

第九章 统计与成对数据的统计分析（测试）

（考试时间：120 分钟 试卷满分：150 分）

注意事项：

1. 答卷前，考生务必将自己的姓名、准考证号填写在答题卡上。
2. 回答选择题时，选出每小题答案后，用铅笔把答题卡对应题目的答案标号涂黑。如需改动，用橡皮擦干净后，再选涂其他答案标号。回答非选择题时，将答案写在答题卡上。写在本试卷上无效。
3. 考试结束后，将本试卷和答题卡一并交回。

第一部分（选择题 共 58 分）

一、选择题：本题共 8 小题，每小题 5 分，共 40 分。在每小题给出的四个选项中，只有一项是符合题目要求的。

1. 已知一组数据：3,5,7,x,9 的平均数为 6，则该组数据的 40% 分位数为（ ）

- A. 4.5 B. 5 C. 5.5 D. 6

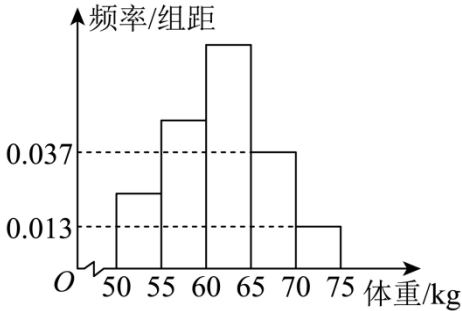
2. 法国当地时间 2024 年 7 月 26 日晚，第三十三届夏季奥林匹克运动会在巴黎举行开幕式。“奥林匹克之父”顾拜旦曾经说过，奥运会最重要的不是胜利，而是参与；对人生而言，重要的不是凯旋，而是拼搏。为弘扬奥运精神，某学校组织高一年级学生进行奥运专题的答题活动。为了调查男生和女生对奥运会的关注程度，在高一年级随机抽取 10 名男生和 10 名女生的竞赛成绩（满分 100 分），按从低到高的顺序排列，得到下表中的样本数据：

男生	82	85	86	87	88	90	90	92	94	96
女生	82	84	85	87	87	87	88	88	90	92

则下列说法错误的是（ ）

- A. 男生样本数据的 25% 分位数是 86
- B. 男生样本数据的中位数小于男生样本数据的众数
- C. 女生样本数据中去掉一个最高分和一个最低分后所得数据的平均数不变
- D. 女生样本数据中去掉一个最高分和一个最低分后所得数据的方差不变
3. 已知一组数据丢失了其中一个，另外六个数据分别是 8, 8, 8, 10, 11, 16，若这组数据的平均数、中位数、众数依次成等差数列，则丢失数据的所有可能值的和为（ ）
- A. 12 B. 20 C. 25 D. 27
4. “绿水青山就是金山银山”的理念深入人心，人民群众的生态环境获得感、幸福感、安全感不断提升。某校高一年级举行环保知识竞赛，共 500 人参加，若参赛学生成绩的第 60 百分位数是 80 分，则关于竞赛成绩不小于 80 的人数的说法正确的是（ ）
- A. 至少为 300 人 B. 至少为 200 人
- C. 至多为 300 人 D. 至多为 200 人
5. 为了解某校今年准备报考飞行员的学生的体重情况，对所得的体重数据（单位：kg）进行分组，区间为

$[50,55), [55,60), [60,65), [65,70), [70,75]$ ，将其按从左到右的顺序分别编号为第一组，第二组，……，第五组。画出频率分布直方图（如图所示），已知第一组，第二组和第三组的频率之比为1:2:3，且第一组的频数为6，则报考飞行员的学生人数是（ ）



- A. 48 B. 5 C. 54 D. 60

6. 已知数据 x_1, x_2, \dots, x_5 ($x_i \in \mathbf{Z}, i=1,2,\dots,5$) 的平均数、中位数、方差均为4，则这组数据的极差为（ ）

- A. 3 B. 4 C. 5 D. 6

7. 已知实数 $x_0, x_1, \dots, x_{2024}$ ，则使 $\sum_{i=0}^{2024} |x_i - k|$ 和 $\sum_{i=0}^{2024} (x_i - k)^2$ 最小的实数 k 分别为 $x_0, x_1, \dots, x_{2024}$ 的（ ）

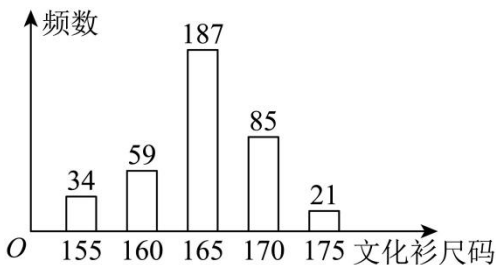
- A. 平均数；平均数 B. 平均数；中位数
C. 中位数；平均数 D. 标准差；平均数

8. 某校积极开展“戏曲进校园”活动，为了解该校各班参加戏曲兴趣小组的人数，从全校随机抽取5个班级，把每个班级参加该小组的人数作为样本数据。已知样本平均数为7，样本标准差为2，且样本数据互不相等，则该样本数据的极差为（ ）

- A. 3 B. 4 C. 5 D. 6

二、选择题：本题共3小题，每小题6分，共18分。在每小题给出的选项中，有多项符合题目要求。全部选对的得6分，部分选对的得部分分，有选错的得0分。

9. 某公司计划组织秋游活动，定制了一套文化衫，女职工需要不同尺码文化衫的频数如图。



根据图中数据，下列结论正确的是（ ）

- A. 文化衫尺码的众数为187 B. 文化衫尺码的平均数为165
C. 文化衫尺码的方差为28 D. 文化衫尺码的中位数为165

10. 已知一组数据 x_1, x_2, \dots, x_m 的平均数为 \bar{x} ，另一组数据 y_1, y_2, \dots, y_n 的平均数为 \bar{y} 。若数据 $x_1, x_2, \dots, x_m, y_1, y_2, \dots, y_n$ 的平均数为 \bar{z} ，则（ ）

- A. 当 $m = n$ 时, $\bar{z} = \frac{\bar{x} + \bar{y}}{2}$ B. 当 $\bar{z} = \frac{\bar{x} + \bar{y}}{2}$ 时, $m = n$
- C. 当 $\bar{x} = \bar{y}$ 时, $\bar{z} = \frac{\bar{x} + \bar{y}}{2}$ D. 当 $\bar{z} > \bar{x}$ 时, $\bar{y} > \bar{x}$

11. 北京时间 2024 年 8 月 12 日凌晨, 第 33 届法国巴黎奥运会闭幕式正式举行, 中国体育代表团以出色的表现再次证明了自己的实力, 最终取得了 40 枚金牌、27 枚银牌和 24 枚铜牌的最佳境外参赛成绩, 也向世界展示了中国体育的蓬勃发展和运动员们顽强拼搏的精神. 某校团委为发扬奥运体育精神举办了竞技比赛, 此比赛共有 5 名同学参加, 赛后经数据统计得到该 5 名同学在此次比赛中所得成绩的平均数为 8, 方差为 4, 比赛成绩 $x \in [0, 15]$, 且 $x \in \mathbb{N}^*$, 则该 5 名同学中比赛成绩的最高分可能为 ()

- A. 13 B. 12 C. 11 D. 10

第二部分 (非选择题 共 92 分)

三、填空题: 本题共 3 小题, 每小题 5 分, 共 15 分.

12. 现利用随机数表发从编号为 00, 01, 02, ..., 18, 19 的 20 支水笔中随机选取 6 支, 选取方法是从下列随机数表第 1 行的第 9 个数字开始由左到右依次选取两个数字, 则选出来的第 6 支水笔的编号为_____.

95226000 49840128 66175168 39682027 43772366 27096623
92580856 43890990 06482834 59741858 29778149 64608925

13. 某公司对来应聘的人进行笔试, 统计出 200 名应聘者的笔试成绩, 整理得到下表:

组号	1	2	3	4	5	6
成绩分组	[40, 50)	[50, 60)	[60, 70)	[70, 80)	[80, 90)	[90, 100]
累积频率	0.05	0.15	a	$a + 0.30$	$2a + 0.10$	$2a + 0.30$

注: 第 n 组的累积频率指的是前 n 组的频率之和.

若公司计划 150 人进入面试, 则估计参加面试的最低分数线为_____.

14. 已知一组统计数据 x_1, x_2, \dots, x_n 的平均数为 \bar{x} , 方差为 s^2 , 则函数 $f(x) = \sum_{i=1}^n (x - x_i)^2$ 的最小值为_____.

四、解答题: 本题共 5 小题, 共 77 分. 解答应写出文字说明、证明过程或演算步骤.

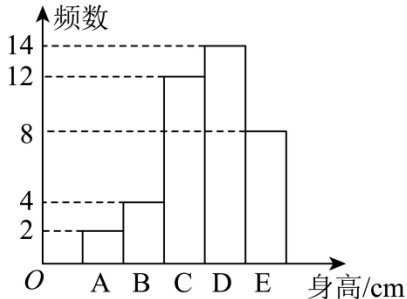
15. (13 分)

为了了解某中学学生的身高情况, 随机对该校男生、女生的身高进行抽样调查, 已知抽取的样本中, 男生、女生的人数相同, 根据所得数据绘制成如图所示的统计图表.

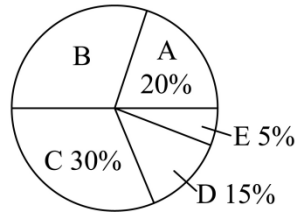
组别	身高 (cm)
A	$x < 150$
B	$150 \leq x < 155$
C	$155 \leq x < 160$

	$160 \leq x <$
	$x \geq 165$

男生身高情况直方图



女生身高情况扇形统计图



根据图表中信息，回答下列问题：

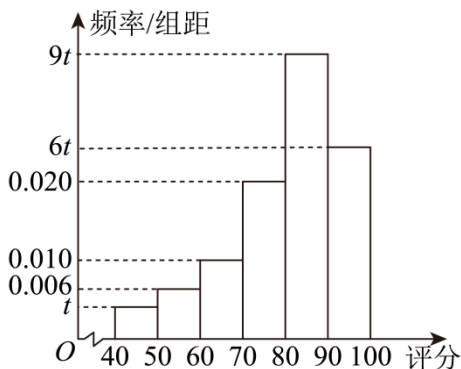
- (1)在样本中，男生身高的中位数落在_____组（填组别序号），女生身高在B组的有_____人；
- (2)在样本中，身高在 $150 \leq x < 155$ 之间的共有_____人，身高人数最多的在_____组（填组别序号）；
- (3)已知该校共有男生500人，女生480人，请估计身高在 $155 \leq x < 165$ 之间的学生约有多少人？

16. (15分)

某公司为了了解顾客对其旗下产品的满意程度，随机抽取 n 名顾客进行满意度问卷调查，按所得评分（满分100分）从低到高将满意度分为四个等级：

调查评分	[40,50)	[50,60)	[60,70)	[70,80)	[80,90)	[90,100]
满意度等级	不满意		一般		良好	满意

并绘制如图所示的频率分布直方图。已知调查评分在 $[70,80)$ 的顾客为80人。



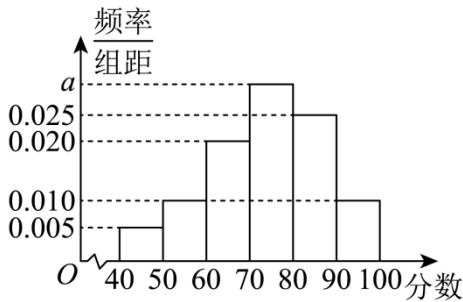
- (1)求 n 的值及频率分布直方图中 t 的值；

(2)若某段时间有 10000 名顾客购买该公司的产品,请估计这 10000 名顾客中对该公司产品满意度达到“满意”的人数;

(3)该公司设定的预案是:以抽取的样本作为参考,若顾客满意度评分的均值低于 80 分,则需要对该公司旗下产品进行调整,否则不需要调整、根据你所学的统计知识,判断该公司是否需要对其旗下产品进行调整,并说明理由。(每组数据以区间的中点值代替)

17. (15 分)

为进一步推动防范电信网络诈骗工作,预防和减少电信网络诈骗案件的发生,某市开展防骗知识大宣传活动.举办了“网络防骗”知识竞赛,从所有答卷中随机抽取 100 份作为样本,将样本的成绩(满分 100 分,成绩均为不低于 40 分的整数)分成六段: [40,50), [50,60), ..., [90,100]得到如图所示的频率分布直方图.



(1)求图中 a 的值,根据频率分布直方图计算样本成绩的平均数和下四分位数;

(2)已知若总体划分为 2 层,通过分层随机抽样,各层抽取的样本量、样本平均数和样本方差分别为: $m, \bar{x}, s_1^2; n, \bar{y}, s_2^2$, 记总的样本平均数为 $\bar{\omega}$, 样本方差为 s^2 .

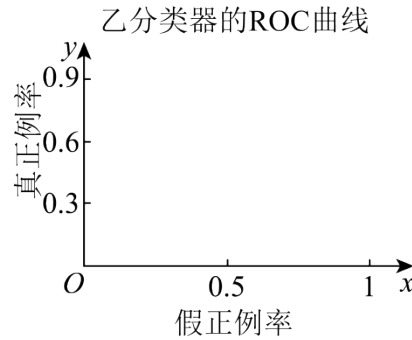
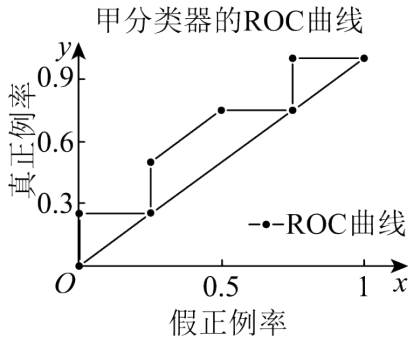
证明: $s^2 = \frac{1}{m+n} \{ m[s_1^2 + (\bar{x} - \bar{\omega})^2] + n[s_2^2 + (\bar{y} - \bar{\omega})^2] \};$

(3)已知落在[50,60)的平均成绩是 59, 方差是 7, 落在[60,70)的平均成绩为 65, 方差是 4, 求两组样本成绩的总平均数 \bar{z} 和总方差 s^2 .

2	正例	0.58	0.53
3	反例	0.15	0.13
4	反例	0.62	0.39
5	正例	0.47	0.87
6	反例	0.47	0.53
7	反例	0.33	0.11

8	正例	0.77	0.63
---	----	------	------

表 2 甲, 乙分类器对于相同 8 个样例的预测数据



(1) 当概率阈值为 0.47 时, 求甲分类器的 ROC 曲线中的对应点;

(2) 在图 2 中绘制乙分类器对应的 ROC 曲线(无需说明绘图过程), 并直接写出甲, 乙两分类器的 ROC 曲线与 x 轴, 直线 $x=1$ 所围封闭图形的面积;

(3) 按照上述思路, 比较甲, 乙两分类器的预测效果, 并直接写出理想分类器的 ROC 曲线与 x 轴, 直线 $x=1$ 所围封闭图形的面积为 1 的充要条件.

19. (17 分)

将 2024 表示成 5 个正整数 x_1, x_2, x_3, x_4, x_5 之和, 得到方程 $x_1 + x_2 + x_3 + x_4 + x_5 = 2024$ ①, 称五元有序数组 $(x_1, x_2, x_3, x_4, x_5)$ 为方程①的解, 对于上述的五元有序数组 $(x_1, x_2, x_3, x_4, x_5)$, 当 $1 \leq i, j \leq 5$ 时, 若 $\max(x_i - x_j) = t (t \in \mathbb{N})$, 则称 $(x_1, x_2, x_3, x_4, x_5)$ 是 t -密集的一组解.

(1) 方程①是否存在一组解 $(x_1, x_2, x_3, x_4, x_5)$, 使得 $x_{i+1} - x_i (i=1, 2, 3, 4)$ 等于同一常数? 若存在, 请求出该常数; 若不存在, 请说明理由;

(2) 方程①的解中共有多少组是 1-密集的?

(3) 记 $S = \sum_{i=1}^5 x_i^2$, 问 S 是否存在最小值? 若存在, 请求出 S 的最小值; 若不存在, 请说明理由.

第九章 统计与成对数据的统计分析（测试）

（考试时间：120 分钟 试卷满分：150 分）

注意事项：

1. 答卷前，考生务必将自己的姓名、准考证号填写在答题卡上。
2. 回答选择题时，选出每小题答案后，用铅笔把答题卡对应题目的答案标号涂黑。如需改动，用橡皮擦干净后，再选涂其他答案标号。回答非选择题时，将答案写在答题卡上。写在本试卷上无效。
3. 考试结束后，将本试卷和答题卡一并交回。

第一部分（选择题 共 58 分）

一、选择题：本题共 8 小题，每小题 5 分，共 40 分。在每小题给出的四个选项中，只有一项是符合题目要求的。

1. 已知一组数据：3,5,7,x,9 的平均数为 6，则该组数据的 40% 分位数为（ ）

- A. 4.5 B. 5 C. 5.5 D. 6

【答案】C

【解析】依题意， $\frac{3+5+7+x+9}{5}=6$ ，解得 $x=6$ ，

将数据从小到大排列可得：3,5,6,7,9，

又 $5 \times 0.4 = 2$ ，则 40% 分位数为 $\frac{5+6}{2} = 5.5$ 。

故选：C。

2. 法国当地时间 2024 年 7 月 26 日晚，第三十三届夏季奥林匹克运动会在巴黎举行开幕式。“奥林匹克之父”顾拜旦曾经说过，奥运会最重要的不是胜利，而是参与；对人生而言，重要的不是凯旋，而是拼搏。为弘扬奥运精神，某学校组织高一年级学生进行奥运专题的答题活动。为了调查男生和女生对奥运会的关注程度，在高一年级随机抽取 10 名男生和 10 名女生的竞赛成绩（满分 100 分），按从低到高的顺序排列，得到下表中的样本数据：

男生	82	85	86	87	88	90	90	92	94	96
女生	82	84	85	87	87	87	88	88	90	92

则下列说法错误的是（ ）

- A. 男生样本数据的 25% 分位数是 86
- B. 男生样本数据的中位数小于男生样本数据的众数
- C. 女生样本数据中去掉一个最高分和一个最低分后所得数据的平均数不变
- D. 女生样本数据中去掉一个最高分和一个最低分后所得数据的方差不变

【答案】D

【解析】对于 A： $10 \times 25\% = 2.5$ ，所以男生样本数据的 25% 分位数是 86，故 A 正确；

对于 B: 男生样本数据的中位数为 $\frac{88+90}{2}=89$, 男生样本数据的众数为 90, 故 B 正确;

对于 C: 女生样本数据的平均数为 $\frac{1}{10}(82+84+85+87\times 3+88\times 2+90+92)=87$,

女生样本数据中去掉一个最高分和一个最低分后所得数据的平均数为 $\frac{1}{8}(84+85+87\times 3+88\times 2+90)=87$, 故

C 正确;

对于 D: 女生样本数据中去掉一个最高分和一个最低分后所得数据的平均数不变,

但是极差变小, 所以方差变小, 故 D 错误.

故选: D

3. 已知一组数据丢失了其中一个, 另外六个数据分别是 8, 8, 8, 10, 11, 16, 若这组数据的平均数、中位数、众数依次成等差数列, 则丢失数据的所有可能值的和为 ()

A. 12

B. 20

C. 25

D. 27

【答案】D

【解析】这 7 个数据的众数是 8, 设丢失的数据是 x ,

则平均数为 $\frac{x+8+8+8+10+11+16}{7}=\frac{x+61}{7}$,

若 $x < 8$, 则中位数是 8, 则 $\frac{x+61}{7}+8=8\times 2$, 解得 $x=-5$.

若 $x=8$, 则中位数是 8, 平均数 $\frac{x+61}{7}=\frac{69}{7}$,

此时 $\frac{69}{7}, 8, 8$ 不成等差数列, 不符合题意.

若 $8 < x < 10$, 则中位数是 x , 则 $\frac{x+61}{7}+8=2x$, 解得 $x=9$.

若 $x=10$, 则中位数是 10, 则 $\frac{x+61}{7}=\frac{71}{7}$,

$\frac{71}{7}, 10, 8$ 不成等差数列, 不符合题意.

若 $x > 10$, 则中位数是 10, 则 $\frac{x+61}{7}+8=10\times 2$, 解得 $x=23$.

所以丢失数据的所有可能值的和为 $-5+9+23=27$.

故选: D

4. “绿水青山就是金山银山”的理念深入人心, 人民群众的生态环境获得感、幸福感、安全感不断提升. 某校高一年级举行环保知识竞赛, 共 500 人参加, 若参赛学生成绩的第 60 百分位数是 80 分, 则关于竞赛成绩不小于 80 分的人数的说法正确的是 ()

A. 至少为 300 人

B. 至少为 200 人

C. 至多为 300 人

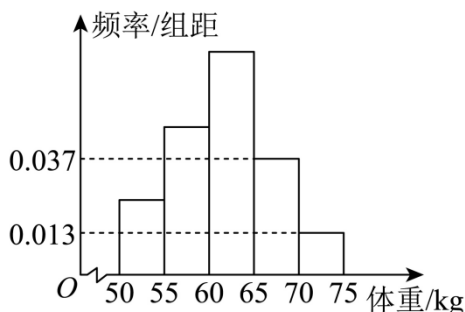
D. 至多为 200 人

【答案】D

【解析】由题意， $500 \times 60\% = 300$ ，因此竞赛成绩不小于80分的人数至多有 $500 - 300 = 200$ 人，

故选：D.

5. 为了解某校今年准备报考飞行员的学生的体重情况，对所得的体重数据（单位：kg）进行分组，区间为 $[50, 55), [55, 60), [60, 65), [65, 70), [70, 75]$ ，将其按从左到右的顺序分别编号为第一组，第二组，……，第五组. 画出频率分布直方图（如图所示），已知第一组，第二组和第三组的频率之比为1:2:3，且第一组的频数为6，则报考飞行员的学生人数是（ ）



A. 48

B. 5

C. 54

D. 60

【答案】A

【解析】由题前三组频率之和为 $1 - (0.013 + 0.037) \times 5 = 0.75$ ，

又第一组、第二组和第三组的频率之比为1:2:3，

所以第一组的频率为 $0.75 \times \frac{1}{1+2+3} = 0.125$ ，又第一组的频数为6，

所以报考飞行员的学生人数为 $\frac{6}{0.125} = 48$ 人.

故选：A.

6. 已知数据 x_1, x_2, \dots, x_5 ($x_i \in \mathbf{Z}$, $i=1, 2, \dots, 5$) 的平均数、中位数、方差均为4，则这组数据的极差为（ ）

A. 3

B. 4

C. 5

D. 6

【答案】D

【解析】根据题意，不妨设 $x_1 < x_2 < x_3 < x_4 < x_5$ ，且 $x_i \in \mathbf{Z}$ ，可得 $x_3 = 4$ ，

由平均数为4，得 $\frac{1}{5}(x_1 + x_2 + x_3 + x_4 + x_5) = 4$ ，即 $x_1 + x_2 + x_3 + x_4 + x_5 = 20$ ；

由方差为4，得 $\frac{1}{5}[(x_1 - 4)^2 + (x_2 - 4)^2 + (x_3 - 4)^2 + (x_4 - 4)^2 + (x_5 - 4)^2] = 4$ ，即 $x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2 = 100$ ；

联立 $\begin{cases} x_1 + x_2 + x_3 + x_4 + x_5 = 20 \\ x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2 = 100 \end{cases}$ ，由 $x_i \in \mathbf{Z}$ 可解得 $x_1 = 1, x_2 = 3, x_4 = 5, x_5 = 7$ ；

根据极差定义可得这组数据的极差为 $x_5 - x_1 = 6$ 。

故选：D

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/247054114154010010>