

专题六 概率与统计

考点限时训练(十五) 第1讲 统计与统计案例

A组 基础演练

1. 某企业在甲、乙、丙、丁四个城市分别有 150 个, 120 个, 190 个, 140 个销售点. 为了调查产品的质量, 需从这 600 个销售点中抽取一个容量为 100 的样本, 记这项调查为①; 在丙城市有 20 个特大型销售点, 要从中抽取 8 个调查, 记这项调查为②, 则完成①, ②这两项调查宜采用的抽样方法依次为

- A. 分层抽样法、系统抽样法
- B. 分层抽样法、简单随机抽样法
- C. 系统抽样法、分层抽样法
- D. 简单随机抽样法、分层抽样法

2. 某餐厅的原料费支出 x (单位: 万元) 与销售额 y (单位: 万元) 之间的有关数据如下表所示, 根据表中提供的全部数据, 用最小二乘法得出 y 与 x 的线性回归方程为 $\hat{y} = 8.5x + 7.5$, 则表中的 m 的值为

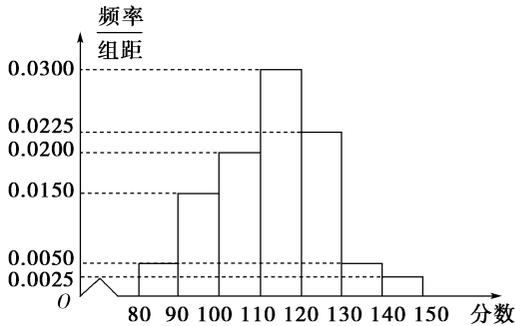
x	2	4	5	6	8
y	25	35	m	55	75

- A. 50
- B. 55
- C. 60
- D. 65

3. 某疾病研究所想知道吸烟与患肺病是否有关, 于是随机抽取 1000 名成年人调查是否吸烟及是否患有肺病, 得到 2×2 列联表, 经计算得 $K^2 = 5.231$, 已知在假设吸烟与患肺病无关的前提条件下, $P(K^2 \geq 3.841) = 0.05$, $P(K^2 \geq 6.635) = 0.01$, 则该研究所可以

- A. 有 95% 以上的把握认为“吸烟与患肺病有关”
- B. 有 95% 以上的把握认为“吸烟与患肺病无关”
- C. 有 99% 以上的把握认为“吸烟与患肺病有关”
- D. 有 99% 以上的把握认为“吸烟与患肺病无关”

4. 如图是某校高三(1)班上学期期末数学考试成绩整理得到的频率分布直方图, 由此估计该班学生成绩的众数、中位数分别为



- A. 105, 103
- B. 115, 125
- C. 125, 113.3
- D. 115, 113.3

5. 图示茎叶图记录了甲、乙两组各五名学生在一次英语听力测试中的成绩(单位: 分). 已知甲组数据的中位数为 15, 乙组数据的平均数为 16.8, 则 x, y 的值分别为

甲组			乙组	
	9	0	9	
x	2	1	5	y
	7	4	2	4

- A. 2, 5
- B. 5, 5
- C. 5, 8
- D. 8, 8

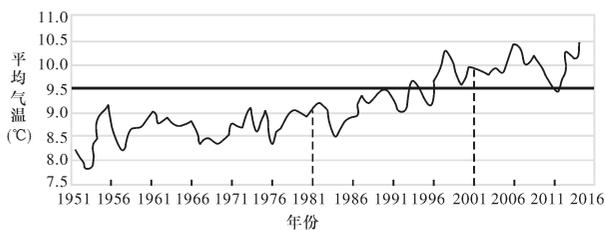
6. 下列说法中正确的个数为

- ①若样本数据 x_1, x_2, \dots, x_n 的平均数 $\bar{x} = 5$, 则样本数据 $2x_1 + 1, 2x_2 + 1, \dots, 2x_n + 1$ 的平均数为 10;
- ②将一组数据中的每个数据都减去同一个数后, 平均数与方差均没有变化;
- ③采用系统抽样法从某班按学号抽取 5 名同学参加活动, 学号为 5, 16, 27, 38, 49 的同学均被选出, 则该班学生人数可能为 60.

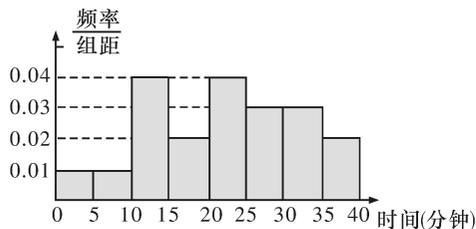
- A. 0
- B. 1
- C. 2
- D. 3

7. 如图是 1951~2016 年中国年平均气温折线图, 虚线处是 1981 年和 2001 年的年平均气温, 图中粗黑线表示 1981~2010 年的平均值. 根据折线图, 可以判断下列结论正确的是

答案	题号
	1
	2
	3
	4
	5
	6
	7
	8



- A. 1951 年以来,我国年平均气温逐年增高
 B. 2001 年以来,我国年平均气温都高于 1981~2010 年的平均值
 C. 1951 年始连续五年年平均气温的方差小于 2001 年始连续五年年平均气温的方差
 D. 2001 年以来,我国年平均气温的平均值高于 1981~2010 年的平均值
8. 某学校随机抽查了本校 20 个同学,调查他们平均每天在课外进行体育锻炼的时间(单位:分钟),根据所得数据的茎叶图,以 5 为组距将数据分为八组,分别是 $[0, 5), [5, 10), \dots, [35, 40]$,作出的频率分布直方图如图所示,则原始的茎叶图可能是

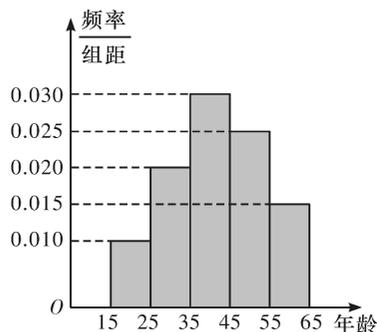


<table style="width: 100%; border-collapse: collapse;"> <tr><td style="border-right: 1px solid black; padding: 2px 5px;">0</td><td style="padding: 2px 5px;">7 6</td></tr> <tr><td style="border-right: 1px solid black; padding: 2px 5px;">1</td><td style="padding: 2px 5px;">7 6 5 4 4 0</td></tr> <tr><td style="border-right: 1px solid black; padding: 2px 5px;">2</td><td style="padding: 2px 5px;">7 5 5 4 2 1 0</td></tr> <tr><td style="border-right: 1px solid black; padding: 2px 5px;">3</td><td style="padding: 2px 5px;">9 5 3 2 0</td></tr> </table> <p style="text-align: center;">A</p>	0	7 6	1	7 6 5 4 4 0	2	7 5 5 4 2 1 0	3	9 5 3 2 0	<table style="width: 100%; border-collapse: collapse;"> <tr><td style="border-right: 1px solid black; padding: 2px 5px;">0</td><td style="padding: 2px 5px;">7 4</td></tr> <tr><td style="border-right: 1px solid black; padding: 2px 5px;">1</td><td style="padding: 2px 5px;">7 6 4 4 1 0</td></tr> <tr><td style="border-right: 1px solid black; padding: 2px 5px;">2</td><td style="padding: 2px 5px;">7 5 5 4 2 1 0</td></tr> <tr><td style="border-right: 1px solid black; padding: 2px 5px;">3</td><td style="padding: 2px 5px;">9 5 3 2 0</td></tr> </table> <p style="text-align: center;">B</p>	0	7 4	1	7 6 4 4 1 0	2	7 5 5 4 2 1 0	3	9 5 3 2 0
0	7 6																
1	7 6 5 4 4 0																
2	7 5 5 4 2 1 0																
3	9 5 3 2 0																
0	7 4																
1	7 6 4 4 1 0																
2	7 5 5 4 2 1 0																
3	9 5 3 2 0																
<table style="width: 100%; border-collapse: collapse;"> <tr><td style="border-right: 1px solid black; padding: 2px 5px;">0</td><td style="padding: 2px 5px;">7 4</td></tr> <tr><td style="border-right: 1px solid black; padding: 2px 5px;">1</td><td style="padding: 2px 5px;">7 7 4 4 4 0</td></tr> <tr><td style="border-right: 1px solid black; padding: 2px 5px;">2</td><td style="padding: 2px 5px;">7 5 5 5 2 1 0</td></tr> <tr><td style="border-right: 1px solid black; padding: 2px 5px;">3</td><td style="padding: 2px 5px;">9 5 3 2 0</td></tr> </table> <p style="text-align: center;">C</p>	0	7 4	1	7 7 4 4 4 0	2	7 5 5 5 2 1 0	3	9 5 3 2 0	<table style="width: 100%; border-collapse: collapse;"> <tr><td style="border-right: 1px solid black; padding: 2px 5px;">0</td><td style="padding: 2px 5px;">7 4</td></tr> <tr><td style="border-right: 1px solid black; padding: 2px 5px;">1</td><td style="padding: 2px 5px;">7 6 4 4 4 0</td></tr> <tr><td style="border-right: 1px solid black; padding: 2px 5px;">2</td><td style="padding: 2px 5px;">8 7 6 5 2 1 0</td></tr> <tr><td style="border-right: 1px solid black; padding: 2px 5px;">3</td><td style="padding: 2px 5px;">9 5 5 2 0</td></tr> </table> <p style="text-align: center;">D</p>	0	7 4	1	7 6 4 4 4 0	2	8 7 6 5 2 1 0	3	9 5 5 2 0
0	7 4																
1	7 7 4 4 4 0																
2	7 5 5 5 2 1 0																
3	9 5 3 2 0																
0	7 4																
1	7 6 4 4 4 0																
2	8 7 6 5 2 1 0																
3	9 5 5 2 0																

9. 假设要考查某公司生产的 500 克袋装牛奶的质量是否达标,现从 800 袋牛奶中抽取 60 袋牛奶进行检验,利用随机数表抽样时,先将 800 袋牛奶按 000,001, \dots ,799 进行编号,如果从随机数表第 8 行第 8 列开始向右读,那么抽取检测的第 5 袋牛奶的编号为_____。(下面摘取了随机数表第 7 行至第 9 行)
- 8442 1753 3157 2455 0688 7704 7447
 6721 7633 5025 8392 1206 76
 6301 6378 5916 9556 6715 5610 5071
 7512 8673 5807 4439 5238 79
 3321 1234 2978 6456 0782 5242 0744
 3815 5100 1342 9966 0279 54

10. 长沙市旅游局为了了解橘子洲景区在大众中的熟知度,随机对 15~65 岁的人群抽样了 n 人,问题是“橘子洲景区是几 A 级旅游景区?”统计结果如下图表.

组号	分组	回答正确的人数	回答正确的人数占本组的频率
第 1 组	$[15, 25)$	a	0.5
第 2 组	$[25, 35)$	18	x
第 3 组	$[35, 45)$	b	0.9
第 4 组	$[45, 55)$	9	0.36
第 5 组	$[55, 65]$	3	y



- (1) 分别求出 a, b, x, y 的值;
 (2) 从第 2, 3, 4 组回答正确的人中用分层抽样的方法抽取 6 人,问第 2, 3, 4 组每组各抽取多少人?
 (3) 在(2)抽取的 6 人中随机抽取 2 人,求所抽取的人中恰好没有第 3 组人的概率.

B组 强化提高

11. 下面的茎叶图记录了甲、乙两代表队各 10 名同学在一次英语听力比赛中的成绩(单位:分). 已知甲代表队数据的中位数为 76, 乙代表队数据的平均数是 75.

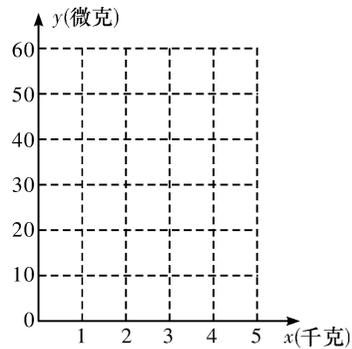
甲队		乙队
	5	6
5 4	6	8 8
7 6 x 1 1	7	0 2 y
8 2 0	8	0 6 8 9

- (1) 求 x, y 的值;
- (2) 若分别从甲、乙两队随机各抽取 1 名成绩不低于 80 分的学生, 求抽到的学生中, 甲队学生成绩不低于乙队学生成绩的概率;
- (3) 判断甲、乙两队谁的成绩更稳定, 并说明理由(方差较小者稳定).

12. 菜农定期使用低害杀虫农药对蔬菜进行喷洒, 以防止害虫的危害, 但蔬菜上市时仍存有少量的残留农药, 食用时需要用清水清洗干净. 下表是用清水 x (单位: 千克) 清洗蔬菜 1 千克后, 蔬菜上残留的农药 y (单位: 微克) 的统计表:

x	1	2	3	4	5
y	58	54	39	29	10

- (1) 在下面的坐标系中, 描出散点图, 并判断变量 x 与 y 是正相关还是负相关;
- (2) 若用解析式 $\hat{y} = cx^2 + d$ 作为蔬菜农药残量 \hat{y} 与用水量 x 的回归方程, 令 $w = x^2$, 计算平均值 \bar{w} 与 \bar{y} , 完成以下表格, 求出 \hat{y} 与 x 的回归方程(c, d 保留两位有效数字);



w	1	4	9	16	25
y	58	54	39	29	10
$w_i - \bar{w}$					
$y_i - \bar{y}$					

(3)对于某种残留在蔬菜上的农药,当它的残留量低于 20 微克时对人体无害,为了放心食用该蔬菜,请评估需要用多少千克的清水清洗 1 千克蔬菜?(精确到 0.1,参考数据 $\sqrt{5}\approx 2.236$)(附:对于一组数据 $(u_1, v_1), (u_2, v_2), \dots, (u_n, v_n)$,其回归直线 $v = \alpha + \beta u$ 的斜率和截距的最小二乘法估计分别为:

$$\hat{\beta} = \frac{\sum_{i=1}^n (u_i - \bar{u})(v_i - \bar{v})}{\sum_{i=1}^n (u_i - \bar{u})^2}, \hat{\alpha} = \bar{v} - \hat{\beta}\bar{u}$$

13. 某高三理科班共有 60 名同学参加某次考试,从中随机挑选出 5 名同学,他们的数学成绩 x 与物理成绩 y 如下表:

数学成绩 x	145	130	120	105	100
物理成绩 y	110	90	102	78	70

数据表明 y 与 x 之间有较强的线性关系.

- (1)求 y 关于 x 的线性回归方程;
- (2)该班一名同学的数学成绩为 110 分,利用(1)中的回归方程,估计该同学的物理成绩;
- (3)本次考试中,规定数学成绩达到 125 分为优秀,物理成绩达到 100 分为优秀.若该班数学优秀率与物理优秀率分别为 50% 和 60%,且除去抽走的 5 名同学外,剩下的同学中数学优秀但物理不优秀的同学共有 5 人.能否在犯错误的概率不超过 0.01 的前提下认为数学优秀与物理优秀有关?

参考数据:回归直线的系数

$$\begin{cases} \hat{b} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^n x_i^2 - n\bar{x}^2}, \\ \hat{a} = \bar{y} - \hat{b}\bar{x}. \end{cases}$$

$$K^2 = \frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)},$$

$$P(K^2 \geq 6.635) = 0.01, P(K^2 \geq 10.828) = 0.001.$$