高样本均值为 170 cm,方差为 17 cm^2 ;女生身高样本均值为 160 cm,方差为 30 cm^2 。下列结论中正确的是()。

- A. 男生样本容量为 20
- B. 每个女生被抽到的概率均为 $\frac{2}{5}$
- C. 所有样本的均值为 166 cm
- D. 所有样本的方差为 22.2 cm^2

解析:男生样本容量为 $300 \times \frac{50}{500} = 30$,故选项 A 错误;每个学生入样的概率均为 $\frac{50}{500} = \frac{1}{10}$,故选项 B 错误;记男生样本为 y_1, y_2, \dots, y_{30} ,均值为 \bar{y} ,方差为 s_y^2 ,女生样本为 z_1, z_2, \dots, z_{20} ,均值为 \bar{z} ,方差为 s_z^2 ,所有样本均值为 \bar{x} ,方差为 s^2 ,则 $\bar{x} = \frac{\sum_{i=1}^{30} y_i + \sum_{j=1}^{20} z_j}{50} = \frac{30\bar{y} + 20\bar{z}}{50} = \frac{3}{5}\bar{y} + \frac{2}{5}\bar{z} = 166(\text{cm})$, $s^2 = \frac{3}{5}[s_y^2 + (\bar{y} - \bar{x})^2] + \frac{2}{5}[s_z^2 + (\bar{z} - \bar{x})^2] = 46.2(\text{cm}^2)$,故选项 C 正确,选项 D 错误。

故选 C。

点评:本题考查了概率、方差、均值等知识,是一道综合性较强的题目。其中分层抽样的均值和方差是新教材中增加的知识,教材中

推导了分层抽样的方差公式。在必修二 216 页拓广探索第 11 题:已知总体划分为三层,采用样本量比例分配的分层随机抽样,得到各层抽取的样本量、样本平均数和样本方差分别为 $l, \bar{x}, s_1^2; m, \bar{y}, s_2^2; n, \bar{z}, s_3^2$ 。记总的样本平均数为 $\bar{\omega}$,样本方差为 s^2 ,则有如下两个结论:

$$(1) \bar{\omega} = \frac{l}{l+m+n}\bar{x} + \frac{m}{l+m+n}\bar{y} + \frac{n}{l+m+n}\bar{z};$$

$$(2) s^2 = \frac{1}{l+m+n} \{l[s_1^2 + (\bar{x} - \bar{\omega})^2] + m[s_2^2 + (\bar{y} - \bar{\omega})^2] + n[s_3^2 + (\bar{z} - \bar{\omega})^2]\}.$$

例 2 (多选题)已知一组样本数据 x_1, x_2, \dots, x_{15} ,其中 $x_i = 2i (i = 1, 2, \dots, 15)$,由这组数据得到另一组新的样本数据 y_1, y_2, \dots, y_{15} ,其中 $y_i = x_i - 20$,则()。

- A. 两组样本数据的样本方差相同
- B. 两组样本数据的样本平均数相同
- C. y_1, y_2, \dots, y_{15} 样本数据的第 30 百分位数为 -10
- D. 将两组数据合成一个样本容量为 30 的新的样本数据,该样本数据的平均数为 5

解析:对于 A,样本数据 $y_i = x_i - 20$,则 $D(y) = D(x)$,即两组样本数据的样本方差相同,A 正确;

对于 B,样本数据 x_1, x_2, \dots, x_{15} ,其平均数 $\bar{x} = \frac{2+4+\dots+30}{15} = 16$,新的样本数据 $y_i = x_i - 20$,则新的样本数据的平均数 $\bar{y} = \bar{x} - 20 = -4$,B 错误;

对于 C, $30 \times 15\% = 4.5$,则样本数据 y_1, y_2, \dots, y_{15} 的第 30 百分位数为第 5 个数据,即 -10,C 正确;

对于 D,将两组数据合成一个样本容量为 30 的新的样本数据,则该样本数据的平均数为 $\frac{15 \times 16 + 15 \times (-4)}{30} = 6$,D 错误。

故选 AC。

点评:分析两组数据,进而由平均数、方差和百分位数的计算公式分析选项,综合可得答案。有关两组数据之间的均值和方差的结论在教材课后习题中亦有证明。由此可见,统计中的一些结论的证明与应用也是高考考查中的重点。比如, $s^2 = \frac{1}{n} [(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2] = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2$ 。一般地,若两个变量 x_i, y_i 满足 $y_i = ax_i + b$, 其中 a, b 为常数, 则有以下两个结论: (1) $\bar{y} = a\bar{x} + b$; (2) $s_y^2 = a^2 s_x^2, s_y = |a| s_x$ 。

考向二、统计中的数据分析

例 3 四名同学各掷骰子 5 次, 分别记录每次骰子出现的点数, 根据四名同学的统计结果, 可以判断出一定没有出现点数 6 的是()。

- A. 平均数为 3, 中位数为 2
- B. 平均数为 3, 方差为 2.6
- C. 中位数为 3, 众数为 2
- D. 中位数为 3, 方差为 1.6

解析:对于 A, 当投掷骰子出现的结果为 1, 1, 2, 5, 6 时, 满足平均数为 3, 中位数为 2, 可以出现点数 6, 所以 A 错误;

对于 B, 若出现点数 6, 且其平均数为 3, 设这五个数从小到大依次为 $a_1, a_2, a_3, a_4, 6$, 则 $\frac{1}{5} [(a_1 - 3)^2 + (a_2 - 3)^2 + (a_3 - 3)^2 + (a_4 - 3)^2 + (6 - 3)^2] = 2.6$, 即 $(a_1 - 3)^2 + (a_2 - 3)^2 + (a_3 - 3)^2 + (a_4 - 3)^2 = 4$, 又 $\frac{1}{5} (a_1 + a_2 + a_3 + a_4 + 6) = 3$, 所以 $a_1 + a_2 + a_3 + a_4 = 9$, 则 $4a_1 \leq 9, a_1 \leq \frac{9}{4}, 4a_4 \geq 9, a_4 \geq \frac{9}{4}$, 若 $a_1 = 1$, 则 $a_2 + a_3 + a_4 = 8$, 此时 $(a_2 - 3)^2 + (a_3 - 3)^2 + (a_4 - 3)^2 = 0$, 得 $a_2 = a_3 = a_4 = 3$, 与 $a_2 + a_3 + a_4 = 8$ 矛盾, 若 $a_1 = 2$, 此时 $a_2 + a_3 + a_4 = 7$, 由 $(a_2 - 3)^2 + (a_3 - 3)^2 + (a_4 - 3)^2 = 3$, 得 $a_2 = 2, a_3 = 2, a_4 = 4$, 或 $a_2 = 2, a_3 = 4, a_4 = 4$, 或 $a_2 = 4, a_3 = 4, a_4 = 4$, 都与 $a_2 + a_3 + a_4 = 7$ 矛盾, 故这样的五个数不存在, 即不可能出现点数 6, 所以 B 正确;

对于 C, 当投掷骰子出现的结果为 2, 2,

3, 4, 6 时, 满足中位数为 3, 众数为 2, 可以出现点数 6, 所以 C 错误;

对于 D, 当投掷骰子出现的结果为 3, 3, 3, 5, 6 时, 满足中位数为 3, 其平均数为 $\bar{x} = \frac{1}{5} (3 + 3 + 3 + 5 + 6) = 4$, 方差为 $s^2 = \frac{1}{5} [(3 - 4)^2 + (3 - 4)^2 + (3 - 4)^2 + (5 - 4)^2 + (6 - 4)^2] = 1.6$, 可以出现点数 6, 所以 D 错误。

故选 B。

点评:本题考查了平均数、中位数、众数和方差在统计中的应用, 以及各个数据对总体的影响, 主要考查同学们的逻辑推理和数学运算能力, 属较难题。

考向三、统计图表估计与计算

例 4 (多选题)某学校开展了针对学生使用手机问题的专项治理, 效果显著, 现随机抽取该校 100 名学生, 调查他们周六使用手机的时间(单位: min), 数据按照 $[0, 25), [25, 50), \dots, [125, 150]$ 分组, 得到如图 1 所示的频率分布直方图, 则下列说法正确的是()。

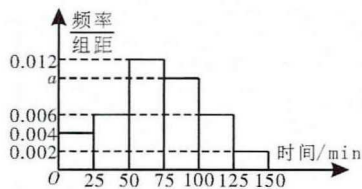


图 1

- A. 这 100 名学生中, 有 25 名学生周六使用手机的时间在 $[75, 100)$ 内
- B. 估计这 100 名学生中, 周六使用手机的平均时间约为 50 min
- C. 估计这 100 名学生中, 周六使用手机时间的第 60 百分位数约为 80
- D. 估计该校周六使用手机时间超过 2 h 的学生比例为 10%

解析: 根据频率分布直方图得 $25a = 1 - 25 \times (0.004 + 0.006 \times 2 + 0.012 + 0.002) = 0.25$, 解得 $a = 0.01$ 。周六使用手机时间在 $[75, 100)$ 内的频率为 $25a = 0.25$, 故有 25 名同学, A 正确; 周六使用手机的平均时间约为 $25 \times (0.004 \times \frac{25}{2} + 0.006 \times \frac{75}{2} + 0.012 \times \frac{125}{2} + 0.01 \times \frac{175}{2} + 0.006 \times \frac{225}{2} + 0.002 \times \frac{275}{2}) = 71.25(\text{min})$,

B 错误;周六使用手机时间在 $[0, 75)$ 内的频率为 $25 \times (0.004 + 0.006 + 0.012) = 0.55$, 则第 60 百分位数是 $75 + \frac{0.6 - 0.55}{0.25} \times 25 = 80$, C 正确;周六使用手机时间在 $[100, 125)$ 内的频率为 $0.006 \times 25 = 0.15$, 使用手机时间在 $[125, 150]$ 内的频率为 $0.002 \times 25 = 0.05$, 所以周六使用手机时间超过 2 h 的学生频率为 $\frac{5}{25} \times 0.006 \times 25 + 0.05 = 0.08$, D 错误。

故选 AC。

点评:此题考查频率分布直方图中的平均数、百分位数的算法,属于基础题。根据频率分布直方图求出图中 a 的值,计算可判断选项 A;直接计算平均数即可判断选项 B;通过频率判断百分位数所在区间,再计算判断选项 C;根据最后两组数据,得出周六使用手机时间超过 2 h 的频率,即可判断选项 D。

考向四、统计图表分析与推断

例 5 (多选题)为了解中学生课外阅读情况,现从某中学随机抽取 200 名学生,收集了他们一年内的课外阅读量(单位:本)等数据,如表 1 所示,图 2 是根据数据绘制的统计图表的一部分。

表 1

学生类别 \ 阅读量		人数				
		$[0, 10)$	$[10, 20)$	$[20, 30)$	$[30, 40)$	$[40, +\infty)$
性别	男	7	31	25	30	4
	女	8	29	26	32	8
学段	初中		25	36	44	11
	高中					

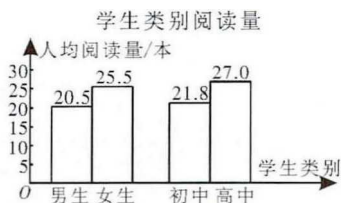


图 2

下面推断合理的是()。

- A. 这 200 名学生阅读量的平均数可能是 26 本
 B. 这 200 名学生阅读量的 75% 分位数在区间 $[30, 40)$ 内

C. 这 200 名学生中的初中生阅读量的中位数一定在区间 $[20, 30)$ 内

D. 这 200 名学生中的初中生阅读量的 25% 分位数可能在区间 $[20, 30)$ 内

解析:由表 1 知,男生共 $7 + 31 + 25 + 30 + 4 = 97$ (人),女生共 $8 + 29 + 26 + 32 + 8 = 103$ (人),这 200 名学生阅读量的平均数为 $20.5 \times \frac{97}{200} + 25.5 \times \frac{103}{200} = 23.075$ (本),故 A 错误;因为 $200 \times 75\% = 150$,所以这 200 名学生阅读量的 75% 分位数是从小到大排序后的第 150 与第 151 个数的平均值,由表格知第 150 与第 151 个数都在区间 $[30, 40)$ 内,故 B 正确;阅读量在区间 $[0, 10)$ 内共有 15 人,若初中生阅读量在区间 $[0, 10)$ 内共有 0 人,则其中位数在区间 $[20, 30)$ 内,若初中生阅读量在区间 $[0, 10)$ 内共有 15 人,则其中位数在区间 $[20, 30)$ 内,故这 200 名学生中的初中生阅读量的中位数一定在区间 $[20, 30)$ 内,故 C 正确;当初中生阅读量在区间 $[0, 10)$ 内共有 0 人时, $(25 + 36 + 44 + 11) \times 25\% = 29$,故这 200 名学生中的初中生阅读量的 25% 分位数是从小到大排序后的第 29 与第 30 个数的平均值,在区间 $[20, 30)$ 内,故 D 正确。

故选 BCD。

点评:本题考查图与表之间的联系,考查平均数、中位数及百分位数的应用。求男女生人数,再由加权平均数求值可判断 A 选项,根据 75% 分位数的定义,结合表格可判断 B 选项,阅读量在区间 $[0, 10)$ 内共有 15 人,可判断无论初中生阅读量在区间 $[0, 10)$ 内是 0 人还是 15 人,其中位数都在区间 $[20, 30)$ 内,从而判断 C 选项,当初中生阅读量在区间 $[0, 10)$ 内共有 0 人时,可判断这 200 名学生中的初中生阅读量的 25% 分位数在区间 $[20, 30)$ 内。

总之,高考中的统计小题常与概率结合在一起,其中方差、平均数、中位数、百分位数的计算是热点。频率分布直方图、扇形图、散点图和折线图几种常见的图表分析也是统计小题中的重点。因此,根据统计图表与数据进行分析 and 计算是统计中的核心问题,也代表着高考对数学核心素养的考查方向。

(责任编辑 王福华)