

GRADUS

数学 III

循迹而上 · 拾级登顶

寻迹 · 梦叶子 编著

目录

第一章 计数原理	1
1.1 分类加法与分步乘法计数原理	1
1.2 排列	2
1.3 组合	4
1.4 二项式定理	6
第二章 概率	9
2.1 随机事件与古典概型	9
2.2 条件概率与全概率	10
2.3 事件的独立性与二项分布	12
2.4 数字特征与正态分布初步	14
第三章 统计	17
3.1 抽样方法	17
3.2 频率分布直方图与统计图表	19
3.3 百分位数与四分位数	21
3.4 集中趋势与离散程度	22
第四章 成对数据的统计分析	27
4.1 相关关系	27
4.2 一元线性回归模型	29
4.3 独立性检验	31
第五章 数学建模与拓展	37
5.1 建模四步：从现实到数学再回现实	37
5.2 模型选择与检验：数据形状决定模型家族	39
5.3 拓展 · 矩阵初步	41

5.4 拓展·逻辑初步与数学史、新技术	43
第六章 卷末综合检测	45

第一章 计数原理

1.1 分类加法与分步乘法计数原理

考点精要

Q1 分类加法计数原理什么时候用？

A “要么…要么…”的问题——选完这一项，整件事当场完成（如“只吃一顿，选一份套餐”）。各类方法互不相同，合起来不重不漏。

Q2 分步乘法计数原理什么时候用？

A “先…再…”的问题——必须把每一步都走完，这件事才算完成（如“先打主食再打饮料”）。每一步的选法数要恒定。

Q3 分类必须满足什么？

A “不重不漏”——两类之间没有同一种方法被数两次（不重），所有可能都被某一类盖住（不漏）。

Q4 分步必须满足什么？

A “层层递进、每步独立”——某一步的选法数不因前面步骤的不同选择而改变；若会变，通常要分类讨论。

Q5 复杂问题怎么判断该分类还是分步？

A 设身处地“完成一件事”，问自己：选完这个项，事情完成了吗？完成了就“加”；没完成还要选就“乘”。复杂问题常“先分类、每类内再分步”。

例题精讲

例题 1.1

（分类与分步的判别）某餐厅提供 3 种主食、4 种荤菜、2 种素菜。（1）任意买 1 份（主食、荤菜、素菜任选其一），共有多少种选择？（2）选“1 种主食 + 1 种荤菜 + 1 种素菜”为一餐，共有多少种搭配？

例题 1.2

(数字组数, 0 不能做首位) 用 0, 1, 2, 3 四个数字组成无重复数字的四位数, 共有多少个?

知识点 (1) 分类加法与分步乘法的判别与应用

A1. 从甲地到乙地有 4 条路, 从乙地到丙地有 3 条路, 从甲地到丙地另有 2 条直达路。从甲地到丙地共有多少种不同的走法? ()

A. 9 B. 12 C. 14 D. 24

A2. 集合 $A = \{a, b, c, d\}$ 共有多少个子集? ()

A. 8 B. 12 C. 16 D. 32

A3. 用红、黄、蓝 3 种颜色给 A、B、C 三个区域涂色 (区域排成一列), 要求相邻区域颜色不同, 共有多少种涂法? ()

A. 6 B. 9 C. 12 D. 27

A4. 从 5 种主食和 3 种饮料中各选一种, 配成 “一主食一饮料”, 共有多少种搭配? ()

A. 8 B. 15 C. 16 D. 30

A5. 从 4 名男生、3 名女生中任选 1 人参加活动, 共有多少种选法? ()

A. 4 B. 7 C. 12 D. 24

A6. 书架上有 3 本不同的数学书和 4 本不同的语文书, 从中任取 1 本, 共有多少种取法? ()

A. 3 B. 4 C. 7 D. 12

M1. 用 0, 1, 2, 3 四个数字组成无重复数字的四位数。小刚按 “首位为 1、2、3” 分三类, 算得 $3 \times 3 \times 2 \times 1 = 18$ 个。小红说: “分类应改为按末位分: 末位为 0、1、2、3 四类, 共 $4 \times 3 \times 2 \times 1 = 24$ 个。” 下列说法正确的是 ()

A. 小刚算错了, 四位数共 24 个

B. 小刚对, 小红错——首位为 0 时得到的是三位数, 不算四位数

C. 两人都对

D. 两人都错, 四位数共 36 个

C1. (密码组合) 某银行要求客户设置 4 位数字密码, 每一位从 0 9 中选取。(1) 若允许数字重复, 共有多少种密码? (2) 若要求 4 位数字互不相同, 共有多少种密码? (3) 若要求密码中至少含 1 个数字 0 (首位可以是 0), 共有多少个这样的密码?

1.2 排列

考点精要

Q1 什么是排列?

A 从 n 个不同元素中取出 m ($m \leq n$) 个, 按一定顺序排成一列。判定标尺: 把两个结果里的元素换位置, 若变成不同的结果 $\rightarrow$ 排列。

Q2 排列数公式是什么?

A $P_n^m = n(n-1)\cdots(n-m+1) = \frac{n!}{(n-m)!}$; 全排列 $P_n^n = n!$ 。

Q3 特殊元素有位置限制时怎么办?

A 特殊元素优先法: 先处理受限制的元素 (确定它的位置), 再排其余不受限元素。

Q4 捆绑法处理什么问题?

A “必须相邻”——把必须相邻的元素捆成一块, 与其余元素一起排, 再乘块内排列数。

Q5 插空法处理什么问题?

A “必须不相邻”——先排其余元素, 再把不能相邻的元素插进空隙 (含两端)。

Q6 定序问题怎么算?

A n 个不同元素全排, 其中 k 个元素顺序固定 $\rightarrow \frac{n!}{k!}$ 。

例题精讲

例题 1.3

(特殊元素优先) 6 名同学站成一排, 甲既不在最左端也不在最右端, 共有多少种站法?

例题 1.4

(捆绑法) 4 名男生、2 名女生站成一排, 要求 2 名女生相邻, 共有多少种排法?

知识点 (1) 特殊元素优先与相邻 (不相邻) 问题

A7. 7 名同学站成一排, 男生甲不能站排头, 共有多少种排法? ()

A. 720 B. 1440 C. 2160 D. 4320

A8. 6 名同学站成一排, 甲、乙两人不相邻, 共有多少种排法? ()

A. 240 B. 360 C. 480 D. 720

A9. 4 名同学站成一排合影, 共有多少种不同的站法? ()

A. 12 B. 16 C. 24 D. 48

A10. 用数字 1, 2, 3, 4 组成没有重复数字的两位数, 共有多少个? ()

A. 8 B. 12 C. 16 D. 24

B1. (2022 · 新高考 II 卷 · 第 5 题) 甲、乙、丙、丁、戊 5 名同学站成一排参加文艺汇演, 若甲不站在两端, 丙和丁相邻, 则不同的排列方式共有 ()

A. 12 种 B. 24 种 C. 36 种 D. 48 种

知识点 (2) 捆绑法与插空法的易错辨析

M2. A、B、C、D、E 5 人排队, 要求 A、B、C 三人相邻, 且 A、B 两人不相邻。小刚说: “三

人相邻就用捆绑法：把三人捆成一块，块内 $3! = 6$ 种，块与 D、E 共 3 个对象全排 $3! = 6$ 种，共 $6 \times 6 = 36$ 种。”他的做法（ ）

- A. 正确，答案 36 种
- B. 错误，答案应为 12 种
- C. 错误，答案应为 24 种
- D. 错误，答案应为 6 种

M3. 5 名同学站成一排，甲既不能站排头也不能站排尾。小华先算全排 $5! = 120$ ，再“除以 2”得 60。他的做法（ ）

- A. 正确，答案 60 种
- B. 错误，答案应为 72 种
- C. 错误，答案应为 96 种
- D. 错误，答案应为 48 种

1.3 组合

考点精要

Q1 什么是组合？

A 从 n 个不同元素中取出 m 个组成一组，不计顺序。判定标尺：换位置后结果不变 $\rightarrow$ 组合。

Q2 组合数公式是什么？

A $C_n^m = \frac{P_n^m}{m!} = \frac{n!}{m!(n-m)!}$ ，即“先排后排，再把组内顺序抹掉”。

Q3 组合数的两个重要性质是什么？

A 对称性 $C_n^m = C_n^{n-m}$ ；杨辉恒等式 $C_n^m + C_n^{m+1} = C_{n+1}^{m+1}$ 。

Q4 排列与组合如何选用？

A 问“换位置后还是不是同一种”：是 $\rightarrow$ 排列（选岗、定名次、排队）；否 $\rightarrow$ 组合（选代表、组队、摸球不看次序）。

Q5 分组分配问题什么时候要除重？

A 组间无区别（均匀分组） $\rightarrow$ 除以相同人数组的组数阶乘；组有名字（如分给甲、乙、丙） $\rightarrow$ 不除重。

例题精讲

例题 1.5

（排列与组合的辨析）（1）从 6 名同学中选 4 名分别担任四个不同的岗位，有多少种选法？（2）从 6 名同学中选 4 名组成一个小组，有多少种选法？

例题 1.6

(分组分配) (1) 6 名学生分成 3 组, 每组 2 人, 共有多少种分法? (2) 6 名学生分给甲、乙、丙三个班, 每班 2 人, 共有多少种分法?

知识点 (1) 组合的基本应用

A11. 从 8 名同学中选 3 名参加数学竞赛, 共有多少种选法? ()

A. 56 B. 112 C. 168 D. 336

A12. 从 6 名男生、4 名女生中选 3 人参加活动, 要求至少 1 名女生, 共有多少种选法? ()

A. 60 B. 80 C. 100 D. 120

A13. 正八边形共有多少条对角线? ()

A. 16 B. 20 C. 24 D. 28

A14. 从 6 名同学中选 2 名参加数学竞赛, 共有多少种选法? ()

A. 12 B. 15 C. 30 D. 36

A15. 已知 $C_n^2 = 21$, 则 n 的值是 ()

A. 6 B. 7 C. 8 D. 9

B2. (2023 · 新高考 I 卷 · 第 13 题) 某学校开设了 4 门体育类选修课和 4 门艺术类选修课, 学生需从这 8 门课中选修 2 门或 3 门课, 并且每类选修课至少选修 1 门, 则不同的选课方案共有_____种 (用数字作答)。

B3. (2023 · 全国甲卷 · 第 9 题) 现有 5 名志愿者报名参加公益活动, 在某一星期的星期六、星期日两天, 每天从这 5 人中安排 2 人参加公益活动, 则恰有 1 人在这两天都参加的不同安排方式共有 ()

A. 120 种 B. 60 种 C. 30 种 D. 20 种

B4. (2022 · 全国甲卷理 · 第 15 题) 从正方体的 8 个顶点中任选 4 个, 则这 4 个点在同一个平面的概率为_____。

知识点 (2) 排列与组合的选用辨析

M4. 从 5 名同学中选 2 名: (1) 分别担任班长和副班长; (2) 当代表参加比赛。关于选法数, 下列说法正确的是 ()

A. (1) (2) 都是 10 种

B. (1) (2) 都是 20 种

C. (1) 为 20 种, (2) 为 10 种

D. (1) 为 10 种, (2) 为 20 种

知识点 (3) 均匀分组与分配问题

M5. 4 名学生分成两组, 每组 2 人, 共有多少种分法? 小明算得 $C_4^2 \times C_2^2 = 6$ 种, 小华算得 $\frac{C_4^2 \times C_2^2}{2!} = 3$ 种。正确的是 ()

- A. 小明对, 6 种
- B. 小华对, 3 种
- C. 都不对, 应为 12 种
- D. 都不对, 应为 4 种

C2. (物流分拣) 某智能分拣仓库收到 5 件互不相同的包裹, 需按目的地分入 A、B、C 三个传送口, 要求 A 口 2 件、B 口 2 件、C 口 1 件。(1) 直接给三个传送口分配包裹, 共有多少种分法? (2) 若不考虑传送口编号, 先把 5 件包裹分成“2、2、1”三堆, 共有多少种分法? (3) 结合 (2) 的结果, 说明“先分堆 (堆无编号) 再分配”时为什么不能简单地乘以 $3!$, 并写出正确的分配总数。

1.4 二项式定理

考点精要

Q1 二项式定理是什么?

A $(a+b)^n = \sum_{r=0}^n C_n^r a^{n-r} b^r$, 共 $n+1$ 项。

Q2 通项公式是什么? 注意什么?

A $T_{r+1} = C_n^r a^{n-r} b^r$ ——下标是 $r+1$, 第 1 项对应 $r=0$, “第 5 项”对应 $r=4$ 。

Q3 二项式系数与某项的系数有什么区别?

A 二项式系数是 C_n^r , 只由 n 、 r 决定; 某项的系数是展开后该项完整的数字系数, 当 a 、 b 带常数因子时, 系数 = $C_n^r \cdot (a \text{ 的常数因子})^{n-r} \cdot (b \text{ 的常数因子})^r$ 。

Q4 二项式系数和是多少? 各项系数和怎么求?

A 二项式系数和 $C_n^0 + C_n^1 + \cdots + C_n^n = 2^n$; 各项系数和令 $x=1$ 代入 (如 $(1+2x)^n$ 令 $x=1$ 得 3^n)。

Q5 二项式系数最大项在哪?

A n 为偶数: 中间一项 $C_n^{n/2}$; n 为奇数: 中间两项 $C_n^{(n-1)/2}$ 、 $C_n^{(n+1)/2}$ 并列最大。

例题精讲

例题 1.7

(通项求指定项) 求 $\left(2x - \frac{1}{x}\right)^6$ 展开式中的常数项。

例题 1.8

(二项式系数与系数) (1) 求 $(1+x)^8$ 展开式中二项式系数最大的项; (2) 求 $(1+2x)^8$ 展开式中系数最大的项。

知识点 (1) 通项公式与指定项

A16. $(1+x)^5$ 的展开式中 x^3 的系数是 ()

A. 5 B. 10 C. 15 D. 20

A17. $(1+x)^6$ 的展开式中 x^2 的系数是 ()

A. 6 B. 10 C. 15 D. 20

A18. 若 $(1+x)^n$ 的展开式共有 10 项, 则 n 的值是 ()

A. 8 B. 9 C. 10 D. 11

B5. (2022 · 新高考 I 卷 · 第 13 题) $\left(1 - \frac{y}{x}\right)(x+y)^8$ 的展开式中 x^2y^6 的系数为____ (用数字作答)。

M6. 求 $\left(2x - \frac{1}{x}\right)^6$ 展开式中的常数项。小明说: “令 x 的指数为 0, $6 - 2r = 0$ 得 $r = 3$, 所以常数项是展开式的第 3 项 $T_3 = C_6^3 \cdot 2^3 \cdot (-1)^3 = -160$ 。” 他的说法 ()

A. 正确

B. 错误, $r = 3$ 对应第 4 项, 答案 -160 正确但 “第 3 项” 的说法错误

C. 错误, 答案应为 $+160$

D. 错误, 答案应为 -20

知识点 (2) 二项式系数与系数的区分

A19. $(1-3x)^6$ 展开式中各项系数和是 ()

A. 0 B. 64 C. 128 D. 729

M7. 求 $(2x+3y)^6$ 展开式中 x^3y^3 的系数。小红说: “ $r = 3$, 系数就是 $C_6^3 = 20$ 。” 她的答案 ()

A. 正确

B. 错误, 应为 $C_6^3 \cdot 2^3 \cdot 3^3 = 4320$

C. 错误, 应为 $C_6^3 \cdot 2^3 = 160$

D. 错误, 应为 $C_6^3 \cdot 3^3 = 540$

本章综合

S1. 用 0, 1, 2, 3, 4 五个数字组成无重复数字的五位数。(1) 共有多少个五位数? (2) 其中偶数有多少个? (3) 其中数字 1 与 2 相邻的有多少个?

S2. 已知 $(1+2x)^n$ 的展开式中, 二项式系数和为 64。(1) 求 n 的值; (2) 求展开式中各项系数和; (3) 求展开式中 x^3 的系数。

第二章 概率

2.1 随机事件与古典概型

考点精要

Q1 什么是古典概型？

A 满足两大前提的试验：(1) 样本空间中只有有限个样本点；(2) 每个样本点出现的可能性相等。两者缺一不可。

Q2 古典概型的概率公式是什么？

A $P(A) = \frac{n(A)}{n(\Omega)}$ ——分子分母都是“数格子”，用上一章的计数原理来数。

Q3 概率的基本性质有哪些？

A $0 \leq P(A) \leq 1$ ； $P(\Omega) = 1$ ， $P(\emptyset) = 0$ ； $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ ；对立事件 $P(\bar{A}) = 1 - P(A)$ 。

Q4 互斥与对立是什么关系？

A 互斥是“不能同时发生” ($A \cap B = \emptyset$)；对立是“你死我活”的最强互斥 ($\bar{A} = \Omega \setminus A$)。对立一定互斥，互斥不一定对立。

Q5 “至少一个”型问题怎么做？

A 优先转补集： $P(\text{至少一次}) = 1 - P(\text{一次都没有})$ ——正面枚举项多，补集通常只算一项。

例题精讲

例题 2.1

(等可能前提) 同时掷两枚质地均匀的骰子，求点数之和为 3 的概率。

例题 2.2

(“至少一个”转补集) 某射手每次射击命中率均为 0.6，各次射击相互独立。连射 4 次，求至少命中 1 次的概率。

知识点 (1) 古典概型与等可能前提

A1. 掷一颗骰子，求点数不小于 4 的概率。()

A. $\frac{1}{2}$ B. $\frac{1}{3}$ C. $\frac{2}{3}$ D. $\frac{1}{6}$

A2. 袋中有 2 个红球、3 个蓝球，任取 2 个球，求恰好 1 个红球的概率。()

A. $\frac{3}{5}$ B. $\frac{1}{2}$ C. $\frac{2}{5}$ D. $\frac{3}{10}$

A3. 抛 3 枚硬币，求至少 2 枚正面的概率。()

A. $\frac{1}{8}$ B. $\frac{1}{4}$ C. $\frac{1}{2}$ D. $\frac{3}{8}$

A4. 盒中有 4 个白球、3 个红球，任取 1 个球，取到红球的概率为 ()

A. $\frac{3}{7}$ B. $\frac{4}{7}$ C. $\frac{1}{2}$ D. $\frac{3}{4}$

A5. 从 1, 2, 3, 4, 5 中随机取 1 个数，取到偶数的概率为 ()

A. $\frac{1}{5}$ B. $\frac{2}{5}$ C. $\frac{3}{5}$ D. $\frac{1}{2}$

B1. (2022 · 新高考 I 卷 · 第 5 题) 从 2 至 8 的 7 个整数中随机取 2 个不同的数，则这 2 个数互质的概率为 ()

A. $\frac{1}{6}$ B. $\frac{1}{3}$ C. $\frac{1}{2}$ D. $\frac{2}{3}$

B2. (2022 · 全国乙卷 · 第 13 题) 从甲、乙等 5 名同学中随机选 3 名参加社区服务工作，则甲、乙都入选的概率为_____。

M1. 同时掷两枚质地均匀的骰子，求点数之和为 5 的概率。小明这样算：“点数之和可能是 2, 3, ..., 12，共 11 种结果，和为 5 是其中 1 种，所以概率为 $\frac{1}{11}$ 。”小明的做法 ()

A. 正确

B. 错误，应为 $\frac{1}{6}$

C. 错误，应为 $\frac{1}{9}$

D. 错误，应为 $\frac{5}{36}$

2.2 条件概率与全概率

考点精要

Q1 条件概率公式是什么？

A $P(B|A) = \frac{P(A \cap B)}{P(A)}$ ($P(A) > 0$)；古典概型下等价地 $P(B|A) = \frac{n(A \cap B)}{n(A)}$ 。

Q2 条件概率的本质是什么？

A 把样本空间从 Ω 缩小到 A ，再在 A 里数 B ——分子分母同步在 A 里数。

Q3 乘法公式是什么？

A $P(A \cap B) = P(A)P(B|A)$ ，可链条化为 $P(ABC) = P(A)P(B|A)P(C|AB)$ 。

Q4 全概率公式是什么？

A 若 $A_1, A_2, \dots, A_n$ 是完备事件组（两两互斥且并集为 Ω ），则 $P(B) = \sum_{i=1}^n P(A_i)P(B|A_i)$ ——“先按来源分类，类内相乘、类间相加”。

Q5 贝叶斯公式解决什么问题？

A “由果溯因”：已知结果 B 发生，反推它来自 A_i 的概率 $P(A_i|B) = \frac{P(A_i)P(B|A_i)}{P(B)}$ 。

例题精讲

例题 2.3

(条件概率) 袋中有 3 个红球、2 个蓝球，不放回地依次摸出 2 个球。已知第一次摸到红球，求第二次也摸到红球的概率。

例题 2.4

(全概率公式) 某产品由甲、乙两家工厂供应：甲厂供货占 70%，次品率 1%；乙厂供货占 30%，次品率 3%。现从中任取一件，求它是次品的概率。

知识点 (1) 条件概率

A6. 袋中有 2 个白球、3 个黑球共 5 个球，不放回连取 2 球。已知第一个是白球，求第二个是黑球的概率。()

- A. $\frac{3}{5}$ B. $\frac{3}{4}$ C. $\frac{1}{2}$ D. $\frac{2}{3}$

A7. 已知 $P(A) = 0.4$, $P(A \cap B) = 0.1$, 求 $P(B|A)$ 。()

- A. 0.25 B. 0.4 C. 0.5 D. 0.04

A8. 10 件产品中有 3 件次品，不放回地依次抽取 2 件。已知第一次抽到次品，则第二次也抽到次品的概率为 ()

- A. $\frac{1}{3}$ B. $\frac{2}{9}$ C. $\frac{3}{10}$ D. $\frac{2}{3}$

B3. (2023 · 全国甲卷 · 第 6 题) 某地的中学生中有 60% 的同学爱好滑冰，50% 的同学爱好滑雪，70% 的同学爱好滑冰或爱好滑雪。在该地的中学生中随机调查一位同学，若该同学爱好滑雪，则该同学也爱好滑冰的概率为 ()

- A. 0.8 B. 0.6 C. 0.5 D. 0.4

M2. 袋中有 3 个红球、2 个蓝球，不放回依次摸 2 个球，求“已知第一次摸到红球，第二次也摸到红球”的概率。小丽这样算：“第二次摸到红球的概率本来就是 $\frac{3}{5}$ ，所以答案是 $\frac{3}{5}$ 。”她的做法 ()

- A. 正确
B. 错误，应为 $\frac{1}{2}$
C. 错误，应为 $\frac{3}{10}$
D. 错误，应为 $\frac{2}{5}$

知识点 (2) 全概率与贝叶斯公式

B4. (2023 · 天津卷 · 第 13 题) 甲、乙、丙三个盒子中装有一定数量的黑球和白球，其总数之比为 5:4:6。这三个盒子中黑球占总数的比例分别为 40%，25%，50%。(1) 现从三个盒子中

各取一个球，取到的三个球都是黑球的概率为____；(2) 将三个盒子混合后任取一个球，取到白球的概率为____。

M3. 某工厂有甲、乙两台机床：甲机床加工 60% 的零件，次品率 2%；乙机床加工 40% 的零件，次品率 4%。现从中任取一件零件。小刚说：“乙机床次品率最高 (4%)，所以抽到次品的概率约为 4%。”小敏说：“应把两台机床分开算： $0.6 \times 0.02 + 0.4 \times 0.04 = 0.028$ ，即 2.8%。”正确的是 ()

- A. 小刚对
- B. 小敏对
- C. 两人都对
- D. 两人都错

C1. (核酸检测混检) 某地区开展核酸检测，已知该地区人群中某病毒携带者比例为 0.1%，检测试剂对携带者检出阳性的概率为 99%，对非携带者误报阳性的概率为 1%。(1) 随机抽取 1 人单独检测，求检测结果为阳性的概率；(2) 已知某人的检测结果为阳性，求他确实是携带者的概率。(结果保留三位小数)

2.3 事件的独立性与二项分布

考点精要

Q1 事件独立怎么判断？

A A 的发生不影响 B 的概率，即 $P(B|A) = P(B)$ ；等价判据 $P(A \cap B) = P(A)P(B)$ 。放回抽样、独立试验、互不干扰的环节通常独立。

Q2 独立与互斥有什么区别？

A 互斥是“不能同时发生” ($A \cap B = \emptyset$ ，空间层面)；独立是“概率上互不影响” ($P(AB) = P(A)P(B)$ ，度量层面)。当 $P(A), P(B) > 0$ 时，二者不可兼得。

Q3 二项分布的三条使用前提是什么？

A (1) 固定 n 次；(2) 每次只有“成功/失败”两个结果；(3) 各次独立且成功概率 p 恒定。三条缺一不可。

Q4 二项分布的概率公式是什么？

A $X \sim B(n, p)$ 时， $P(X = k) = C_n^k p^k (1 - p)^{n-k}$ ， $k = 0, 1, \dots, n$ 。

Q5 “至少一次成功”怎么算？

A $P(\text{至少一次}) = 1 - P(\text{零次}) = 1 - (1 - p)^n$ 。

例题精讲

例题 2.5

(独立与互斥的辨析) 掷一颗骰子, 设 $A =$ “点数为奇数”, $B =$ “点数为偶数”。判断 A 与 B 是否互斥? 是否独立?

例题 2.6

(二项分布) 某射手射击命中率 0.8, 各次射击相互独立。射击 5 次, 求恰好命中 3 次的概率。

知识点 (1) 事件独立性的判定

A9. 掷两枚骰子, 设 $A =$ “第一枚为偶数”, $B =$ “第二枚为奇数”。 A 与 B 的关系是 ()

A. 独立 B. 不独立 C. 互斥 D. 无法判断

B5. (2024 · 新高考 I 卷 · 第 8 题) 有 6 个相同的球, 分别标有数字 1, 2, 3, 4, 5, 6, 从中有放回地随机取两次, 每次取 1 个球。甲表示事件 “第一次取出的球的数字是 1”, 乙表示事件 “第二次取出的球的数字是 2”, 丙表示事件 “两次取出的球的数字之和是 8”, 丁表示事件 “两次取出的球的数字之和是 7”, 则 ()

A. 甲与丙相互独立
B. 甲与丁相互独立
C. 乙与丙相互独立
D. 丙与丁相互独立

B6. (2022 · 全国乙卷 · 第 10 题) 某棋手与甲、乙、丙三位棋手各比赛一盘, 各盘比赛结果相互独立。已知该棋手与甲、乙、丙比赛获胜的概率分别为 p_1, p_2, p_3 , 且 $p_3 > p_2 > p_1 > 0$ 。记该棋手连胜两盘的概率为 p , 则 ()

A. p 与该棋手和甲、乙、丙的比赛次序无关
B. 该棋手在第二盘与甲比赛, p 最大
C. 该棋手在第二盘与乙比赛, p 最大
D. 该棋手在第二盘与丙比赛, p 最大

M4. 掷一颗骰子, 设 $A =$ “点数为奇数”, $B =$ “点数为偶数”。下列说法正确的是 ()

A. A 与 B 互斥且独立
B. A 与 B 互斥但不独立
C. A 与 B 独立但不互斥
D. A 与 B 既不互斥也不独立

知识点 (2) 二项分布与 “至少一次” 问题

A10. 抛一枚骰子 5 次, 求 “至少出现一次 6 点” 的概率。()

A. $1 - \left(\frac{5}{6}\right)^5$ B. $\left(\frac{5}{6}\right)^5$ C. $\frac{1}{6}$ D. $1 - \left(\frac{1}{6}\right)^5$

A11. 某人射击命中率 0.7, 各次射击相互独立, 射击 4 次, 求恰好命中 2 次的概率。()

A. 0.2401 B. 0.2646 C. 0.28 D. 0.49

A12. 已知 $X \sim B(3, 0.5)$, 则 $P(X = 2)$ 等于 ()

A. $\frac{1}{8}$ B. $\frac{3}{8}$ C. $\frac{1}{2}$ D. $\frac{3}{4}$

A13. 抛一枚硬币 4 次, 恰好 2 次正面朝上的概率为 ()

A. $\frac{1}{8}$ B. $\frac{1}{4}$ C. $\frac{3}{8}$ D. $\frac{1}{2}$

M5. 一箱产品共 10 件, 其中 3 件次品。(1) 不放回地抽 4 件, 求恰好 2 件次品的概率; (2) 有放回地抽 4 件, 求恰好 2 件次品的概率。小张说: “两道题答案一样, 都可以用二项分布算。” 他的说法 ()

- A. 正确, 两道题答案都是 0.2646
B. 错误, (1) 应为 $\frac{3}{10}$, (2) 才是 0.2646
C. 错误, (1) 应为 $\frac{1}{3}$, (2) 才是 0.2646
D. 错误, 两道题都不该用二项分布

M6. 某人射击命中率 0.7, 各次射击相互独立, 射击 4 次。小红直接枚举 “中 1 次、2 次、3 次、4 次” 四种情况求 “至少命中 1 次” 的概率。下列说法正确的是 ()

- A. 枚举繁琐但结果正确, 概率为 0.9919
B. 枚举繁琐且易漏项, 应改用补集 $1 - 0.3^4$, 概率也为 0.9919
C. 枚举是最简做法, 正确
D. 补集应算 0.3^4 , 答案为 0.0081

C2. (疫苗有效率) 某传染病疫苗的三期临床数据显示: 未接种人群感染率为 4%, 接种人群感染率为 0.8%。某社区随机选取 100 名未接种居民作为对照组, 设各人感染与否相互独立。(1) 设 X 为对照组中感染人数, 写出 “恰有 2 人感染” 的概率表达式 (用组合数与幂表示); (2) 求对照组中 “至少 1 人感染” 的概率 (结果保留两位小数); (3) 疫苗有效率常定义为 $\frac{p_{\text{未}} - p_{\text{接}}}{p_{\text{未}}} \times 100\%$, 求该疫苗的有效率。

2.4 数字特征与正态分布初步

考点精要

Q1 期望是什么?

A 分布的 “重心”: $E(X) = x_1p_1 + x_2p_2 + \cdots + x_np_n$, 是取值按概率加权的平均。

Q2 方差是什么?

A 分布的 “离散度”: $D(X) = \sum (x_i - E(X))^2 p_i$, 是关于重心的平均偏离平方; $\sqrt{D(X)}$ 叫标准差。

Q3 二项分布的期望和方差是什么?

A $X \sim B(n, p)$ 时, $E(X) = np$, $D(X) = np(1 - p)$ ——期望不带 $(1 - p)$, 方差带 $(1 - p)$ 。

Q4 正态分布怎么写？有什么对称性？

A $X \sim N(\mu, \sigma^2)$ ，曲线关于 $x = \mu$ 对称，在 $x = \mu$ 处最高； μ 是均值， σ 是标准差。

Q5 3σ 原则是什么？

A $P(\mu - \sigma < X < \mu + \sigma) \approx 0.6827$ ； $P(\mu - 2\sigma < X < \mu + 2\sigma) \approx 0.9545$ ； $P(\mu - 3\sigma < X < \mu + 3\sigma) \approx 0.9973$ 。

例题精讲

例题 2.7

(二项分布的期望与方差) 某射手命中率 0.2，各次射击相互独立，射击 10 次，设 X 为命中次数。求 $E(X)$ 与 $D(X)$ 。

例题 2.8

(正态分布的对称性与 3σ 原则) 某厂生产的一批零件直径 (单位: mm) 服从正态分布 $N(10, 0.04)$ 。(1) 求直径落在 $[9.8, 10.2]$ 内的概率近似值；(2) 求直径小于 10 mm 的概率。

知识点 (1) 期望与方差

A14. 设 $X \sim B(10, 0.2)$ ，则 $E(X)$ 与 $D(X)$ 分别是 ()

- A. $E(X) = 2$, $D(X) = 1.6$ B. $E(X) = 2$, $D(X) = 2$
C. $E(X) = 1.6$, $D(X) = 2$ D. $E(X) = 2$, $D(X) = 0.4$

A15. 设随机变量 X 的分布列为 $P(X = 0) = 0.2$, $P(X = 1) = 0.5$, $P(X = 2) = 0.3$ ，则 $E(X)$ 等于 ()

- A. 0.8 B. 1.0 C. 1.1 D. 1.2

B7. (2022 · 全国甲卷 · 第 19 题) 甲、乙两个学校进行体育比赛，比赛共设三个项目，每个项目胜方得 10 分，负方得 0 分，没有平局。三个项目比赛结束后，总得分高的学校获得冠军。已知甲学校在三个项目中获胜的概率分别为 0.5, 0.4, 0.8，各项目的比赛结果相互独立。(1) 求甲学校获得冠军的概率；(2) 用 X 表示乙学校的总得分，求 X 的分布列与期望。

M7. 设 $X \sim B(10, 0.2)$ ，小丽算得 $E(X) = 2$, $D(X) = 2$ 。她的方差结果 ()

- A. 正确
B. 错误，应为 1.6
C. 错误，应为 0.4
D. 错误，应为 3.2

知识点 (2) 正态分布与 3σ 原则

A16. 某班学生身高 $X \sim N(170, 25)$ ($\sigma = 5$)。身高在 $[165, 175]$ 内的概率近似值为 ()

- A. 0.6827 B. 0.9545 C. 0.5 D. 0.9973

B8. (2022 · 新高考 II 卷 · 第 13 题) 已知随机变量 X 服从正态分布 $N(2, \sigma^2)$ ，且 $P(2 < X \leq$

$2.5) = 0.14$, 则 $P(X > 2.5) = \underline{\hspace{2cm}}$ 。

本章综合

S1. 甲、乙两人独立地射击同一目标, 命中率分别为 0.8、0.7。(1) 求两人都命中的概率; (2) 求至少一人命中的概率; (3) 求恰好一人命中的概率。

S2. 某校从 6 名男生、4 名女生中选派 4 人参加活动。(1) 共有多少种选派方案? (2) 求“恰好 2 男 2 女”的概率; (3) 设 X 为选派的 4 人中女生的人数, 求 X 的分布列和数学期望。

第三章 统计

3.1 抽样方法

考点精要

Q1 普查与抽样调查怎么选？

A 普查结果准但费时费力还可能破坏对象（如检测灯泡寿命）；抽样调查省时省力但可能偏。关键判断：查了会不会坏、查完用不用得上、查全代价高不高——触发任何一个，就偏向抽样。

Q2 简单随机抽样的定义与判断标准？

A 从含有 N 个个体的总体中逐个不放回地抽取 n 个，每次抽取时每个个体被抽到的概率都相等。判断标准一条：是不是“每个个体机会均等 + 随机决定选谁”。凡有“规律、习惯、图方便”掺进去的都不是。

Q3 分层抽样的人数公式？

A 总体 N 个，某层 N_i 个，样本量 n ，则该层抽取 $n_i = \frac{N_i}{N} \times n$ 。比例是**人数比**，不是层数比。

Q4 系统抽样的间隔与首个体？

A 间隔 $k = \left\lfloor \frac{N}{n} \right\rfloor$ ，先随机确定第一个个体，再按间隔依次抽取。总体存在周期性规律时，起点必须随机，否则整段样本被带偏。

Q5 抽样里“多”和“准”是什么关系？

A 样本量再大，没有代表性也是白搭。统计的第一原则不是“多”，而是“代表”。

例题精讲

例题 3.1

（普查还是抽样）下列调查中，应采用普查还是抽样调查？为什么？（1）检验某厂生产的一批乒乓球（共 5000 个）是否合格（需逐个检测弹性）；（2）统计全校学生本学期获得的校级表彰名单；（3）检测一批疫苗的保存有效期（需开瓶检测）。

例题 3.2

(分层抽样求人数与验证) 某中学高一、高二、高三年级人数分别为 800、600、600。采用分层抽样抽取样本量为 100 的样本。(1) 三个年级各应抽取多少人?(2) 验证各层抽取比例与总体比例一致。

知识点 (1) 普查与抽样调查

A1. 下列调查中, 适合用普查的是 ()

- A. 检测一批灯泡的使用寿命 B. 调查全国中学生的视力状况
C. 统计本班同学的身高 D. 检验一批种子的发芽率

A2. 抽样调查与普查相比, 最主要的优点是 ()

- A. 结果更准确 B. 省时省力, 适用于破坏性检测或大范围调查
C. 一定能反映总体情况 D. 不需要设计抽样方案

知识点 (2) 简单随机抽样与系统抽样

A3. 下列抽样方式中, 属于简单随机抽样的是 ()

- A. 按学号每隔 10 名抽 1 名 B. 把 1000 名学生编号后随机抽取 100 个号码
C. 按年级分层各抽 30 人 D. 在校门口拦前 100 名学生

A4. 某校 2000 名学生编号为 1 至 2000, 用系统抽样抽取 100 名。抽样间隔 k 为 (), 若第一段抽中 5 号, 则样本中的最后一个编号为 ()

- A. 20, 1985 B. 10, 1995 C. 20, 1995 D. 10, 1985

知识点 (3) 分层抽样

A5. 某厂生产零件 600 个, 其中甲车间 240 个、乙车间 360 个。用分层抽样抽 60 个检验, 甲、乙车间各应抽 ()

- A. 24, 36 B. 30, 30 C. 20, 40 D. 36, 24

M1. 某校 1200 名学生, 高一、高二、高三人数之比为 5:3:2, 采用分层抽样抽取样本量为 120 的样本。下列说法正确的是 ()

- A. 三个年级各抽 40 人 B. 高一抽 60 人、高二抽 36 人、高三抽 24 人
C. 高一抽 50 人、高二抽 30 人、高三抽 20 人 D. 高三抽 40 人

M2. 三个年级人数分别为 300、200、100 (共 600), 抽取样本量 60。小明说“三层都抽 20 人才公平”。这种说法对吗? 正确应各抽多少人? 请说明理由。

C1. (手机使用时长分层抽样调查) 某校共有学生 2400 人, 其中高一 960 人、高二 720 人、高三 720 人。学校欲调查全校学生平均每天使用手机的时间。(1) 若采用分层抽样抽取样本量为 120 的样本, 三个年级各应抽取多少人?(2) 若改按住宿情况分层: 住宿生 1500 人、走读生 900 人, 样本量仍为 120, 各应抽取多少人?(3) 有同学建议“在校门口随机拦截 120 人”, 请指出该方案的不足, 并说明为什么分层抽样更合理。

3.2 频率分布直方图与统计图表

考点精要

Q1 直方图的纵轴是什么？频率怎么读？

A 纵轴是 $\frac{\text{频率}}{\text{组距}}$ ，不是频率本身。小长方形面积 = 组距 $\times$ 纵轴高度 = 频率。所有小长方形面积之和等于 1。

Q2 作直方图的步骤？

A 求极差 $\rightarrow$ 定组距与组数 $\rightarrow$ 分组（左闭右开） $\rightarrow$ 列表（频数/频率） $\rightarrow$ 画图。

Q3 由直方图怎么估中位数、平均数、众数区间？

A 中位数：总面积平分线所在位置；平均数： $\bar{x} \approx \sum(\text{组中值} \times \text{频率})$ ；众数区间：频率（面积）最大的区间，即最高的矩形。

Q4 平均数与中位数在偏态数据里的位置关系？

A 右偏（长尾在右）时 $\bar{x} > M_e$ ；左偏时相反。平均数往长尾方向偏，中位数稳稳站在面积平分线上。

例题精讲

例题 3.3

（读直方图求频率）某班 40 名同学的体重数据分组（单位：kg，组距 5）及频数为：[40, 45)：4，[45, 50)：8，[50, 55)：12，[55, 60)：10，[60, 65)：6。（1）求各组频率；（2）求体重不低于 55 kg 的频率与人数；（3）验证“面积和 = 1”。

例题 3.4

（由直方图估计统计量）利用例题 3.3 的 40 人体重数据：（1）估计中位数；（2）估计平均数；（3）指出众数所在区间。

知识点 (1) 直方图的读图与频率

A6. 某样本 50 人，分组 [0, 10)、[10, 20)、[20, 30)、[30, 40) 的频数分别为 5、15、20、10，则第二组的频率为（ ）

A. 0.1 B. 0.3 C. 0.4 D. 0.2

A7. 频率分布直方图中，纵轴表示（ ）

A. 频数 B. 频率 C. $\frac{\text{频率}}{\text{组距}}$ D. 组距

A8. 一个直方图共 4 组，前三组的频率分别为 0.15、0.25、0.35，则第 4 组的频率为（ ）

A. 0.35 B. 0.25 C. 0.2 D. 0.4

M3. 某直方图组距为 10，第一个矩形高 0.02，第二个矩形高 0.04。下列说法正确的是（ ）

A. 第二组的频率是 0.04 B. 第一组的频率是 0.02

C. 第二组的频率是第一组的 2 倍 D. 两组的频率之和是 0.06

M4. 有两个直方图：第一个组距为 1，某矩形高 0.3；第二个组距为 4，某矩形高 0.1。小华说“第一个的矩形高，所以频率大”。对吗？请通过计算说明理由。

知识点 (2) 由直方图估计统计量

A9. 频率分布直方图中，若组距为 2，某个矩形的面积为 0.16，则该组的频率为 ()

A. 0.08 B. 0.16 C. 0.32 D. 0.64

A10. 已知直方图各组的频率为 0.1、0.2、0.4、0.3，对应组中值分别为 10、30、50、70，则样本平均数约为 ()

A. 40 B. 42 C. 46 D. 50

B1. (2024 · 新课标 II 卷 · 第 4 题) 某农业研究部门在面积相等的 100 块稻田上种植一种新型水稻，得到各块稻田的亩产量 (单位: kg) 并部分整理如下表:

亩产量	[900, 950)	[950, 1000)	[1000, 1050)	[1050, 1100)	[1100, 1150)	[1150, 1200)
频数	6	12	18		24	10

据表中数据，结论中正确的是 ()

- A. 100 块稻田亩产量的中位数小于 1050 kg
B. 100 块稻田中亩产量低于 1100 kg 的稻田所占比例超过 80%
C. 100 块稻田亩产量的极差介于 200 kg 至 300 kg 之间
D. 100 块稻田亩产量的平均值介于 900 kg 至 1000 kg 之间

B2. (2022 · 新高考 II 卷 · 第 19 题) **【数据表文字化】** 在某地区进行流行病学调查，随机调查了 100 位某种疾病患者的年龄，得到频率分布直方图。直方图分组 (组距 10，左闭右开) 及纵轴高度 ($\frac{\text{频率}}{\text{组距}}$) 如下 **【数据表文字化】**:

区间	[0, 10)	[10, 20)	[20, 30)	[30, 40)	[40, 50)	[50, 60)	[60, 70)	[70, 80)	[80, 90)
高度	0.001	0.002	0.012	0.017	0.023	0.020	0.017	0.006	0.002

(1) 估计该地区这种疾病患者的平均年龄 (同一组中的数据用该组区间的中点值为代表); (2) 估计该地区一位这种疾病患者的年龄位于区间 [20, 70) 的概率; (3) 已知该地区这种疾病的患病率为 0.1%，该地区年龄位于区间 [40, 50) 的人口占该地区总人口的 16%。从该地区任选一人，若此人的年龄位于区间 [40, 50)，求此人患这种疾病的概率 (以样本数据中患者的年龄位于各区间的频率作为患者的年龄位于该区间的概率，精确到 0.0001)。

C2. (空气质量数据) 某市环保监测站记录了 30 天的 PM_{2.5} 日均浓度 (单位: $\mu\text{g}/\text{m}^3$)，分组及频数如下 (组距 25):

浓度	[25, 50)	[50, 75)	[75, 100)	[100, 125)	[125, 150)
天数	3	6	9	7	5

(1) 求各组频率并验证面积和等于 1; (2) 估计这 30 天日均浓度的中位数; (3) 若规定日均浓度不低于 $100 \mu\text{g}/\text{m}^3$ 视为“轻度污染以上”, 估计这 30 天中有多少天处于该水平。

3.3 百分位数与四分位数

考点精要

Q1 第 p 百分位数的定义?

A 一组数据中从小到大排序后, 小于或等于它的数据约占 $p\%$, 它就是第 p 百分位数。它用于排名比较, 与分数本身数值无关。

Q2 计算第 p 百分位数的算法?

A 排序 $\rightarrow$ 算 $i = \frac{p}{100} \times n \rightarrow$ 若 i 不是整数, 取大于 i 的最小整数 j , 第 j 个数据即所求; 若 i 是整数, 取第 i 个与第 $i+1$ 个数据的平均数。

Q3 四分位数与箱线图?

A 第 25 百分位数 = 下四分位数 Q_1 , 第 50 百分位数 = 中位数, 第 75 百分位数 = 上四分位数 Q_3 。五数概括: 最小值、 Q_1 、中位数、 Q_3 、最大值, 画成图即箱线图。

Q4 百分位数最常犯的错?

A 不排序直接数; 以及 i 恰为整数时只取第 i 个而不取平均数。

例题精讲

例题 3.5

(求百分位数) 某班 16 名同学一次数学测验的成绩如下 (未排序):

65, 78, 90, 58, 72, 84, 66, 95, 51, 80, 69, 88, 74, 61, 77, 83

求第 25 百分位数与第 60 百分位数。

例题 3.6

(求五数概括) 利用例题 3.5 的 16 个成绩数据, 求最小值、 Q_1 、中位数、 Q_3 、最大值, 写出五数概括, 并判断 95 分是否像是离群点。

知识点 (1) 百分位数

A11. 已排序的 10 个数据为 10, 12, 15, 18, 20, 21, 23, 25, 26, 28, 则第 80 百分位数为 ()

A. 25 B. 26 C. 25.5 D. 24

A12. 已排序的 10 个数据为 3, 5, 7, 8, 10, 12, 15, 18, 20, 24, 则下四分位数 Q_1 为 ()

A. 6 B. 7 C. 7.5 D. 8

A13. 一组数据共有 15 个 (已排序), 其第 70 百分位数对应的位置是 ()

A. 第 10 个数据 B. 第 11 个数据 C. 第 10 与第 11 个数据的平均数 D. 第 7 个数据

M5. 未排序数据为 5, 2, 8, 9, 3, 7, 1, 6, 4, 10。小丽不排序, 直接数第 7 个数据 (6) 作为第 70 百分位数。她做得对吗? 正确结果是多少?

M6. 已排序数据 12, 15, 18, 20, 25 ($n = 5$)。某同学算第 80 百分位数: $i = 0.8 \times 5 = 4$, 于是取第 4 个数据 20。他的做法对吗? 正确结果是多少?

知识点 (2) 四分位数与箱线图

A14. 一组数据 (已排序) 为 2, 5, 8, 10, 14, 18, 20, 则下四分位数 Q_1 为 ()

A. 5 B. 8 C. 6.5 D. 7

A15. 一组数据的五数概括为 (最小值 = 12, $Q_1 = 20$, 中位数 = 24, $Q_3 = 30$, 最大值 = 40), 则四分位距 $Q_3 - Q_1$ 为 ()

A. 10 B. 12 C. 16 D. 18

3.4 集中趋势与离散程度

考点精要

Q1 平均数、中位数、众数各自的适用场景?

A 谈“大多数人的水平”用中位数 (工资、房价); 谈“总水平”用平均数 (总分、总量); 谈“最常见”用众数 (进货尺码)。平均数对极端值敏感, 中位数、众数抗极端值。

Q2 方差与标准差?

A 方差 $s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$, 是“离均平方的平均”; 标准差 $s = \sqrt{s^2}$ 与数据同单位。方差越大数据越分散、越不稳定。简算公式: $s^2 = \frac{1}{n} \sum x_i^2 - \bar{x}^2$ 。

Q3 方差的两条性质?

A 所有数据加同一个常数 c , 方差不变; 所有数据乘常数 k , 方差变 k^2 倍, 标准差变 $|k|$ 倍。

Q4 方差除以 n 还是 $n - 1$?

A 描述“这组数据本身”用 $\frac{1}{n}$; 若把这组数据当作样本估计总体, 样本方差常用 $\frac{1}{n-1}$ (无偏修正)。两种口径对“离散程度大小”的判断结论一致, 但用途不同不能混用。

Q5 比较两组数据的稳定性能只看平均分吗?

A 不能。平均数相同看方差: 方差小 = 稳定。“平均 + 方差”一起看才完整。

例题精讲

例题 3.7

(极端值陷阱) 某公司 11 名员工, 10 人月薪 4000 元, 1 人月薪 48000 元。(1) 求平均月薪与月薪中位数; (2) 招聘广告若写“平均月薪 8000 元”是否有误导性? 描述大多数员工收入该用哪个指标?

例题 3.8

(计算方差与标准差) 甲、乙两名运动员 6 次射击成绩 (环) : 甲: 6, 7, 7, 8, 8, 9 乙: 3, 6, 7, 8, 9, 12。(1) 求两人的平均成绩; (2) 求两人的方差并判断谁更稳定。

知识点 (1) 平均数、中位数与众数

A16. 一组数据为 2, 3, 4, 5, 6, 加入一个极端值 100 后, 下列说法正确的是 ()

- A. 中位数明显变大 B. 平均数明显变大, 中位数基本不变
C. 众数明显变大 D. 中位数与平均数变化程度相同

A17. 某班 20 名同学身高 (cm) 已排序, 第 10 个为 165、第 11 个为 167, 则这组数据的中位数为 ()

- A. 165 B. 166 C. 167 D. 无法确定

B5. (2022 · 全国甲卷 (理) · 第 2 题) 【文字化还原】某社区通过公益讲座以普及社区居民的垃圾分类知识。为了解讲座效果, 随机抽取 10 位社区居民, 让他们在讲座前和讲座后各回答一份垃圾分类知识问卷, 这 10 位社区居民在讲座前和讲座后问卷答题的正确率 (原题散点图) 还原为如下数据表【文字化还原】:

居民编号	1	2	3	4	5	6	7	8	9	10
讲座前正确率	65%	65%	70%	65%	65%	75%	90%	90%	80%	95%
讲座后正确率	90%	85%	80%	90%	85%	85%	95%	100%	85%	100%

则 ()

- A. 讲座前问卷答题的正确率的中位数小于 70%
B. 讲座后问卷答题的正确率的平均数大于 85%
C. 讲座前问卷答题的正确率的标准差小于讲座后正确率的标准差
D. 讲座后问卷答题的正确率的极差大于讲座前正确率的极差

M7. 某市发布 “平均月薪 9000 元”, 而月薪中位数只有 4500 元。下列说法正确的是 ()

- A. 平均月薪被少数高收入者拉高, 中位数更能代表大多数人的收入水平
B. 平均月薪远大于中位数, 说明收入分布右偏 (长尾在高收入端)
C. 平均月薪与中位数都可以直接作为每个人的收入水平
D. 平均数越大, 说明数据质量越高

知识点 (2) 方差与标准差

A18. 数据 2, 4, 6, 8, 10 的方差为 ()

- A. 8 B. 40 C. $2\sqrt{2}$ D. 6

A19. 若一组数据的方差为 9, 则其标准差为 ()

- A. 81 B. 3 C. 9 D. 4.5

A20. 甲班平均分 80、方差 16; 乙班平均分 78、方差 4。下列说法正确的是 ()

- A. 甲班成绩整体水平更高且更稳定
- B. 乙班成绩整体水平更高且更稳定
- C. 甲班平均分更高，乙班成绩更稳定
- D. 两班稳定性相同

A21. 若一组数据的方差为 4，把每个数据都加 3，则新数据的方差为 ()

- A. 4 B. 7 C. 12 D. 19

B3. (2023 · 新课标 I 卷 · 第 9 题) 有一组样本数据 $x_1, x_2, \dots, x_6$ ，其中 x_1 是最小值， x_6 是最大值，则 ()

- A. x_2, x_3, x_4, x_5 的平均数等于 $x_1, x_2, \dots, x_6$ 的平均数
- B. x_2, x_3, x_4, x_5 的中位数等于 $x_1, x_2, \dots, x_6$ 的中位数
- C. x_2, x_3, x_4, x_5 的标准差不小于 $x_1, x_2, \dots, x_6$ 的标准差
- D. x_2, x_3, x_4, x_5 的极差不大于 $x_1, x_2, \dots, x_6$ 的极差

B4. (2023 · 全国乙卷 (理) · 第 17 题) 某厂为比较甲、乙两种工艺对橡胶产品伸缩率的处理效应，进行 10 次配对试验，每次试验选用材质相同的两个橡胶产品，随机地选其中一个用甲工艺处理，另一个用乙工艺处理，测量处理后的橡胶产品的伸缩率。甲、乙两种工艺处理后的橡胶产品的伸缩率分别记为 x_i, y_i ($i = 1, 2, \dots, 10$)。试验结果如下：

试验序号	1	2	3	4	5	6	7	8	9	10
伸缩率 x_i	545	533	551	522	575	544	541	568	596	548
伸缩率 y_i	536	527	543	530	560	533	522	550	576	536

记 $z_i = x_i - y_i$ ($i = 1, 2, \dots, 10$)，记 $z_1, \dots, z_{10}$ 的样本平均数为 $\bar{z}$ ，样本方差为 s^2 。(1) 求 $\bar{z}, s^2$ ；(2) 判断甲工艺处理后的橡胶产品的伸缩率较乙工艺处理后的橡胶产品的伸缩率是否有显著提高。(如果 $\bar{z} \geq 2\sqrt{\frac{s^2}{10}}$ ，则认为甲工艺处理后的橡胶产品的伸缩率较乙工艺处理后的橡胶产品的伸缩率有显著提高，否则不认为有显著提高。)

M8. 数据 1, 2, 3, 4, 5 的平均数为 3。下列说法正确的是 ()

- A. 这组数据的方差为 2 (按 $\frac{1}{n}$ 计算)
- B. 若把这组数据当作样本估计总体，样本方差按 $\frac{1}{n-1}$ 计算为 2.5
- C. 高中阶段“描述这组数据”统一按 $\frac{1}{n}$ ，答案是 2
- D. 同一道题里 n 与 $n-1$ 可以随意换用

本章综合

S1. 某校高一年级共有 600 名学生，其中男生 360 人、女生 240 人。为调查学生平均每日睡眠时间 (单位：小时)，采用分层抽样抽取 60 名学生，测得数据经整理得到频率分布直方图，分组 (组距 1) 及纵轴高度如下 【数据表文字化】：

区间	[4, 5)	[5, 6)	[6, 7)	[7, 8)	[8, 9)	[9, 10)
高度	0.02	0.10	0.28	0.34	0.20	0.06

(1) 男生、女生各应抽取多少人? (2) 求样本中睡眠时间低于 6 小时的学生所占比例; (3) 估计睡眠时间的中位数; (4) 若规定睡眠时间不少于 8 小时为“睡眠充足”, 估计该校高一学生中睡眠充足的人数。

S2. 甲、乙两名同学 5 次数学周测成绩 (分): 甲: 85, 90, 92, 88, 95 乙: 80, 85, 100, 78, 97。(1) 求两人的平均成绩; (2) 求两人的方差, 判断谁的成绩更稳定; (3) 若老师要表彰“成绩整体水平高且稳定”的同学, 应推荐谁? 说明理由。

第四章 成对数据的统计分析

4.1 相关关系

考点精要

Q1 函数关系、相关关系、无关系怎么区分？

A 函数关系：一个变量完全确定另一个变量（电压一定时， $I = \frac{U}{R}$ ，给定电阻电流确定）；相关关系：一个变量变化时另一个有变化趋势，但同一个 x 对应多个 y （身高与体重）；无关系：看不出任何相互影响。

Q2 “相关”等于“因果”吗？

A 不等于。相关只说明同向/反向变化，不能说明谁导致谁——冰淇淋销量与溺水人数同涨，背后是共同原因“夏天”。

Q3 散点图读哪四件事？

A 方向（右上 → 正相关，右下 → 负相关）、形状（线性/曲线/无关）、强度（点离直线近不近）、离群点（偏离整体的怪点）。

Q4 相关系数 r 怎么读？

A $-1 \leq r \leq 1$ ；符号决定方向（ $r > 0$ 正、 $r < 0$ 负）， $|r|$ 决定强度（越接近 1 线性越强）。经验约定： $|r| \geq 0.75$ 强， $0.3 \leq |r| < 0.75$ 中等， $|r| < 0.3$ 弱。 $r \approx 0$ 只说明无**线性**相关，仍可能有曲线相关。

Q5 任何回归计算前先做什么？

A 先画散点图。数据是弯的还硬套直线， r 再漂亮也白搭。

例题精讲

例题 4.1

（判断关系类型）判断下列每组变量之间是函数关系、相关关系还是无关系：（1）电阻一定时，电路中的电流与电压；（2）某班同学的身高与体重；（3）某市冬季的供暖费与当月平均气温；（4）某天的降雨量与当天的彩票开奖号码。

例题 4.2

(读散点图与相关因果辨析) 某研究统计了 12 名高三学生“每周课外锻炼时长 x (小时)”与“体质测试得分 y (分)”，散点图显示点云从左下角斜向右上角且较为密集。(1) 判断相关方向与大致强度；(2) 能否据此得出“锻炼能提高体质测试成绩”的结论？说明理由。

知识点 (1) 相关关系与函数关系

A1. 下列各组变量中，属于相关关系的是 ()

- A. 圆的半径与面积 B. 电压一定时，电路中的电流与电阻
C. 某班同学的身高与体重 D. 某天天气与彩票开奖号码

A2. 下列各组变量中，属于函数关系的是 ()

- A. 某班同学的身高与体重 B. 圆的半径与面积
C. 气温与售出的羽绒服数量 D. 降雨量与出门人数

知识点 (2) 散点图与“相关不等于因果”

A3. 某散点图的点云从左下角斜向右上角且较为密集，则 ()

- A. 正相关，线性相关较强 B. 正相关，线性相关较弱
C. 负相关，线性相关较强 D. 无相关

A4. 某散点图的点云从左上角斜向右下角且较分散，则 ()

- A. 正相关，线性相关较强 B. 负相关，线性相关较弱
C. 负相关，线性相关较强 D. 无相关

M1. 某研究显示“阅读量高的学生数学成绩也高”，相关系数 $r = 0.88$ 。下列结论正确的是 ()

- A. 提高阅读量必然导致数学成绩提高
B. 阅读量与数学成绩存在较强的正相关，但不能据此断言因果
C. 数学成绩高导致阅读量高
D. 两者没有关系

C1. (冰淇淋与溺水的“假因果”) 某市统计了最近 10 年的数据：人均冰淇淋消费量 x (kg/年) 与当年溺水人数 y (人)，散点图显示点云从左上角斜向右下角，样本相关系数 $r \approx -0.82$ 。(1) 指出相关方向与强度；(2) 某媒体据此报道“少吃冰淇淋可以降低溺水风险”。请从统计角度分析该报道是否合理；(3) 若要研究“冰淇淋消费”与“溺水”之间是否存在因果关系，应当如何设计研究？

知识点 (3) 相关系数

A5. 甲、乙两组数据的相关系数分别为 $r_1 = 0.6$ 、 $r_2 = -0.8$ ，则 ()

- A. 甲组线性相关更强 B. 乙组线性相关更强
C. 两组线性相关强度相同 D. 无法比较强度

A6. 甲、乙两组数据的样本相关系数分别为 $r_1 = 0.9$ 、 $r_2 = -0.95$ ，则 ()

- A. 甲组线性相关更强 B. 乙组线性相关更强
C. 两组线性相关强度相同 D. 无法比较强度

A7. 若两组数据的样本相关系数 $r \approx 0$ ，下列说法正确的是 ()

- A. 两个变量一定无关 B. 两个变量没有线性相关关系，但可能存在曲线相关
C. 计算一定有误 D. 两个变量负相关

M2. 一组数据满足 $y = (x - 3)^2$ ($x = 1, 2, 3, 4, 5$)，散点图为 U 形，计算得线性相关系数 $r \approx 0$ 。下列说法正确的是 ()

- A. 两个变量没有关系 B. 两个变量没有线性相关关系，但可能存在曲线相关
C. 相关系数计算有误 D. $r = 0$ 说明完全无关

B4. (2022 · 全国乙卷 (理) · 第 19 题) 某地经过多年的环境治理，已将荒山改造成了绿水青山。为估计一林区某种树木的总材积量，随机选取了 10 棵这种树木，测量每棵树的根部横截面积 (单位: m^2) 和材积量 (单位: m^3)，得到如下数据:

样本号 i	1	2	3	4	5	6	7	8	9	10	总和
根部横截面积 x_i	0.04	0.06	0.04	0.08	0.08	0.05	0.05	0.07	0.07	0.06	0.6
材积量 y_i	0.25	0.40	0.22	0.54	0.51	0.34	0.36	0.46	0.42	0.40	3.9

并计算得 $\sum_{i=1}^{10} x_i^2 = 0.038$, $\sum_{i=1}^{10} y_i^2 = 1.6158$, $\sum_{i=1}^{10} x_i y_i = 0.2474$ 。(1) 估计该林区这种树木平均一棵的根部横截面积与平均一棵的材积量; (2) 求该林区这种树木的根部横截面积与材积量的样本相关系数 (精确到 0.01); (3) 现测量了该林区所有这种树木的根部横截面积, 并得到所有这种树木的根部横截面积总和为 186 m^2 。已知树木的材积量与其根部横截面积近似成正比。利用以上数据给出该林区这种树木的总材积量的估计值。

$$\text{附: 样本相关系数 } r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot \sum_{i=1}^n (y_i - \bar{y})^2}}.$$

4.2 一元线性回归模型

考点精要

Q1 回归直线是什么?

A $\hat{y} = bx + a$, 其中 $\hat{y}$ 是预测值。最小二乘法选 a, b 使残差平方和 $\sum (y_i - bx_i - a)^2$ 最小, 此时

$$b = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}, \quad a = \bar{y} - b\bar{x}$$

Q2 回归直线必过哪个点?

A 样本中心 $(\bar{x}, \bar{y})$ 。初中“正比例函数过原点”的印象在这里不适用, 回归直线一般不过原点, 也不一定过某个数据点。

Q3 残差是什么?

A $e_i = y_i - \hat{y}_i$, 是观测值减预测值的**纵向距离**, 不是“错误”。点在线上方残差为正, 下方为负; $|e_i|$ 越大, 该点拟合越差。

Q4 内插与外推?

A 回归直线只在样本 x 的取值范围内预测 (内插) 才可靠; 取到范围之外预测 (外推) 没有数据支撑, 不可靠。

Q5 r 与 b 的符号关系?

A r 与 b 同号: $r < 0$ 则 $b < 0$, 回归直线从左到右下降。

例题精讲

例题 4.3

(求回归方程并预测) 某实验室记录了某化学反应中温度 x (单位: $^{\circ}\text{C}$) 与反应时间 y (分钟) 的 5 组数据:

x	20	25	30	35	40
y	60	54	50	45	41

(1) 求回归方程 $\hat{y} = bx + a$; (2) 预测温度为 32°C 时的反应时间。

例题 4.4

(残差与外推) 利用例题 4.3 的回归方程 $\hat{y} = -0.94x + 78.2$: (1) 求 5 个样本点的残差; (2) 哪个点拟合最差? (3) 用该方程预测 $x = 60$ 时的反应时间, 合理吗?

知识点 (1) 回归直线与最小二乘法

A8. 回归直线必过点 ()

A. 原点 B. 第一个样本点 C. 样本中心 $(\bar{x}, \bar{y})$ D. 所有样本点

A9. 已知回归直线过样本中心 $(4, 9)$, 斜率 $b = 1.5$, 则回归方程为 ()

A. $\hat{y} = 1.5x + 3$ B. $\hat{y} = 1.5x + 9$ C. $\hat{y} = 1.5x - 3$ D. $\hat{y} = 1.5x + 15$

A10. 已知回归直线过样本中心 $(6, 20)$, 且斜率 $b = 2$, 则回归方程为 ()

A. $\hat{y} = 2x + 8$ B. $\hat{y} = 2x + 20$ C. $\hat{y} = 2x - 8$ D. $\hat{y} = 2x + 12$

M3. 关于回归直线 $\hat{y} = bx + a$, 下列说法正确的是 ()

A. 回归直线必过原点 B. 回归直线必过某一个样本点

C. 回归直线必过样本中心 $(\bar{x}, \bar{y})$ D. 回归直线必过所有样本点

B1. (2023 · 天津卷 · 第 7 题) 鸢是鹰科的一种鸟, 鸢尾花因花瓣形如鸢尾而得名。通过随机抽样, 收集了若干朵某品种鸢尾花的花萼长度和花瓣长度 (单位: cm), 绘制散点图, 计算得样本相关系数为 0.8642, 利用最小二乘法求得相应的经验回归方程为 $\hat{y} = 0.7501x + 0.6105$ (其中 x 为花萼长度, y 为花瓣长度)。根据以上信息, 如下判断正确的为 ()

- A. 花萼长度和花瓣长度不存在相关关系
- B. 花萼长度和花瓣长度负相关
- C. 花萼长度为 7 cm 的该品种鸢尾花的花瓣长度的平均值约为 5.8612 cm
- D. 若选取其他品种鸢尾花进行抽样, 所得花萼长度与花瓣长度的样本相关系数一定为 0.8642

知识点 (2) 残差与预测 (内插与外推)

A11. 样本数据中 $x \in [1, 6]$, 回归方程 $\hat{y} = 2x + 1$ 。用它对 $x = 10$ 做预测 ()

- A. 合理, 回归方程可以预测任意 x
- B. 不合理, 属于外推, 没有数据支撑
- C. 合理, 只要代入即可
- D. 无法判断

A12. 观测值 $y_i = 10$, 回归方程预测值 $\hat{y}_i = 8.5$, 则残差为 ()

- A. 1.5
- B. -1.5
- C. 8.5
- D. 10

A13. 关于残差, 下列说法正确的是 ()

- A. 残差越大说明拟合越好
- B. 残差是观测值减预测值, 其绝对值越大拟合越差
- C. 残差一定为正
- D. 残差为零说明数据有误

M4. 某回归方程由 $x \in [40, 60]$ 的样本数据建立, 现在要用它预测 $x = 100$ 时的 y 值。下列说法正确的是 ()

- A. 可以放心预测, 方程是万能的
- B. 属于外推, 没有数据支撑, 预测结果不可靠
- C. 预测结果一定错误
- D. 需要先把 x 除以 2 再预测

C2. (学习时长与成绩的回归) 某班 5 名同学每周课外学习数学的时长 x (小时) 与周测成绩 y (分) 如下:

x	2	3	4	5	6
y	60	65	72	78	85

(1) 求回归方程 $\hat{y} = bx + a$; (2) 预测每周学习 7 小时的同学的周测成绩; (3) 某同学每周学习 6.5 小时, 实际成绩 80 分, 求其残差; 能否用该回归方程预测每周学习 12 小时的同学的成绩? 说明理由。

4.3 独立性检验

考点精要

Q1 列联表与期望频数?

A 把两个分类变量的数据交叉排成 2×2 表, 行合计、列合计、总合计必须三线对应。在假设“两变量无关”下, 某格期望频数 $E = \frac{\text{行合计} \times \text{列合计}}{\text{总人数}}$ 。

Q2 χ^2 统计量怎么算?

A $\chi^2 = \sum \frac{(\text{观测频数} - \text{期望频数})^2}{\text{期望频数}}$, 四格全部代入求和。

Q3 χ^2 与临界值怎么比较?

A $\chi^2 > k$: 拒绝原假设 (原假设 = “两变量无关”), 认为两变量有关联, 犯错误的概率不超过 α ; $\chi^2 \leq k$: 不拒绝原假设, 没有充分证据认为有关联。临界值: $\alpha = 0.05$ 时 $k = 3.841$, $\alpha = 0.01$ 时 $k = 6.635$ 。

Q4 结论的规范表述?

A “在犯错误的概率不超过 α 的前提下, 认为……有关联” 或 “没有充分证据认为……有关联”。注意: “没有充分证据” 不等于 “证明无关”。

例题精讲

例题 4.5

(完善列联表与求期望) 某中学调查了 120 名学生 “是否参加课后体育锻炼” 与 “体质是否达标” 的关系, 得到不完整列联表:

	达标	未达标	合计
参加锻炼	50		
不锻炼		20	
合计			120

且 “不锻炼且达标” 的人数比 “参加锻炼且未达标” 的人数多 10 人。(1) 补全列联表; (2) 求四格的期望频数。

例题 4.6

(计算 χ^2 并下结论) 为研究 “课间是否做眼保健操” 与 “是否近视” 的关系, 调查了 200 名学生, 得到列联表:

	近视	不近视	合计
做操	50	70	120
不做	60	20	80
合计	110	90	200

(1) 求四格期望频数; (2) 计算 χ^2 ; (3) 在 $\alpha = 0.05$ (临界值 3.841) 与 $\alpha = 0.01$ (临界值 6.635) 下分别下结论。

知识点 (1) 列联表与期望频数

A14. 在 2×2 列联表中, 某格所在行的合计为 60、所在列的合计为 40、总人数为 200, 则该格的期望频数为 ()

A. 24 B. 12 C. 3 D. 20

A15. 在 2×2 列联表中, 某格观测频数为 18, 期望频数为 12, 则 $\frac{(O - E)^2}{E}$ 的值为 ()

A. 3 B. 4 C. 1.5 D. 6

A16. 关于期望频数，下列说法正确的是（ ）

- A. 期望频数必须是整数 B. 期望频数是“两变量无关”假设下的理论频数，可以为小数
C. 期望频数越大， χ^2 越大 D. 期望频数由观测数据直接读出

M5. 关于 2×2 列联表与期望频数，下列说法正确的是（ ）

- A. 期望频数 $E = \frac{\text{行合计} \times \text{列合计}}{\text{总人数}}$
B. 期望频数代表“两变量无关”假设下应当出现的频数
C. 补全列联表时，行合计、列合计、总合计必须三线对应
D. 期望频数必须是整数

知识点 (2) 卡方统计量与临界值

A17. 关于独立性检验，下列说法正确的是（ ）

- A. 原假设是“两变量有关联”
B. 若 χ^2 大于临界值，则拒绝原假设，认为两变量有关联
C. 若 χ^2 小于临界值，则证明两变量无关
D. χ^2 越小，越有理由认为两变量有关联

A18. 某调查算得 $\chi^2 = 2.6$ ，临界值 ($\alpha = 0.05$) 为 3.841，则正确的结论是（ ）

- A. 两变量无关 B. 有充分证据认为两变量有关联
C. 没有充分证据认为两变量有关联 D. 需要再算一次

A19. 某疫苗试验，接种组 150 人中感染 15 人，对照组 100 人中感染 25 人。(1) 完善 2×2 列联表 (接种/未接种 $\times$ 感染/未感染)；(2) 计算 χ^2 ；(3) 在 $\alpha = 0.05$ (临界值 3.841) 下判断“接种疫苗与是否感染是否有关联”。

A20. 某调查的 2×2 列联表四格观测频数分别为 30、20、20、30 (总人数 100，行、列合计均各为 50)，则 χ^2 为（ ）

- A. 4 B. 2 C. 8 D. 1

B2. (2022 · 新高考 I 卷 · 第 20 题) 一医疗团队为研究某地的一种地方性疾病与当地居民的卫生习惯 (卫生习惯分为良好和不够良好两类) 的关系，在已患该疾病的病例中随机调查了 100 例 (称为病例组)，同时在未患该疾病的人群中随机调查了 100 人 (称为对照组)，得到如下数据：

	不够良好	良好	合计
病例组	40	60	100
对照组	10	90	100
合计	50	150	200

(1) 能否有 99% 的把握认为患该疾病群体与未患该疾病群体的卫生习惯有差异？($\alpha = 0.01$ 时临界值 6.635) (2) 从该地的人群中任选一人， A 表示事件“选到的人卫生习惯不够良好”， B 表示事件“选到的人患有该疾病”。 $\frac{P(B|A)}{P(\bar{B}|A)} \cdot \frac{P(\bar{B}|\bar{A})}{P(B|\bar{A})}$ 是卫生习惯不够良好对患该

疾病风险程度的一项度量指标，记该指标为 R 。(i) 证明： $R = \frac{P(A|B)}{P(\bar{A}|B)} \cdot \frac{P(\bar{A}|\bar{B})}{P(A|\bar{B})}$ ；(ii) 利用该调查数据，给出 $P(A|B)$ 、 $P(A|\bar{B})$ 的估计值，并利用 (i) 的结果给出 R 的估计值。(注：本题统计量在原卷中记作 K^2 ，与新课标教材中的 χ^2 是同一统计量，计算与判定规则完全一致。)

B3. (2023 · 全国甲卷 (理) · 第 19 题) 为探究某药物对小鼠的生长抑制作用，将 40 只小鼠均分为两组，分别为对照组 (不加药物) 和实验组 (加药物)。(1) 设其中两只小鼠中对照组小鼠数目为 X ，求 X 的分布列和数学期望。(本小题涉及超几何分布，属第 6 章概率内容，可暂缓。)(2) 测得 40 只小鼠体重如下 (单位：g，已按从小到大排好)：

对照组：17.3, 18.4, 20.1, 20.4, 21.5, 23.2, 24.6, 24.8, 25.0, 25.4, 26.1, 26.3, 26.4, 26.5, 26.8, 27.0, 27.4, 27.5, 27.6, 28.3

实验组：5.4, 6.6, 6.8, 6.9, 7.8, 8.2, 9.4, 10.0, 10.4, 11.2, 14.4, 17.3, 19.2, 20.2, 23.6, 23.8, 24.5, 25.1, 25.2, 26.0

(i) 求 40 只小鼠体重的中位数 m ，并完成下面 2×2 列联表：

	体重小于 m	体重不小于 m	合计
对照组			
实验组			
合计			

(ii) 根据列联表，能否有 95% 的把握认为药物对小鼠生长有抑制作用？($\alpha = 0.05$ 时临界值 3.841)

B5. (2022 · 全国甲卷 (文) · 第 17 题) 甲、乙两城之间的长途客车均由 A 和 B 两家公司运营。为了解这两家公司长途客车的运行情况，随机调查了甲、乙两城之间的 500 个班次，得到下面列联表：

	准点班次数	未准点班次数
A 公司	240	20
B 公司	210	30

(1) 根据上表，分别估计这两家公司甲、乙两城之间的长途客车准点的概率；(2) 能否有 90% 的把握认为甲、乙两城之间的长途客车是否准点与客车所属公司有关？

附： $K^2 = \frac{n(ad - bc)^2}{(a + b)(c + d)(a + c)(b + d)}$ ，其中 $n = a + b + c + d$ 。临界值表： $P(K^2 \geq k) = 0.100$ 时 $k = 2.706$ ， $= 0.050$ 时 $k = 3.841$ ， $= 0.010$ 时 $k = 6.635$ 。

M6. 某调查算得 $\chi^2 = 5.02$ ，临界值 ($\alpha = 0.05$) 为 3.841。下列结论正确的是 ()

- A. 不拒绝原假设，两变量无关
- B. 拒绝原假设，在犯错误的概率不超过 0.05 的前提下认为两变量有关联
- C. χ^2 越大说明原假设越成立
- D. 无法判断

M7. 某调查算得 $\chi^2 = 2.1$ ，小于临界值 3.841。下列说法正确的是 ()

- A. 证明了两个变量无关
- B. 没有充分证据认为两变量有关联
- C. 两个变量一定无关
- D. 样本量不够，需要增大样本量重做

本章综合

S1. 某机构调查了 8 名高三学生的“周学习总时长 x (小时)”与“周测数学成绩 y (分)”:

x	34	38	42	46	50	54	58	62
y	68	74	80	86	92	98	104	110

(1) 求 $\bar{x}, \bar{y}$; (2) 求回归方程 $\hat{y} = bx + a$; (3) 预测每周学习 47 小时的学生的成绩; (4) 求样本点 (60, 96) 处的残差。

S2. 某商场调查了顾客的“购物袋使用习惯”与“是否支持限塑令”的关系，得到列联表:

	支持	反对	合计
自带购物袋	70	30	100
不自带购物袋	50	50	100
合计	120	80	200

(1) 求四格期望频数; (2) 计算 χ^2 ; (3) 在 $\alpha = 0.05$ (临界值 3.841) 与 $\alpha = 0.01$ (临界值 6.635) 下分别下结论，并说明结论的意义。

第五章 数学建模与拓展

5.1 建模四步：从现实到数学再回现实

考点精要

Q1 建模四步法是什么？

A 审题（找变量 + 写清现实定义域）、建模（按数据规律选函数模型）、求解（方程、不等式、函数性质）、检验与回写（结果合不合理？把数学答案翻译回现实）。四步缺一不可，检验环节最常被偷工减料。

Q2 现实定义域怎么定？

A 人数必为整数、时间非负、质量有上限。数学顶点不一定可取——卖不了“半台”打印机，算完要回到现实约束检查。

Q3 模型选择的基本判据？

A 后项 - 前项 $\approx$ 常数对应线性；后项 $\div$ 前项 $\approx$ 常数对应指数；差再差 $\approx$ 常数对应二次；不同区间规律不同对应分段。

Q4 检验的两道关？

A 数值检验（回代已知点看残差）+ 现实检验（定义域、值域、趋势是否违背常识）。检验不过就回头改模型，迭代。

例题精讲

例题 5.1

（完整四步：咖啡降温）一杯 86°C 的咖啡放在 22°C 的房间里，每 3 分钟测一次温度，得 $t = 0, 3, 6, 9$ (min) 时 $T = 86, 70, 58, 49$ ($^{\circ}\text{C}$)。 (1) 判断该用线性还是指数模型； (2) 建立 T 关于 t 的模型； (3) 预测 $t = 12$ 、 $t = 15$ 时的温度； (4) 检验模型是否合理。

例题 5.2

（定义域的现实约束）某商店每月销售 x 台打印机 (x 为自然数)，月利润 y (万元) $= -x^2 + 15x - 44$ 。问月利润最大时卖出多少台？最大利润是多少？

知识点 (1) 建模四步与模型选择

A1. 数据 $x = 0, 1, 2, 3$ 对应 $y = 7, 10, 13, 16$ 。适合的模型是 (), 解析式为 ()

A. 线性, $y = 3x + 7$ B. 指数, $y = 7 \times 2^x$ C. 二次, $y = x^2 + 7$ D. 线性, $y = 3x$

A2. 用 30 m 长的栅栏围一块矩形菜园, 设长为 x m。面积 S 关于 x 的关系式为 (), 定义域为 (), 最大面积为 () m^2

A. $S = x(15 - x)$, $0 < x < 15$, 56.25 B. $S = x(30 - x)$, $0 < x < 30$, 225

C. $S = x(15 - x)$, $0 < x < 15$, 60 D. $S = x(30 - x)$, $0 \leq x \leq 30$, 225

A3. 下列选项能完整体现“数学建模四步法”合理顺序的是 ()

A. 求解、审题、建模、检验 B. 审题、建模、求解、检验与回写

C. 建模、审题、检验、求解 D. 检验、建模、审题、求解

A4. 在“把现实问题翻译成数学”这一步, 下列做法不恰当的是 ()

A. 确定自变量与因变量, 并写清现实定义域

B. 根据数据特征选择函数模型

C. 直接套用一个最复杂的模型, 不考虑数据形状

D. 先统一单位, 再列关系式

M1. 下列做法属于“套模板解应用题”而非“数学建模”的是 ()

A. 先审题找变量, 再按数据选模型

B. 看到“增长”就套指数公式, 不检验

C. 算完把结果翻译回现实

D. 数据不符合模型时回头修改模型

C1. (咖啡冷却模型) 一杯 88°C 的咖啡放在 24°C 的房间里, 每 5 分钟测一次温度, 得 $t = 0, 5, 10, 15$ (min) 时 $T = 88, 72, 60, 51$ ($^\circ\text{C}$)。 (1) 先看“高出室温的部分” $T - 24$, 判断该用线性还是指数模型, 并说明依据; (2) 建立 T 关于 t 的模型; (3) 预测 $t = 25$ 时的温度; (4) 检验: 当 t 越来越大时, 模型给出的温度趋向多少? 这符合实际吗?

知识点 (2) 现实定义域与模型检验

A5. 某模型预测某商品第 8 个月的销量为 -5 件。这个结果不合理, 因为 ()

A. 数学算对了就应采信 B. 销量不可能为负, 应检查数据趋势并修正模型

C. 模型越复杂越好 D. 负值说明商品滞销

A6. 某商品进价 20 元, 售价 x 限制在 $[20, 50]$ (元), 日销量 $y = 100 - x$ (件), 则日利润最大时的售价为 ()

A. 20 元 B. 50 元 C. 60 元 D. 40 元

M2. 某商店每月销售 x 台打印机 (x 为自然数), 月利润 $y = -x^2 + 15x - 44$ (万元)。某同学算出对称轴 $x = 7.5$, 直接回答“当 $x = 7.5$ 时利润最大, 最大利润 12.25 万元”。他的答案 ()

A. 正确, $x = 7.5$ 就是最佳值

B. 错误, 打印机台数必须是整数, 应在 $x = 7$ 与 $x = 8$ 处比较, 最大利润 12 万元

- C. 正确，利润可以是小数
D. 无法判断

5.2 模型选择与检验：数据形状决定模型家族

考点精要

Q1 差稳定还是比稳定？

A 后项 - 前项 $\approx$ 常数对应线性；后项 $\div$ 前项 $\approx$ 常数对应指数；差再差 $\approx$ 常数对应二次。模型选择不是猜，要实际算差与比。

Q2 什么是过拟合？

A 模型过于复杂，把已知点全部精确穿过（残差为零），但对新数据预测反而更差。评判模型的标准不是“对已知点拟合多完美”，而是“对新数据预测多靠谱”。

Q3 什么是奥卡姆剃刀？

A 解释力相当的情况下，选择更简单的模型。复杂不是优点，恰当才是。

Q4 指数增长能无限持续吗？

A 不能。现实有资源约束，长期看增长会放缓（S 型修正）。检验环节就是专门拦下“数学漂亮、现实荒谬”的模型。

例题精讲

例题 5.3

（三组数据选模型）判断下列三组数据各适合什么模型，并说明依据：A: $x = 1, 2, 3, 4, 5$, $y = 4, 7, 10, 13, 16$ ；B: $x = 1, 2, 3, 4, 5$, $y = 3, 6, 12, 24, 48$ ；C: 某快递计费——首重 1 kg 内 10 元，续重每 1 kg 加 4 元（不足 1 kg 按 1 kg 计）。

例题 5.4

（过拟合对比）某产品连续 3 个月的销量数据：(0, 1.2), (1, 2.5), (2, 2.0) (x 为月， y 为万件)。两个候选模型：模型甲（线性）： $\hat{y}_甲 = 0.4x + 1.4$ ；模型乙（二次，精确穿过三点）： $y_乙 = -0.9x^2 + 2.2x + 1.2$ 。(1) 分别求两模型对三个已知点的残差平方和；(2) 分别预测 $x = 3$ 时的销量；(3) 哪个模型更值得用？说明理由。

知识点 (1) 数据形状决定模型家族

A7. 数据 3, 6, 12, 24, 48 适合 () 模型，依据是 ()

- A. 线性，差恒为 3 B. 指数，相邻比恒为 2
C. 二次，二阶差恒定 D. 无法判断

A8. 数据 2, 5, 8, 11, 14 适合 () 模型，依据是 ()

- A. 线性，差恒为 3 B. 指数，相邻比恒定

C. 二次 D. 分段

A9. 数据 $x = 1, 2, 3, 4$, $y = 1, 4, 9, 16$ 适合 () 模型, 依据是 ()

A. 线性, 差恒为 3 B. 指数, 比恒为 2 C. 二次, 一阶差为 3, 5, 7 (二阶差恒为 2) D. 分段

A10. 某城市年用电量每年比上一年增长约 8%, 用模型刻画用电量随年份 t 的变化, 最适合的是 ()

A. 线性模型 B. 指数模型 C. 二次模型 D. 反比例模型

M3. 两组递增数据: 甲组 4, 7, 10, 13, 乙组 4, 8, 16, 32。下列说法正确的是 ()

A. 都适合线性模型 B. 甲适合线性 (差恒为 3), 乙适合指数 (比恒为 2)
C. 都适合指数模型 D. 数据都递增, 模型无法区分

B1. (2023 · 新课标 I 卷 · 第 10 题) 噪声污染问题越来越受到重视。用声压级来度量声音的强弱, 定义声压级 $L_p = 20 \lg \frac{p}{p_0}$, 其中常数 p_0 ($p_0 > 0$) 是听觉下限阈值, p 是实际声压。下表为不同声源的声压级:

声源	与声源的距离/m	声压级/dB
燃油汽车	10	60 90
混合动力汽车	10	50 60
电动汽车	10	40

已知在距离燃油汽车、混合动力汽车、电动汽车 10 m 处测得实际声压分别为 p_1, p_2, p_3 , 则 ()

A. $p_1 \geq p_2$ B. $p_2 > 10p_3$ C. $p_3 = 100p_0$ D. $p_1 \leq 100p_2$

知识点 (2) 拟合、过拟合与检验

A11. 某同学用高次多项式模型拟合 10 个已知点, 已知点残差全为 0, 但预测第 11 个点时结果剧烈波动。这说明 ()

A. 模型很好, 应继续使用 B. 模型过拟合, 应改用更简单的模型并重新评估
C. 数据有问题 D. 高次多项式一定比低次好

A12. 最小二乘法用“残差平方和”而不用“残差和”来度量拟合好坏, 最主要的原因是 ()

A. 残差都是正数, 不用平方
B. 残差有正有负, 直接相加会互相抵消, 看不出拟合优劣
C. 平方后计算更简单
D. 平方可以消除数据的单位

A13. 某线性回归对 5 个已知点的残差依次为 \$0.1, -0.2, 0.1, -0.3, 0.3\$, 则残差平方和为 ()

A. 0.24 B. 1.0 C. 0.10 D. 0.024

M4. 某同学用指数模型预测某地人口: $N = 100 \times 1.05^t$ (万人), t 为年份。算得 $t = 50$ 时约 1147 万人, 而该地面积狭小、资源有限。下列说法正确的是 ()

- A. 数学算对了就应采信
 B. 应结合现实检验：人口不可能长期指数增长，模型应修正为 S 型
 C. 预测值很大说明模型很好
 D. 只要预测值不为负数就合理

C2. (共享单车投放模型) 某城市共享单车投放初期快速增长，投放量 (万辆) 逐年记录：第 0~4 年分别为 1, 2, 4, 8, 16。(1) 判断适合什么模型并建立投放量 N 关于年份 t 的模型；(2) 预测第 7 年的投放量；(3) 检验：该城市人口约 200 万，第 7 年预测值是否合理？说明理由，并指出模型应如何修正。

B2. (2023 · 新课标 II 卷 · 第 19 题) 某研究小组经过研究发现某种疾病的患病者与未患病者的某项医学指标有明显差异。经过大量调查，得到如下频率分布直方图 (【文字化还原】直方图关键数据)：患病者该项指标的频率分布直方图：区间 $[95, 100]$ 、 $[100, 105]$ 的频率与组距之比分别为 0.002、0.012；未患病者该项指标的频率分布直方图：区间 $[95, 100]$ 、 $[100, 105]$ 的频率与组距之比分别为 0.01、0.002；假设数据在组内均匀分布。利用该指标制定一个检测标准，需要确定临界值 c ，将该指标大于 c 的人判定为阳性，小于或等于 c 的人判定为阴性。此检测标准的漏诊率是将患病者判定为阴性的概率，记为 $p(c)$ ；误诊率是将未患病者判定为阳性的概率，记为 $q(c)$ 。以事件发生的频率作为相应事件发生的概率。(1) 当漏诊率 $p(c) = 0.5\%$ 时，求临界值 c 和误诊率 $q(c)$ ；(2) 设函数 $f(c) = p(c) + q(c)$ 。当 $c \in [95, 105]$ 时，求 $f(c)$ 的解析式，并求 $f(c)$ 在区间 $[95, 105]$ 的最小值。

5.3 拓展 · 矩阵初步

考点精要

Q1 矩阵怎么加减、数乘？

A 同型矩阵对应元素相加；数乘是每个元素都乘这个数。

Q2 矩阵乘法怎么算？

A $m \times n$ 矩阵乘 $n \times p$ 矩阵得 $m \times p$ 矩阵，结果第 i 行第 j 列元素 = 左矩阵第 i 行与右矩阵第 j 列对应相乘后求和 (行乘列求和)。

Q3 矩阵乘法满足交换律吗？

A 不满足。 $AB \neq BA$ 一般成立，与数的乘法完全不同。

Q4 线性方程组怎么写成分矩阵形式？

A
$$\begin{cases} a_{11}x + a_{12}y = b_1 \\ a_{21}x + a_{22}y = b_2 \end{cases} \quad \text{可写作} \quad \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \text{——矩阵是“一批线性关系的打包”。}$$

例题精讲

例题 5.5

(矩阵运算) $A = \begin{pmatrix} 1 & 3 \\ 2 & 0 \end{pmatrix}$, $B = \begin{pmatrix} 4 & 1 \\ 1 & 5 \end{pmatrix}$ 。求 $A+B$ 、 $2A$ 、 AB 、 BA ，并比较 AB 与 BA 。

例题 5.6

(方程组的矩阵形式) 把方程组 $\begin{cases} 3x + 2y = 12 \\ 2x + y = 7 \end{cases}$ 写成矩阵形式，并求其解。

知识点 (1) 矩阵的加减与数乘

A14. 计算: $\begin{pmatrix} 1 & 3 \\ 2 & 0 \end{pmatrix} + \begin{pmatrix} 4 & 1 \\ 1 & 5 \end{pmatrix} = (\quad)$

A. $\begin{pmatrix} 5 & 4 \\ 3 & 5 \end{pmatrix}$ B. $\begin{pmatrix} 5 & 4 \\ 1 & 5 \end{pmatrix}$ C. $\begin{pmatrix} 4 & 3 \\ 2 & 0 \end{pmatrix}$ D. $\begin{pmatrix} 5 & 3 \\ 4 & 5 \end{pmatrix}$

A15. 计算: $2 \begin{pmatrix} 1 & -3 \\ 0 & 4 \end{pmatrix} = (\quad)$

A. $\begin{pmatrix} 2 & -6 \\ 0 & 8 \end{pmatrix}$ B. $\begin{pmatrix} 2 & -3 \\ 0 & 4 \end{pmatrix}$ C. $\begin{pmatrix} 1 & -6 \\ 0 & 8 \end{pmatrix}$ D. $\begin{pmatrix} 2 & -6 \\ 0 & 4 \end{pmatrix}$

A16. 已知 $\begin{pmatrix} a & 1 \\ 2 & b \end{pmatrix} + \begin{pmatrix} 1 & c \\ d & 3 \end{pmatrix} = \begin{pmatrix} 3 & 2 \\ 4 & 5 \end{pmatrix}$, 则 ()

A. $a = 2, b = 2, c = 1, d = 2$ B. $a = 2, b = 5, c = 1, d = 2$ C. $a = 4, b = 8, c = 3, d = 6$
D. 无法确定

知识点 (2) 矩阵乘法与线性方程组

A17. 计算: $\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 1 & 1 \end{pmatrix} = (\quad)$

A. $\begin{pmatrix} 4 & 2 \\ 10 & 4 \end{pmatrix}$ B. $\begin{pmatrix} 4 & 2 \\ 7 & 3 \end{pmatrix}$ C. $\begin{pmatrix} 2 & 2 \\ 6 & 4 \end{pmatrix}$ D. $\begin{pmatrix} 5 & 1 \\ 10 & 4 \end{pmatrix}$

A18. 方程组 $\begin{cases} 2x + 3y = 8 \\ x - y = 1 \end{cases}$ 的矩阵形式为 $\begin{pmatrix} 2 & 3 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 8 \\ 1 \end{pmatrix}$, 其解为 ()

A. $x = 2, y = \frac{4}{3}$ B. $x = \frac{11}{5}, y = \frac{6}{5}$ C. $x = 3, y = \frac{2}{3}$ D. $x = 1, y = 2$

M5. 计算 $\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix}$ 。某同学按“对应元素相乘”算得 $\begin{pmatrix} 5 & 12 \\ 21 & 32 \end{pmatrix}$ 。他的算法 ()

A. 正确 B. 错误，矩阵乘法是“行乘列求和”
C. 有时正确 D. 两种算法结果相同

M6. 设 $A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, $B = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, 则 ()

A. $AB = BA$ B. $AB \neq BA$, 矩阵乘法一般不满足交换律

C. $AB = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, BA 为零矩阵 D. 无法相乘

5.4 拓展 · 逻辑初步与数学史、新技术

考点精要

Q1 原命题与逆否命题什么关系?

A 原命题 $p \Rightarrow q$ 与逆否命题 $\neg q \Rightarrow \neg p$ 同真同假 (互为等价)。

Q2 反证法的思路?

A 要证 p , 先假设 p 的反面成立, 推出与已知条件或公理矛盾, 从而证明 p 成立。

Q3 “最小二乘”为什么要用平方?

A 残差有正有负会互相抵消, 平方让正负都变正、且放大离得远的点——这也是方差思想的来源。

Q4 软件/计算器输出的结果可以直接用吗?

A 不能。工具负责“算”, “选什么模型、结论怎么读”始终是人做的判断。对软件输出先问一句“散点图长什么样”。

例题精讲

例题 5.7

(反证法) 证明: 若 x, y 为实数且 $x + y > 2$, 则 $x > 1$ 或 $y > 1$ 。

例题 5.8

(新技术输出解读) 某同学用软件对一组“施肥量—产量”数据做回归, 软件输出线性回归 $\hat{y} = 3.2x + 50$, $R^2 = 0.97$ 。但他在生成散点图时发现数据随施肥量增加先增后减 (抛物线形状)。(1) 这个 $R^2 = 0.97$ 能说明线性模型合适吗? (2) 应当怎么办?

知识点 (1) 命题逻辑与反证法

A19. 命题“若 $x > 1$, 则 $x^2 > 1$ ”是 () 命题, 它的逆否命题是“若 $x^2 \leq 1$, 则 x ()”

A. 真, ≤ 1 B. 真, > 1 C. 假, ≤ 1 D. 假, > 1

A20. 命题“若 $x > 2$, 则 $x^2 > 4$ ”的逆否命题是 ()

A. 若 $x^2 > 4$, 则 $x > 2$ B. 若 $x \leq 2$, 则 $x^2 \leq 4$ C. 若 $x^2 \leq 4$, 则 $x \leq 2$ D. 若 $x^2 \leq 4$, 则 $x < 2$

A21. 用反证法证明“质数有无穷多个”, 正确的第一步是 ()

- A. 假设质数只有有限个 B. 假设存在一个最大的质数 C. 假设质数不存在 D. 直接列举出所有的质数

知识点 (2) 数学史与新技术输出解读

A22. 软件对某组数据输出指数拟合 $N = 800 \times 1.5^t$, $R^2 = 0.98$, 但该场景是培养皿细菌且营养早已耗尽。这个输出能直接用吗? 为什么?

- A. 能用, R^2 高说明模型好
B. 不能用, 指数增长在现实上不可持续, 应结合背景改用 S 型模型
C. 能用, 软件算出来的一定正确
D. 不能用, 必须手算

M7. 某同学用软件对一组“投入—产出”数据拟合出线性回归, $R^2 = 0.99$, 直接采信并用于决策。下列说法正确的是 ()

- A. R^2 高说明模型一定合适
B. 应先检查散点图判断模型类型, 再决定是否采信
C. 软件输出一定正确
D. 工具可以完全代替人的原理判断

B3. (2022 · 全国甲卷 · 理 · 第 8 题) 沈括的《梦溪笔谈》是中国古代科技史上的杰作, 其中收录了计算圆弧长度的“会圆术”。如图 (【文字化还原】): $\widehat{AB}$ 是以 O 为圆心、 OA 为半径的圆弧, C 是 AB 的中点, D 在 $\widehat{AB}$ 上, $CD \perp AB$ 。“会圆术”给出 $\widehat{AB}$ 的弧长的近似值 s 的计算公式: $s = AB + \frac{CD^2}{OA}$ 。当 $OA = 2$, $\angle AOB = 60^\circ$ 时, $s = ()$

- A. $\frac{11 - 3\sqrt{3}}{2}$ B. $\frac{11 - 4\sqrt{3}}{2}$ C. $\frac{9 - 3\sqrt{3}}{2}$ D. $\frac{9 - 4\sqrt{3}}{2}$

本章综合

S1. 某手机品牌发布一款新机, 上市后前 4 个月的销量 (万部) 分别为 4, 8, 16, 32。(1) 判断适合什么模型, 并建立销量 N 关于月份 x 的模型; (2) 预测第 5、6 个月的销量; (3) 检验: 若该型号手机的目标用户约 5000 万且市场逐渐饱和, 用第 (2) 问的模型预测第 10 个月的销量是否合理? 应如何修正?

S2. (矩阵调运模型) 某工厂用 A 、 B 两种原料生产甲、乙两种产品: 每件甲产品需 A 原料 2 kg、 B 原料 1 kg; 每件乙产品需 A 原料 1 kg、 B 原料 3 kg。(1) 用矩阵 M 表示“每件产品消耗的原料量”(行对应原料, 列对应产品); (2) 若本月计划生产甲产品 100 件、乙产品 200 件, 用矩阵乘法求两种原料的总需求量; (3) 若仓库现有 A 原料 400 kg、 B 原料 500 kg, 按第 (2) 问的生产计划是否可行? 说明理由。

第六章 卷末综合检测

一、单项选择题（本大题共 5 小题，每小题 5 分，共 25 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。）

Z1. 从 4 名男生和 3 名女生中选出 3 人组成志愿者小组，要求至少包含 1 名女生，不同的选法数为（ ）

A. 31 B. 34 C. 18 D. 45

Z2. 将一枚质地均匀的骰子连续抛掷两次，两次点数之和为 8 的概率是（ ）

A. $\frac{1}{6}$ B. $\frac{5}{36}$ C. $\frac{1}{9}$ D. $\frac{7}{36}$

Z3. 某校高一、高二、高三人数之比为 3:2:1，采用分层抽样抽取样本量为 60 的样本，则高二年级应抽取（ ）

A. 20 人 B. 30 人 C. 10 人 D. 15 人

Z4. 已知一组数据的方差为 4，若每个数据都乘以 3 后再加 2，则新数据的方差为（ ）

A. 12 B. 14 C. 36 D. 38

Z5. 已知两个变量 x, y 的回归直线过点 $(3, 8)$ ，回归方程为 $\hat{y} = bx + a$ ，且 $b = 2$ ，则 a 的值为（ ）

A. 2 B. 1 C. -2 D. 5

二、多项选择题（本大题共 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，全部选对得 6 分，部分选对得部分分，有选错的得 0 分。）

Z6. 已知一组样本数据 $x_1, x_2, \dots, x_8$ 的平均数为 10，方差为 4，则下列说法正确的是（ ）

- A. 该组数据的中位数一定为 10
- B. 若该组数据的最大值为 20，则极差不小于 10
- C. 数据 $2x_1 + 1, 2x_2 + 1, \dots, 2x_8 + 1$ 的平均数为 21
- D. 数据 $2x_1 + 1, 2x_2 + 1, \dots, 2x_8 + 1$ 的方差为 16

Z7. 关于相关系数 r 与回归分析，下列说法正确的是（ ）

- A. 若 $r = -0.9$ ，则两变量负相关且线性相关较强
- B. 回归直线必过样本中心点 $(\bar{x}, \bar{y})$
- C. 若两变量的相关系数 r 接近 0，则两变量一定没有关系
- D. 用回归方程对样本范围外的 x 做预测，结果不一定可靠

Z8. 某生物种群数量 N 随时间 t (周) 的变化数据为: $t = 0, 1, 2, 3, 4$ 时 $N = 100, 140, 196, 274, 384$ 。下列说法正确的是 ()

- A. 该数据相邻两项的比约为 1.4, 适合用指数模型近似
- B. 该种群数量会一直按此比例无限增长
- C. 用指数模型预测 $t = 10$ 时的数量约为 100×1.4^{10}
- D. 若培养空间与营养有限, 长期看模型应修正为 S 型

三、填空题 (本大题共 3 小题, 每小题 5 分, 共 15 分。)

Z9. 从 0, 1, 2, 3, 4 五个数字中任取三个不同数字组成三位数, 则能组成_____ 个不同的三位数。

Z10. 一组数据从小到大排列为 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 则该组数据的第 70 百分位数为_____。

Z11. 甲、乙两人独立地射击同一目标, 甲命中的概率为 0.8, 乙命中的概率为 0.7, 则目标被击中的概率为_____。

四、解答题 (本大题共 3 小题, 共 92 分。解答应写出文字说明、证明过程或演算步骤。)

Z12. (统计大题, 30 分)

某校 100 名学生参加引体向上体质测试, 成绩 (次) 分组及频数如下:

次数	[0, 5)	[5, 10)	[10, 15)	[15, 20)	[20, 25)
人数	10	25	35	20	10

(1) 求各组频率, 并写出频率分布直方图各矩形的高度 (组距为 5); (2) 估计这 100 名学生成绩的中位数; (3) 估计这 100 名学生成绩的平均数; (4) 若规定 15 次及以上为“优秀”, 估计该校学生引体向上的优秀率。

Z13. (概率大题, 30 分)

某班要从 5 名男生、4 名女生中选出 3 人参加数学竞赛。(1) 共有多少种不同的选法? (2) 求选出的 3 人中恰有 1 名女生的概率; (3) 求选出的 3 人中至少有 1 名女生的概率。

Z14. (统计与成对数据综合, 32 分)

某公司统计了过去 5 个月的广告投入 x (万元) 与当月销售额 y (万元):

x	1	2	3	4	5
y	11	15	19	23	27

(1) 求回归方程 $\hat{y} = bx + a$; (2) 预测广告投入 6 万元时的销售额; (3) 该公司还调查了“是否投放短视频广告”与“销售额是否较上月提升”的关系, 得到列联表 (单位: 家): 投放短视频广告的 60 家中 45 家销售额提升; 未投放短视频广告的 40 家中 20 家提升。请完善 2×2 列联表并计算 χ^2 , 判断能否在犯错误的概率不超过 0.05 的前提下认为“投放短视频广告与销售额提升有关联” ($\alpha = 0.05$ 时临界值 3.841); (4) 有同事提出“用回归方程预测广告投入 20 万元时的销售额”, 你认为是否合理? 说明理由。