

考点 2.4.2 用样本估计总体

【知识梳理】

1. 用样本估计总体

1) 画频率分布直方图的步骤：求极差，决定组距与组数，将数据分组，列频率分布表，画频率分布直方图；

2) 第 p 百分位数：一个值，使得这组数据至少有 $p\%$ 的数据小于等于这个值，且至少有 $(100 - p)\%$ 的数据大于或等于这个值；

2. 数字特征

1) 平均数： $\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$ ；取值为 $x_1, x_2, \cdots, x_n$ 的频率分别为 $p_1, p_2, \cdots, p_n$ ，则其

平均数为 $x_1p_1 + x_2p_2 + \cdots + x_np_n$ ；

注意：频率分布表计算平均数要取组中值；

2) 方差与标准差：一组样本数据 $x_1, x_2, \cdots, x_n$ ，

$$\text{方差: } s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2;$$

$$\text{标准差: } s = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$

注：方差与标准差越小，说明样本数据越稳定。

【典型例题】

2.4.2.1 概率的分布与计算

难度☆☆☆

【例 1】 难度☆☆☆

一个容量 100 的样本，其数据的分组与各组的频数如表

组别	(0, 10]	(10, 20]	(20, 30]	(30, 40]	(40, 50]	(50, 60]	(60, 70]
频数	12	13	24	15	16	13	7

则样本数据落在(0, 40] 上的频率为 ()

A. 0.42 B. 0.39 C. 0.52 D. 0.64

【例 2】 难度☆☆☆

一个容量为 100 的样本，将其数据按从小到大的顺序分为 8 组，如下表：

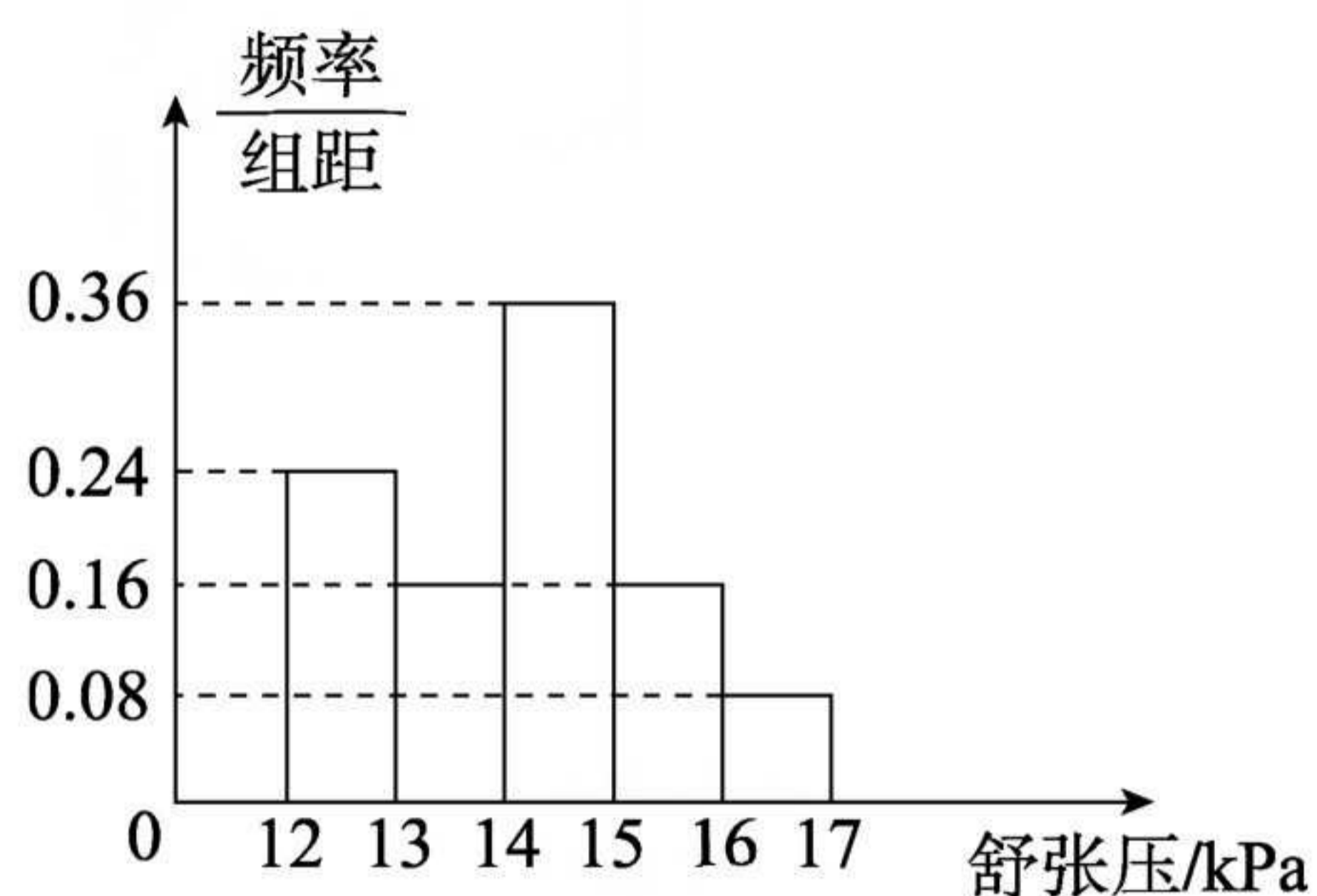
组号	1	2	3	4	5	6	7	8
频数	10	13	x	14	15	13	12	9

则第三组的频数和频率分别是 ()

A. 14 和 0.14 B. 0.14 和 14 C. $\frac{1}{14}$ 和 0.14 D. 14 和 $\frac{1}{14}$

3 ☆☆☆

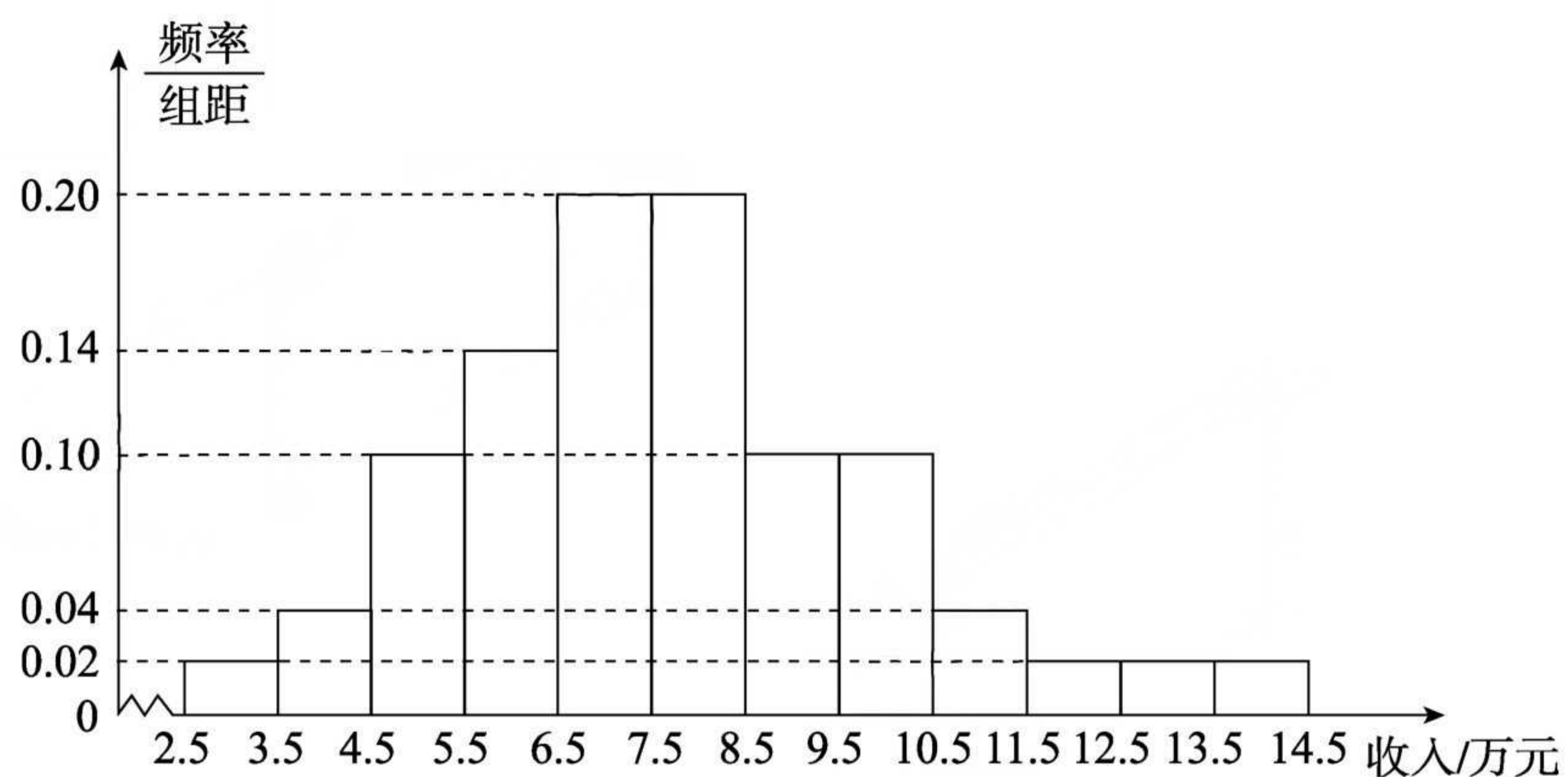
(2022·天津)为研究某药品的疗效,选取若干名志愿者进行临床试验,所有志愿者的舒张压数据(单位: kPa)的分组区间为 $[12, 13)$, $[13, 14)$, $[14, 15)$, $[15, 16)$, $[16, 17]$, 将其按从左到右的顺序分别编号为第一组, 第二组, $\dots$, 第五组, 右图是根据试验数据制成的频率分布直方图. 已知第一组与第二组共有 20 人, 第三组中没有疗效的有 6 人, 则第三组中有疗效的人数为 ()



- A. 8 B. 12 C. 16 D. 18

【例 4】 难度 ★★☆☆

(2021·甲卷)为了解某地农村经济情况,对该地农户家庭年收入进行抽样调查,将农户家庭年收入的调查数据整理得到如下频率分布直方图:

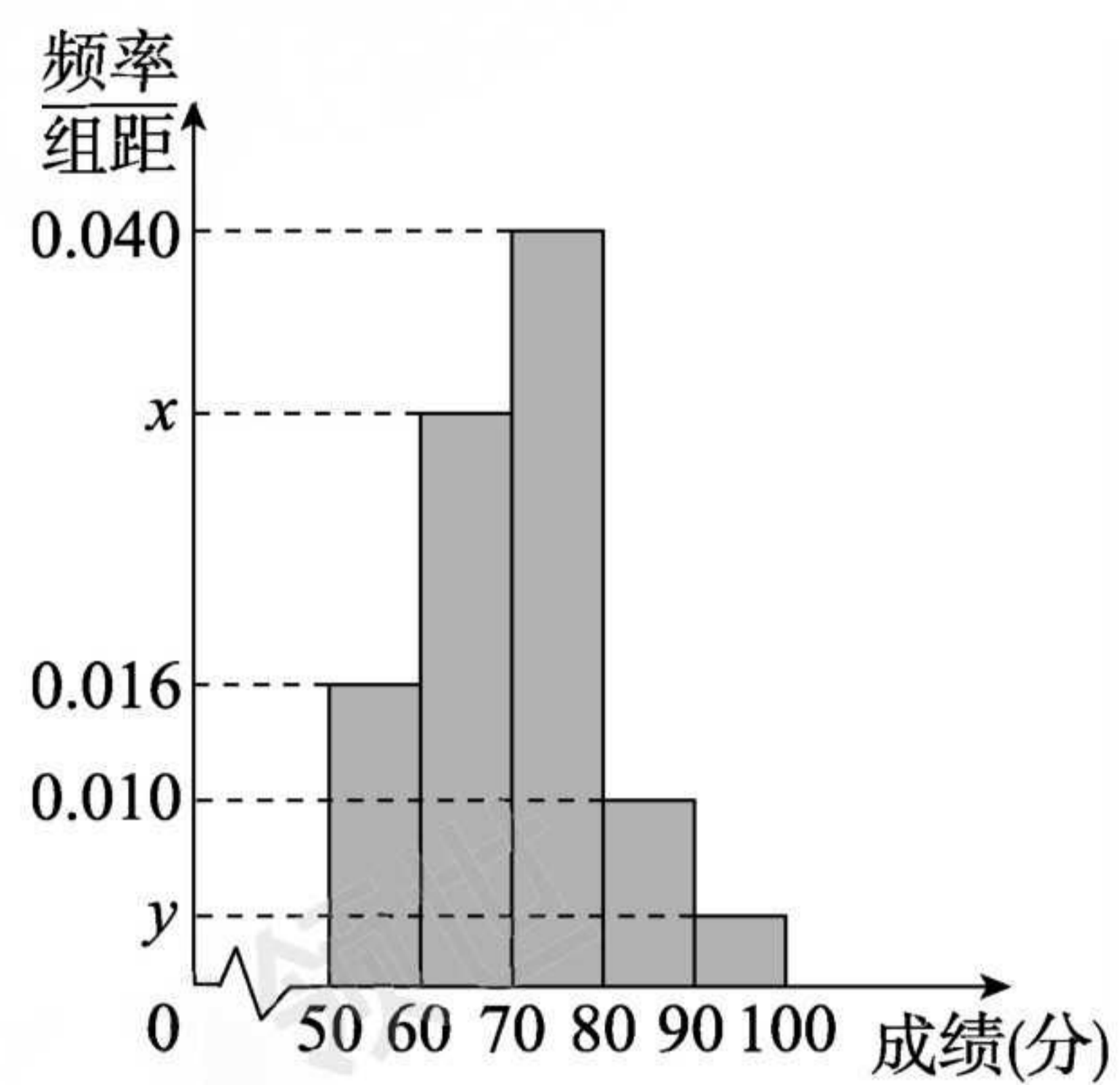


根据此频率分布直方图, 下面结论中不正确的是 ()

- A. 该地农户家庭年收入低于 4.5 万元的农户比率估计为 6%
 B. 该地农户家庭年收入不低于 10.5 万元的农户比率估计为 10%
 C. 估计该地农户家庭年收入的平均值不超过 6.5 万元
 D. 估计该地有一半以上的农户, 其家庭年收入介于 4.5 万元至 8.5 万元之间

5 ☆☆☆

某中学举行了一次“环保知识竞赛”活动. 为了了解本次竞赛学生成绩情况, 从中抽取了部分学生的分数(得分取正整数, 满分为 100 分)作为样本(样本容量为 n)进行统计. 按照 $[50, 60)$, $[60, 70)$, $[70, 80)$, $[80, 90)$, $[90, 100]$ 的分组作出如图所示的频率分布直方图, 但由于不慎丢失了部分数据. 已知得分在 $[50, 60)$ 的有 8 人, 在 $[90, 100]$ 的有 2 人, 由此推测频率分布直方图中的 $x =$ _____.



6 ☆☆☆

某校为了对学生的阅读素养进行监测,随机抽取了 M 名学生进行阅读素养评分.评分规则规定实行百分制计分,现将所得的成绩按照 $[40, 50)$, $[50, 60)$, $[60, 70)$, $[70, 80)$, $[80, 90)$, $[90, 100]$ 分成 6 组,并根据所得数据作出了如下所示的频数与频率的统计表和频率分布直方图.由于一些特殊原因,统计表和直方图都已残缺,请对照图中现有信息按要求还原数据.

分组	频数	频率
$[40, 50)$	5	
$[50, 60)$	25	p
$[60, 70)$		0.30
$[70, 80)$		
$[80, 90)$		
$[90, 100)$		
合计	M	1

- (1) 求出表中 M , p 及图中 a , b 的值;
 (2) 估计该校学生阅读素养的成绩中位数以及平均数(同一组中的数据用该组区间的中点值代表该组数据平均值).

