

PATHWAY

数学 III

循迹而上 · 拾级登顶

寻迹 · 梦叶子 编著

目录

第一章 怎么数清楚每一种可能：计数原理	1
1.1 分类加法与分步乘法计数原理：要么“分叉”，要么“接力”	1
1.2 排列：有座位的挑人	5
1.3 组合：只管打包	9
1.4 二项式定理：组合数的“展开舞台”	12
第二章 怎么用数字度量“运气”：概率	21
2.1 随机事件与古典概型：给“可能”排队编号	21
2.2 条件概率与全概率：信息来了，空间缩了	25
2.3 事件的独立性与二项分布：互不干扰的重复	29
2.4 数字特征与正态分布初步：分布的“平均数”与“钟形曲线”	32
第三章 从数据里看世界：统计	41
3.1 抽样：从“全查”到“抽查”	41
3.2 频率分布直方图：面积才是频率	45
3.3 百分位数：排名是怎么定的	47
3.4 中心与离散：数据稳不稳	50
第四章 两个变量是“伙伴”还是“路人”：成对数据的统计分析	57
4.1 相关关系：是“伙伴”还是“路人”	57
4.2 回归：用一条线做预测	61
4.3 独立性检验：证据够不够	63
第五章 把现实装进模型：数学建模与拓展	71
5.1 建模四步：从现实到数学再回现实	71
5.2 模型选择：数据形状决定模型家族	73
5.3 拓展·矩阵初步：多元关系的一张表	75

5.4 拓展·逻辑初步与数学史	78
第六章 卷末综合检测	85

第一章 怎么数清楚每一种可能：计数原理

章首情境

1.1 分类加法与分步乘法计数原理：要么“分叉”，要么“接力”

学习目标

- 理解分类加法计数原理，会用加法数“要么…要么…”的可能；
- 理解分步乘法计数原理，会用乘法数“先…再…”的可能；
- 会辨别一个计数问题该分类还是分步，能综合运用两个原理；
- 会检查分类的“不重不漏”与分步的“层层独立”。

前置复习 (Be Prepared)

- 从 3 件上衣中选 1 件，有几种选法？
- 从 3 件上衣、4 条裤子中各选 1 件，有几种选法？
- 用列举法写出集合 $\{1, 2, 3\}$ 的所有子集，数一数共有几个。

1.1.1 分类加法计数原理：“要么…要么…”

从生活到数学

先回到“只吃一顿”：你有 3 类来源，每类内部再细分，但无论选哪一类中的哪一种，“吃一顿饭”这件事立刻完成。选套餐、选面条、选沙拉，三条路径是“平行”的——走到任何一条路径的终点，任务就完成了。

分类加法计数原理

完成一件事有 k 类方案，第 1 类有 m_1 种方法，第 2 类有 m_2 种方法，……，第 k 类有 m_k 种方法，并且**每一类中的每一种方法都能独立完成这件事**（各类方法互不相同），那么完成这件事共有

$$N = m_1 + m_2 + \cdots + m_k$$

种不同的方法。

How To: 判断能否用分类加法计数原理

- 问自己：选完这一项，“整件事”完成了吗？
- 若**完成了** → 这件事被拆成了“若干类”，各类方法**相加**；
- 检查各类是否**互不重叠**（不重）、合起来是否**覆盖全部**（不漏）。

分类加法示意图（三类方案平行，任选其一即完成“报名”这件事）

方案一音乐社：3 个方向 方案二科技社：4 个方向 → $3 + 4 + 2 = 9$ （“要么…要么…”，用加法） 方案三美术社：2 个方向

三条路径是平行的——无论走哪一条，走到终点任务就完成，因此把各类方法数**相加**。

例题 1.1

（分类加法计数原理）某班进行社团报名，一人限报一个社团：音乐社有 3 个可报方向，科技社有 4 个可报方向，美术社有 2 个可报方向。小明任意报一个方向，共有多少种选择？**解**：选报任何一个方向，都**立即完成**“报名”这件事，所以属于分类加法问题。三类方向互不重叠、合起来覆盖全部方向：

$$N = 3 + 4 + 2 = 9$$

答：共有 **9** 种选择。**Try It 1.1**：从甲地到乙地，可以乘火车（有 4 趟）、汽车（有 3 趟）、飞机（有 2 班）。一天中从甲地到乙地共有多少种不同的走法？

例题 1.2

（分类的“不重不漏”检查）从 1, 2, 3, 4, 5 中任取两个不同的数相加。小华这样分类：“和为偶数的”与“和为奇数的”。请问这样的分类是否“不重不漏”？**解**：任意两个数之和不是偶数就是奇数，两类**覆盖了全部**情况（不漏）；且一个“和”不可能既是偶数又是奇数，两类**不重叠**（不重）。因此分类满足“不重不漏”，两类的数量之和正好等于所有“取两个数相加”的总数 $C_5^2 = 10$ 。（分类标准的正确与否，直接决定后续计数是否正确——分类一旦交叉，和就会重复计数。）**Try It 1.2**：把“从 1, 2, 3, 4, 5 中取两个数”分类为“含 1 的”与“不含 1 的”，是否不重不漏？**想一想（对抗 M3）**：用 0, 1, 2, 3 四个数字组成无重复数字的四位数。小刚说：“四位数的个数 = 首位为 1 的 + 首位为 2 的 + 首位为 3 的，共 3 类。”他的分类漏掉了哪一类？如果他想“不漏”，应怎么修正？（提示：首位不能是几？四个数字里哪个数字没有资格当首位？）

1.1.2 分步乘法计数原理：“先…再…”

从生活到数学

再回到“先打主食、再打饮料”：主食有 5 种、饮料有 3 种。只选主食，“这一顿饭”还没完成——你还没拿到饮料；只选饮料，饭也没影。**必须把两步都走完，这件事才算完成。**两步像接力赛，缺任何一步都“没完成”。

分步乘法计数原理

完成一件事需要分成 k 个步骤，做第 1 步有 m_1 种方法，做第 2 步有 m_2 种方法，……，做第 k 步有 m_k 种方法，那么完成这件事共有

$$N = m_1 \times m_2 \times \cdots \times m_k$$

种不同的方法。

How To: 判断能否用分步乘法计数原理

- 问自己：选完这一项，“整件事”完成了吗？
- 若**没有完成**，还需要继续选 → 这件事被拆成了“若干步”，各步方法**相乘**；
- 检查每一步的选法数是否**不随前面步骤变化**（若会变，通常要分类讨论或换模型）。

分步乘法示意图（两步接力，走完才算完成“先打主食、再打饮料”这件事）

第 1 步主食：5 种 → 第 2 步饮料：3 种

$5 \times 3 = 15$ （“先…再…”，用乘法——两步像接力赛，缺任何一步都“没完成”）

例题 1.3

（分步乘法计数原理）从甲地到丙地必须先经过乙地。甲地到乙地有 3 条路可走，乙地到丙地有 2 条路可走。从甲地经乙地到丙地，共有多少种不同的走法？**解：**这件事要“先走甲 → 乙，再走乙 → 丙”，分两步完成：第 1 步：甲 → 乙，有 3 种；第 2 步：乙 → 丙，有 2 种。

$$N = 3 \times 2 = 6$$

答：共有 **6** 种不同的走法。（画一张树形图：从 3 个“枝头”各分出 2 个“枝梢”，共 $3 \times 2 = 6$ 个枝梢。）**Try It 1.3:** 穿衣服出门：上衣 3 件、裤子 4 条、鞋 2 双，每次各穿一件，共有多少种不同的搭配？

例题 1.4

（乘法原理的数位问题）用 1, 2, 3, 4 四个数字可以组成多少个**无重复数字的四位数**？**解：**组成四位数要依次确定千位、百位、十位、个位，共 4 步：第 1 步（千位）：4 种；第 2 步（百位）：剩 3 种；第 3 步（十位）：剩 2 种；第 4 步（个位）：剩 1 种。

$$N = 4 \times 3 \times 2 \times 1 = 24$$

答：可以组成 **24** 个无重复数字的四位数。（这正是后面要学的排列数 $A_4^4 = 4! = 24$ ——乘法原理是排列公式的源头。）**Try It 1.4:** 若数字**可以重复使用**，用 1, 2, 3 组成三位

数，共有多少个？（提示：每一位都有 3 种选择，与前面是否用过无关）

Q&A

- **问：**乘法原理里“第 2 步的选法数会随第 1 步变化”时，为什么有时还能乘？

答：关键看**数量**变不变。比如 4 人抽签、两人依次抽：无论第 1 人抽走哪支，第 2 人可抽的都是 3 支——数量恒定，可以乘 4×3 。但若第 2 步的选法**数量**随第 1 步变化（比如第 2 步的可选项取决于第 1 步选了哪个），就必须分类讨论，不能直接乘。

- **问：**分类加法与分步乘法会不会“混着用”？

答：会，而且大量题目是混用的。复杂问题通常“先分类、每类内部再分步”，或“先分步、某一步内部再分类”。1.1.3 就专门练这个。

1.1.3 两个原理的辨别与综合：先问“完成了吗”

How To：分类还是分步——判断流程

- 设身处地，“完成一件事”；
- 选完这个项，**事情完成了吗**？
- 完成了 → 这是“一类”，与其他类**相加**；
- 没完成，还要选 → 这是“一步”，与其他步**相乘**；
- 复杂问题常常“先分类、再分步”，或“先分步、每步内再分类”。

例题 1.5

（分类与分步的综合）书架上有 3 本不同的语文书、4 本不同的数学书、5 本不同的英语书。（1）从中任取 1 本，有多少种取法？（2）从中各取 1 本语文书、1 本数学书，有多少种取法？（3）从三类书中各取 1 本，有多少种取法？**解：**（1）取 1 本就完成 → 分类加法： $3 + 4 + 5 = 12$ 种。（2）要取语文书**且**数学书，两步都完成才算 → 分步乘法： $3 \times 4 = 12$ 种。（3）三步： $3 \times 4 \times 5 = 60$ 种。（注意（1）（2）答案碰巧都是 12，但一个是加、一个是乘——判断标准不是“结果像不像”，而是“完成一件事需要几步”。）**Try It 1.5：**从 A 地到 C 地：可直接乘大巴（有 3 班），也可乘高铁先到 B 地再转车（A→B 有 2 趟，B→C 有 3 趟）。从 A 到 C 共有多少种走法？（提示：先分类——“直达”与“经 B”，再在“经 B”这一类里分步）

例题 1.6

(集合子集个数) 集合 $A = \{a, b, c\}$ 共有多少个子集? **解:** 每个元素 a, b, c 都有两种“命运”: 在子集中 / 不在子集中。确定一个子集 = 依次决定 3 个元素的命运, 共 3 步, 每步 2 种:

$$N = 2 \times 2 \times 2 = 8$$

答: A 共有 8 个子集 (含空集与它本身)。(推广: n 元集合共有 2^n 个子集——这是分步乘法的著名应用。) **Try It 1.6:** 集合 $\{1, 2, 3, 4, 5\}$ 共有多少个子集?

例题 1.7

(涂色问题) 用红、黄、蓝、绿 4 种颜色给 A、B、C、D 四个区域涂色 (区域排成一列), 要求相邻区域颜色不同, 共有多少种不同的涂法? **解:** 按从左到右依次给四个区域涂色, 分 4 步: A: 4 种; B: 不能与 A 同色, 3 种; C: 不能与 B 同色, 3 种; D: 不能与 C 同色, 3 种。

$$N = 4 \times 3 \times 3 \times 3 = 108$$

答: 共有 108 种涂法。(相邻约束只影响“下一步”, 每一步的选法数恒定, 因此可以连乘。) **Try It 1.7:** 把上题改为“A、B、C 三个区域排成一列, 用 3 种颜色涂, 相邻不同色”, 共有多少种涂法? **想一想 (对抗 M1):** 从 5 名同学中选 2 名分别担任班长和副班长。小明这样算: “先选班长 (5 种), 再选副班长 (4 种), 共 $5 \times 4 = 20$ 种。” 小红说: “应该先选 2 个人 ($C_5^2 = 10$ 种), 再任命正副班长。” 两人谁对? 为什么“选 2 人”这一步本身不是“完成”? (提示: 任命正副班长需要两步——先选出 2 人这一步只完成了“候选”, 还没完成“谁正谁副”。 $10 \times 2 = 20$, 殊途同归。) **本节练习** (答案见章末)

- 某同学去书店买一本书: 书店有 4 种语文辅导书、3 种数学辅导书、2 种英语辅导书, 任意买 1 本, 共有多少种选择?
- 某餐厅提供 3 种主食、4 种荤菜、2 种素菜, 选“1 种主食 + 1 种荤菜 + 1 种素菜”为一餐, 共有多少种搭配?
- 用 0, 1, 2, 3, 4 五个数字组成无重复数字的五位数, 共有多少个? (注意: 0 不能做首位)
- 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?
- 集合 A 共有 $2^7 = 128$ 个子集, 问集合 A 中有多少个元素?

1.2 排列: 有座位的挑人

学习目标

- 理解排列与排列数的概念，掌握排列数公式 $A_n^m = n(n-1)\cdots(n-m+1)$ ；
- 会用排列数计算全排列与选排列；
- 掌握特殊元素优先、捆绑法、插空法、定序问题等常用方法。

前置复习

- 计算 $4 \times 3 \times 2$ 与 $5 \times 4 \times 3$ 。
- 用 1, 2, 3, 4 组成无重复数字的四位数，共多少个？（回顾例题 1.4）
- “甲站第 1 位、乙站第 2 位”和“乙站第 1 位、甲站第 2 位”是同一种站法吗？

1.2.1 排列与排列数：顺序算数的“挑人”

从生活到数学

5 名选手争夺冠亚季军（3 个不同的名次）：第 1 名有 5 种人选，第 2 名剩 4 种，第 3 名剩 3 种，共 $5 \times 4 \times 3 = 60$ 种结果。这里的“结果”是一串有顺序的名字——“甲冠、乙亚、丙季”和“乙冠、甲亚、丙季”是两个不同的结果。

排列与排列数

一般地，从 n 个不同元素中取出 m ($m \leq n$) 个元素，并按照一定的顺序排成一行，叫做从 n 个不同元素中取出 m 个元素的一个排列。所有不同排列的个数，叫做从 n 个不同元素中取出 m 个元素的排列数，记作 A_n^m ：

$$A_n^m = n(n-1)(n-2)\cdots(n-m+1) = \frac{n!}{(n-m)!}$$

当 $m = n$ 时， $A_n^n = n!$ ，称为 n 个元素的全排列（规定 $0! = 1$ ）。

How To：用排列数公式

- 确认元素是否互不相同、是否要排顺序；
- 确定 n （元素总数）与 m （取出的个数）；
- 从 n 开始连乘 m 个数： $n, n-1, \dots, n-m+1$ ；
- 结果就是“依次选位”的乘法原理展开。

例题 1.8

（排列数计算）计算：(1) A_5^3 ；(2) A_5^2 ；(3) A_4^4 ；(4) A_6^2 。解：从 n 开始连乘 m 个：(1) $A_5^3 = 5 \times 4 \times 3 = 60$ ；(2) $A_5^2 = 5 \times 4 = 20$ ；(3) $A_4^4 = 4! = 4 \times 3 \times 2 \times 1 = 24$ ；(4) $A_6^2 = 6 \times 5 = 30$ 。Try It 1.8：计算 A_7^3 、 A_8^2 、 A_3^3 。

例题 1.9

(用排列数解决排队问题) 6 名同学站成一排照相, 共有多少种不同的站法? 若其中甲必须站在最左端, 又有多少种站法? **解:** (1) 6 人全排: $A_6^6 = 6! = 720$ 种。(2) 甲固定在最左端, 其余 5 人任意排: $A_5^5 = 5! = 120$ 种。**答:** (1) **720** 种; (2) **120** 种。**Try It 1.9:** 5 名同学排成一排, 其中某两名同学必须相邻, 有多少种排法? (提示: 先把两人“捆”成 1 块——这是下一小节的方法, 先试试)

Q&A

- **问:** A_n^m 和 n^m 有什么区别?

答: A_n^m 是**不重复**地取 m 个并排序 (如从 10 人里选 3 人排值班表, 同一个人不能出现两次); n^m 是**可以重复**地选 (如 4 位数字密码, 每一位从 0 9 任选, 共 10^4 种)。判断标尺: 同一个元素能被用两次吗?

- **问:** 排列数 A_n^m 一定是从 1 开始乘吗?

答: 不是, 是从 n 开始**往前**乘 m 个数 ($n, n-1, \dots$)。 A_5^3 是 $5 \times 4 \times 3$, 不是 $1 \times 2 \times 3$ 。

1.2.2 有限制条件的排列问题: 四件常用“武器”

从生活到数学

全排列好算, 难的是“有限制”: 甲必须站中间、乙丙必须挨着、丁戊不能挨着、甲乙丙顺序必须固定……约束一个个加进来, 硬排就乱了。下面四件“武器”专门对付四类约束。

武器一: 特殊元素优先法

How To: 先处理受限制的元素, 再排其余不受限的元素。

例题 1.10

(特殊元素优先, 对抗 M9) 6 名同学站成一排, 甲**既不在最左也不在最右**, 共有多少种站法? **解:** 甲受限制, 先安排甲: 第 1 步: 甲站中间 4 个位置之一: 4 种; 第 2 步: 其余 5 人任意站剩余 5 个位置: $A_5^5 = 120$ 种。

$$N = 4 \times 120 = 480$$

答: 共有 **480** 种站法。(检验: 无限制 720 种; 甲在最左 120 种, 甲在最右 120 种, $720 - 240 = 480$ ✓) **Try It 1.10:** 5 名同学站成一排, 甲不能站排头, 共有多少种站法? **武器二: 捆绑法** (处理“必须相邻”) **How To:** 把必须相邻的元素捆成“一大块”, 与其余元素一起排, 再乘块内的排列数。

例题 1.11

(捆绑法) 3 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? **解**: 把 2 名女生捆成 1 块: 第 1 步: 把“女生块”与 3 名男生共 4 个元素排列: $A_4^4 = 24$ 种; 第 2 步: 块内 2 名女生交换顺序: $A_2^2 = 2$ 种。

$$N = 24 \times 2 = 48$$

答: 共有 48 种排法。Try It 1.11: 4 名男生、2 名女生站成一排, 2 名女生相邻, 有多少种排法? **武器三: 插空法** (处理“必须不相邻”) **How To**: 先排其余元素, 再把不能相邻的元素插进空隙 (含两端)。

例题 1.12

(插空法) 3 名男生、2 名女生站成一排, 要求 2 名女生**不相邻**, 共有多少种排法? **解**: 先排 3 名男生: $A_3^3 = 6$ 种, 形成 4 个空 (含两端); 再把 2 名女生插进这 4 个空中的 2 个: $A_4^2 = 4 \times 3 = 12$ 种。

$$N = 6 \times 12 = 72$$

答: 共有 72 种排法。(对照: 无限制 $5! = 120$ 种; 相邻 48 种、不相邻 72 种, $48 + 72 = 120$ ✓——“相邻”与“不相邻”恰好把全部情况一分为二, 这也是“不重不漏”分类的一次实战。) Try It 1.12: 3 名男生、3 名女生站成一排, 3 名女生不相邻, 有多少种排法? **武器四: 定序问题** (部分元素顺序固定) **How To**: n 个不同元素全排, 其中 k 个元素的顺序固定 $\rightarrow \frac{n!}{k!}$ 。

例题 1.13

(定序问题) 6 名同学站成一排, 其中甲、乙、丙三人的相对顺序**固定** (比如规定甲在乙左边、乙在丙左边, 三人之间不必相邻), 共有多少种站法? **解**: 先不考虑限制: $A_6^6 = 720$ 种。在任一排列中, 甲、乙、丙三人有 $3! = 6$ 种相对顺序, 其中只有 1 种符合“甲在乙左、乙在丙左”, 故

$$N = \frac{720}{3!} = \frac{720}{6} = 120$$

答: 共有 120 种站法。Try It 1.13: 7 名同学站成一排, 甲、乙两人顺序固定 (甲在乙左边), 有多少种站法? **想一想 (对抗 M5)**: A、B、C、D、E 5 人排队, 其中 A、B、C 三人**必须相邻**, 但 A、B 两人又**不能挨在一起**。小刚说: “既然必须相邻, 就用捆绑法, 把三人捆成一块, 块内 $3!$ 种。”这个做法错在哪? (提示: 块内三人任意排, A、B 会挨着吗? 要同时满足“A、B、C 相邻”和“A、B 不相邻”, 三人块内只能是谁夹在中间?)——算一算, 正确结果是多少? (见章末答案)

Q&A

- 问: 定序问题为什么能用“除以 $k!$ ”?

答：全排列里，那 k 个元素的 $k!$ 种相对顺序等可能地出现，每种顺序对应同样多的排列；固定其中一种，就留下全部排列的 $\frac{1}{k!}$ 。

- 问：捆绑法和插空法能同时用吗？

答：能。多个限制条件依次处理即可，但每一步都要确认“被处理过的条件不会在后面的步骤里被破坏”——详见本节练习第 4 题与章末的复合限制题。

本节练习（答案见章末）

- 计算： A_5^4 、 A_7^2 、 A_4^2 。
- 7 名同学站成一排，男生甲不能站排头，共有多少种排法？
- 3 名同学站成一排，其中甲、乙两名同学必须相邻，共有多少种排法？
- 5 名同学站成一排，甲、乙两人不相邻，共有多少种排法？
- 5 名同学站成一排，甲、乙两人顺序固定（甲在乙前），共有多少种排法？

1.3 组合：只管打包

学习目标

- 理解组合与组合数的概念，掌握组合数公式 $C_n^m = \frac{n!}{m!(n-m)!}$ ；
- 会区分排列与组合，能根据问题正确选择模型；
- 掌握组合数的两个重要性质： $C_n^m = C_n^{n-m}$ 与 $C_n^m + C_n^{m+1} = C_{n+1}^{m+1}$ ；
- 会解决选人组队、分组分配、几何计数等典型问题。

前置复习

- $A_5^3 = ?$
- 从 5 人中选 3 人排成一排，有多少种？从 5 人中选 3 人组成一组，有多少种？
- 计算 $\frac{4 \times 3}{2 \times 1}$ 。

1.3.1 组合与组合数：顺序不算数的“挑人”

从生活到数学

同样从 5 人里选 3 人：

- 排成名次（冠、亚、季军）： $A_5^3 = 60$ 种；

- 组成一支队伍：每 3 人小组内部有 $3! = 6$ 种排列，在“队伍”视角下全是同一种。

于是队伍数 $= 60 \div 6 = 10$ 。这就是组合与排列的关系：**先按排列数，再把组内的顺序抹掉。**

组合与组合数

一般地，从 n 个**不同元素**中取出 m ($m \leq n$) 个元素**组成一组**，**不计顺序**，叫做从 n 个不同元素中取出 m 个元素的一个**组合**。所有不同组合的个数，叫做从 n 个不同元素中取出 m 个元素的**组合数**，记作 C_n^m ：

$$C_n^m = \frac{A_n^m}{m!} = \frac{n!}{m!(n-m)!}$$

组合数的性质

(1) 对称性： $C_n^m = C_n^{n-m}$ ；(2) 杨辉恒等式： $C_n^m + C_n^{m+1} = C_{n+1}^{m+1}$ 。

例题 1.14

(组合数计算与性质) 计算：(1) C_5^2 ；(2) C_8^3 ；(3) 利用性质化简并计算 C_{10}^7 。**解：**(1) $C_5^2 = \frac{5 \times 4}{2!} = 10$ ；(2) $C_8^3 = \frac{8 \times 7 \times 6}{3!} = \frac{336}{6} = 56$ ；(3) 由对称性 $C_{10}^7 = C_{10}^3 = \frac{10 \times 9 \times 8}{3!} = 120$ 。**Try It 1.14：**计算 C_6^2 、 C_9^2 、 C_{10}^3 。

例题 1.15

(排列与组合的辨析，对抗 M4) (1) 从 5 名同学中选 3 名分别担任班长、学习委员、体育委员，有多少种选法？(2) 从 5 名同学中选 3 名参加数学竞赛，有多少种选法？**解：**(1) 3 个职务**不同**，谁当什么职务有区别 $\rightarrow$ 排列： $A_5^3 = 60$ 种。(2) 3 人参加同一个比赛，**先念到谁无所谓** $\rightarrow$ 组合： $C_5^3 = \frac{5 \times 4 \times 3}{3!} = 10$ 种。(同一个“选 3 人”，一个 60、一个 10，差的正是职务带来的顺序。**判定标尺：换位置后结果变不变。**) **Try It 1.15：**从 6 名同学中选 4 名：(1) 分别担任四个不同的岗位；(2) 组成一个小组。各有多少种选法？**想一想 (对抗 M2)：**从 5 名同学中选 2 名当代表参加比赛。小明说：“先选 1 人 (5 种)，再选 1 人 (4 种)，共 $5 \times 4 = 20$ 种。”小华说：“ $C_5^2 = 10$ 种。”谁对？小明错在哪？(提示：代表不分先后——“先选甲再选乙”和“先选乙再选甲”是同一种选择，分步把它数成了 2 次，重复了一倍。)

Q&A

- **问：**为什么规定 $C_n^0 = 1$ 、 $C_n^n = 1$ ？

答：从 n 个元素中“选 0 个” (空组) 只有 1 种；“全选”也只有 1 种。这与对称性 $C_n^0 = C_n^n$ 一致。

- **问：** C_n^m 一定是整数吗？

答：一定。它数的是“组合的个数”，个数当然是整数；公式算出的分数约分后必为整数。

1.3.2 组合的应用：组队、分组、几何计数

从生活到数学

组合数一旦配上条件，就变成概率题里最常用的“数格子”：8 个人选 3 个，含特定的人有几种？6 个人分成两组，要不要除以 2？十边形有多少条对角线？这一小节把组合的应用练到“闭眼会分”。

例题 1.16

(组队选人，含特殊人) 从 8 名同学中选 3 名参加辩论赛。(1) 共有多少种选法？(2) 必须包含队长甲，有多少种选法？(3) 不包含甲，有多少种选法？验证 (2) (3) 之和等于 (1)。**解：**(1) $C_8^3 = \frac{8 \times 7 \times 6}{3!} = 56$ 种。(2) 甲必选，再从其余 7 人中选 2 人： $C_7^2 = 21$ 种。(3) 甲不选，从其余 7 人中选 3 人： $C_7^3 = 35$ 种。验证： $21 + 35 = 56 = C_8^3$ ✓——“含甲”与“不含甲”是**不重不漏**的分类，这正是分类加法原理在组合中的应用。**Try It 1.16：**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？(提示：正面分类有点多，试试“总数 - 全是男生”的补集)

例题 1.17

(分组分配问题，对抗 M6/M10) (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？(3) 5 名学生分成三组，人数分别为 2、2、1，共有多少种分法？**解：**(1) **组间无区别**（三组都叫“组”，没有名字）：先分组、再除“组间顺序”：

$$N = \frac{C_6^2 C_4^2 C_2^2}{3!} = \frac{15 \times 6 \times 1}{6} = 15$$

(2) **组有名字**（甲、乙、丙不同）：分步乘即可，无需除重：

$$N = C_6^2 C_4^2 C_2^2 = 15 \times 6 \times 1 = 90$$

(3) 两组 2 人、一组 1 人，只有“2 人组”彼此无区别，**只除 2!**：

$$N = \frac{C_5^2 C_3^2 C_1^1}{2!} = \frac{10 \times 3 \times 1}{2} = 15$$

答：(1) 15 种；(2) 90 种；(3) 15 种。(对照 (1) (2)：同样的“每组 2 人”，给三组贴上“甲、乙、丙”的标签，就多出 $3! = 6$ 倍。“均匀分组”不除重、或“有名字的组”误除重，是这片最常见的坑。) **Try It 1.17：**4 名学生分成两组，每组 2 人，共有多少种分法？若分给两个不同的班级呢？

例题 1.18

(几何计数) (1) 平面上有 10 个点, 任意三点不共线, 共可确定多少条直线? 多少个三角形? (2) 正六边形的对角线共有多少条? **解:** (1) 两点确定一条直线: $C_{10}^2 = 45$ 条; 三点确定一个三角形: $C_{10}^3 = 120$ 个。 (2) 任意两顶点连线 $C_6^2 = 15$ 条, 其中 6 条是边, 其余为对角线:

$$N = C_6^2 - 6 = 15 - 6 = 9$$

答: (1) 45 条、120 个; (2) 9 条。(一般地, 正 n 边形有 $C_n^2 - n$ 条对角线——“总数减去边数”。) **Try It 1.18:** 正八边形的对角线有多少条? **想一想 (对抗 M6):** 5 名同学分别去三个不同的场馆做志愿者, 每馆至少 1 人, 有多少种方案? 小明说: “先分成三组 ($\frac{C_5^2 C_3^2 C_1^1}{2!} = 15$), 再分给三个场馆 ($3! = 6$), 共 90 种。” 小红说: “直接每人选场馆 $3^5 = 243$, 再减去有空馆的。” 谁对? (提示: 分组时 “2、2、1” 已除过 $2!$, 分配时三组有名字再乘 $3!$ ——小明的 90 是对的; 小红用容斥也能算, 但小明的 “先除重、后贴标签” 更直接。) **本节练习** (答案见章末)

- 计算: C_7^3 、 C_9^2 、 C_6^4 。
- 从 10 名同学中选 4 名参加知识竞赛: (1) 共有多少种选法? (2) 必须包含指定的甲? (3) 甲、乙两人至多选一人?
- 6 名学生分成两组, 每组 3 人, 共有多少种分法?
- 正七边形有多少条对角线?
- 从 4 名男生、3 名女生中选 3 人, 要求男、女生都有, 共有多少种选法?

1.4 二项式定理: 组合数的 “展开舞台”

学习目标

- 理解并掌握二项式定理 $(a + b)^n = \sum_{r=0}^n C_n^r a^{n-r} b^r$;
- 掌握通项公式 $T_{r+1} = C_n^r a^{n-r} b^r$, 会求指定项与特定系数;
- 掌握二项式系数的性质: 对称性、增减性、各系数和 2^n ;
- 会用赋值法求系数和。

前置复习

- 展开 $(a + b)^2$ 与 $(a + b)^3$, 数一数各项的系数。
- 计算 $C_4^0 + C_4^1 + C_4^2 + C_4^3 + C_4^4$ 。

- C_n^m 表示什么?

1.4.1 二项式定理与通项公式：每个系数都是“选位置”

从生活到数学

$(a+b)^3 = (a+b)(a+b)(a+b)$ 。展开式的每一项，是从**每个括号里各取一项**相乘的结果。比如项 a^2b ：3 个括号里，有 2 个取 a 、1 个取 b 。取 b 的那个括号可以是第 1 个、第 2 个或第 3 个——有 $C_3^1 = 3$ 种“**选位置**”的方式，所以 a^2b 的系数是 3。所有系数，都是这么“数”出来的。

“选位置”示意 ($(a+b)^3$ 中项 a^2b 的三种来源)

第 1 个括号取 b ： $b \cdot a \cdot a$ 第 2 个括号取 b ： $a \cdot b \cdot a$ 第 3 个括号取 b ： $a \cdot a \cdot b$

从 3 个括号中选 1 个“取 b ”的位置，共 $C_3^1 = 3$ 种——项 a^2b 的系数就是 3。

二项式定理

$$(a+b)^n = C_n^0 a^n + C_n^1 a^{n-1}b + C_n^2 a^{n-2}b^2 + \cdots + C_n^n b^n = \sum_{r=0}^n C_n^r a^{n-r} b^r$$

其中 C_n^r ($r = 0, 1, \dots, n$) 叫做**二项式系数**。展开式中第 $r+1$ 项为

$$T_{r+1} = C_n^r a^{n-r} b^r \quad (r = 0, 1, \dots, n)$$

叫做二项展开式的**通项**。

How To：求指定项 / 指定系数

- 写出通项 $T_{r+1} = C_n^r a^{n-r} b^r$ (把 a 、 b 的式子原样代入)；
- 合并同底数幂，写出 x 的指数表达式；
- 令指数等于目标值，解出 r ；
- 代回通项，计算系数（别忘了系数里的常数因子）。

例题 1.19

(求指定项) 求 $(1+x)^6$ 的展开式，并求展开式中 x^3 的系数。**解**：由二项式定理，

$$(1+x)^6 = \sum_{r=0}^6 C_6^r x^r = 1 + 6x + 15x^2 + 20x^3 + 15x^4 + 6x^5 + x^6$$

(系数 1, 6, 15, 20, 15, 6, 1 正是杨辉三角第 7 行。) 其中 x^3 的系数为 $C_6^3 = 20$ 。**Try It 1.19**：求 $(1+x)^5$ 的展开式中 x^2 的系数。

例题 1.20

(通项求指定项, 对抗 M7) 求 $\left(2x - \frac{1}{x}\right)^6$ 展开式中的常数项。解: 通项 $T_{r+1} = C_6^r (2x)^{6-r} \left(-\frac{1}{x}\right)^r = C_6^r \cdot 2^{6-r} \cdot (-1)^r \cdot x^{6-r} \cdot x^{-r} = C_6^r \cdot 2^{6-r} \cdot (-1)^r \cdot x^{6-2r}$ 。常数项要求 x 的指数为 0: $6 - 2r = 0$, 得 $r = 3$ 。

$$T_4 = C_6^3 \cdot 2^3 \cdot (-1)^3 = 20 \times 8 \times (-1) = -160$$

答: 常数项为 -160 。(注意: 常数项是第 4 项——用 $r = 3$, 不是 $r = 4$, 正是通项下标陷阱。) Try It 1.20: 求 $\left(x + \frac{1}{x}\right)^6$ 展开式中的常数项。

例题 1.21

(求特定系数, 对抗 M8) 求 $(2x + 3y)^6$ 展开式中 $x^3 y^3$ 的系数。解: 通项 $T_{r+1} = C_6^r (2x)^{6-r} (3y)^r = C_6^r \cdot 2^{6-r} \cdot 3^r \cdot x^{6-r} y^r$ 。令 x, y 的指数相等且均为 3: $6 - r = 3$ 且 $r = 3$, 得 $r = 3$ 。系数 $= C_6^3 \cdot 2^3 \cdot 3^3 = 20 \times 8 \times 27 = 4320$ 。答: $x^3 y^3$ 的系数为 4320。(若只写 $C_6^3 = 20$, 就漏掉了 $2^3 \cdot 3^3$ 这两个“藏在底数里的系数”。) Try It 1.21: 求 $(1 - 2x)^7$ 展开式中 x^3 的系数。想一想: $(a + b)^n$ 展开式一共有 $n + 1$ 项还是 n 项? 第 1 项对应 r 等于几? (提示: r 从 0 到 n , 共 $n + 1$ 项; 第 1 项 $r = 0$ 。)

1.4.2 二项式系数的性质: 杨辉三角与赋值法

从生活到数学

把 $(a + b)^n$ 的系数按 $n = 0, 1, 2, \dots$ 排成三角形, 就是杨辉三角: 每行两端是 1, 中间的每个数等于它“左肩 + 右肩”之和——这正是恒等式 $C_n^m + C_n^{m+1} = C_{n+1}^{m+1}$ 。这个三角藏着二项式系数的全部脾气。

杨辉三角前 7 行 (每行两端为 1, 中间数 = 左肩 + 右肩)

$n=0$: 1 $n=1$: 1 1 $n=2$: 1 2 1 $n=3$: 1 3 3 1 $n=4$: 1 4 6 4 1 $n=5$: 1 5 10 10 5 1 $n=6$: 1 6 15 20 15 6 1

从第 3 行起, 中间的每个数都等于它上一行“左肩 + 右肩”之和, 例如 $10 = 4 + 6$ 、 $15 = 5 + 10$ ——这正是组合恒等式 $C_n^m + C_n^{m+1} = C_{n+1}^{m+1}$ 的直观来源。

二项式系数的性质

(1) 对称性: $C_n^r = C_n^{n-r}$ (杨辉三角左右对称); (2) 增减与最大: 二项式系数先增后减; n 为偶数时, 中间一项 $C_n^{n/2}$ 最大; n 为奇数时, 中间两项 $C_n^{(n-1)/2}$ 、 $C_n^{(n+1)/2}$ 并列最大; (3) 各二项式系数之和: $C_n^0 + C_n^1 + \dots + C_n^n = 2^n$; (4) 奇数项系数和 = 偶数项系数和 = 2^{n-1} (由赋值法得出)。

性质 (3) 的证明: 在 $(1 + x)^n = \sum_{r=0}^n C_n^r x^r$ 中令 $x = 1$, 得

$$2^n = (1 + 1)^n = C_n^0 + C_n^1 + \dots + C_n^n$$

故各二项式系数之和为 2^n 。■

例题 1.22

(系数之和, 赋值法) (1) 求 $(1+2x)^7$ 展开式中各项的**系数和**; (2) 求 $(1+2x)^7$ 展开式中各项的**二项式系数和**。解: (1) 令 $x=1$: $(1+2)^7 = 3^7 = 2187$, 即各项系数和为 2187。(2) 二项式系数和与底数里的 2 无关: $2^7 = 128$ 。(对比: 一个 2187、一个 128——“系数和”里要令 $x=1$, “二项式系数和”固定是 2^n 。) Try It 1.22: 求 $(1-3x)^6$ 展开式中各项系数和。

例题 1.23

(最大项, 对抗 M12) (1) $(1+x)^8$ 展开式中二项式系数最大的项是什么? (2) $(1+2x)^8$ 展开式中**系数**最大的项是什么? 解: (1) $n=8$ 为偶数, 中间一项二项式系数最大: $T_5 = C_8^4 x^4 = 70x^4$ 。(2) 设第 $r+1$ 项系数为 $a_r = C_8^r \cdot 2^r$, 考察相邻项之比:

$$\frac{a_{r+1}}{a_r} = \frac{C_8^{r+1} 2^{r+1}}{C_8^r 2^r} = \frac{8-r}{r+1} \cdot 2$$

当 $\frac{8-r}{r+1} \cdot 2 > 1$, 即 $r < 5$ 时系数递增; $r \geq 5$ 时递减。比较 $r=4$ 与 $r=5$: $a_4 = C_8^4 \cdot 2^4 = 70 \times 16 = 1120$, $a_5 = C_8^5 \cdot 2^5 = 56 \times 32 = 1792$ 。故系数最大的项为 $a_5 = 1792$, 即 $T_6 = 1792x^5$ 。(同样是 $n=8$: 二项式系数最大在 T_5 , 系数最大却在 T_6 ——“**二项式系数**”与“**系数**”是两码事, 别混。) Try It 1.23: $(1+x)^9$ 展开式中二项式系数最大的项有几项? 想一想 (对抗 M12): $(1+2x)^n$ 展开式中系数最大的项一定在正中间吗? 观察例题 1.23 (2): $n=8$ 时最大系数落在 T_6 , 不在正中间的 T_5 ——因为 2^r 把峰值“往右推”了。若要确定 $(1+kx)^n$ 的系数最大项, 最稳妥的做法是什么? (提示: 比较相邻项系数比 $\frac{a_{r+1}}{a_r}$ 与 1 的大小, 找到峰值。)

Q&A

• 问: $(a+b)^n$ 中“二项式系数”与“某项的系数”到底怎么区分?

答: 二项式系数是 C_n^r , 由 n, r 决定, 与 a, b 具体是什么无关; 某项的系数是展开后该项完整的数字系数, 可能含 a, b 里自带的常数因子。求“系数最大项”时讨论的是后者。

• 问: 赋值法求系数和时, 什么情况下要取 $x=-1$?

答: 求“奇数项系数和 = 偶数项系数和”或它们的差时取 $x=-1$ ——因为 $(-1)^r$ 正负交替, 正好把奇偶项分开。

本节练习 (答案见章末)

• 求 $(1+x)^7$ 展开式中 x^2 的系数。

- 求 $\left(x - \frac{2}{x}\right)^5$ 展开式的常数项（如果存在的话）。
- 求 $\left(x^2 + \frac{1}{x}\right)^9$ 展开式中 x^9 的系数。
- 求 $(1-x)^8$ 展开式中各项系数和。
- 求 $(1+x)^{10}$ 展开式中二项式系数最大的项。

本章回顾

概念地图

· 计数原理 · 分类加法计数原理： $N = m_1 + m_2 + \cdots + m_k$ （“要么…要么…”，分类要**不重不漏**） · 分步乘法计数原理： $N = m_1 \times m_2 \times \cdots \times m_k$ （“先…再…”，分步要**层层独立**） · 判定标尺：选完这项，事情完成了吗？完成 $\rightarrow$ 加；未完成 $\rightarrow$ 乘 · 排列 $A_n^m = \frac{n!}{(n-m)!}$ ：有顺序的挑人（换位置结果变） · 组合 $C_n^m = \frac{n!}{m!(n-m)!}$ ：只打包不排序（换位置不变） · 组合 = 排列 $\div$ 组内全排列： $C_n^m = \frac{A_n^m}{m!}$ · 组合数性质： $C_n^m = C_n^{n-m}$ ； $C_n^m + C_n^{m+1} = C_{n+1}^{m+1}$ · 常用方法：特殊元素优先 · 捆绑（相邻） · 插空（不相邻） · 定序除以 $k!$ · 均匀分组除重 · 二项式定理： $(a+b)^n = \sum_{r=0}^n C_n^r a^{n-r} b^r$ ，通项 $T_{r+1} = C_n^r a^{n-r} b^r$ · 二项式系数性质：对称 · 中间最大（ n 奇偶分论） · 系数和 = 2^n · 奇数位和 = 偶数位和 = 2^{n-1}

关键公式速查

内容	公式
分类加法	$N = m_1 + m_2 + \cdots + m_k$
分步乘法	$N = m_1 \times m_2 \times \cdots \times m_k$
排列数	$A_n^m = n(n-1)\cdots(n-m+1) = \frac{n!}{(n-m)!}$
全排列	$A_n^n = n!$ ，规定 $0! = 1$
组合数	$C_n^m = \frac{n!}{m!(n-m)!}$
组合性质	$C_n^m = C_n^{n-m}$ ， $C_n^m + C_n^{m+1} = C_{n+1}^{m+1}$
子集个数	n 元集合的子集有 2^n 个
二项式定理	$(a+b)^n = \sum_{r=0}^n C_n^r a^{n-r} b^r$
通项	$T_{r+1} = C_n^r a^{n-r} b^r$ （下标 $r+1$ ）
系数和	各项系数和：令 $x=1$ ；二项式系数和： 2^n

分层练习

A 基础层

- 书架上有 5 本不同的数学书、6 本不同的语文书。任取 1 本，有多少种取法？各取 1 本呢？
- 计算： A_6^3 、 C_7^4 、 C_{10}^2 。

- 用 1, 2, 3, 4 四个数字可以组成多少个无重复数字的四位数? 其中末位是 4 的有多少个?
- 4 名同学站成一排, 甲必须站排头, 有多少种站法?
- 从 6 名同学中选 2 名参加比赛, 有多少种选法?
- 求 $(1+x)^5$ 的展开式中 x^3 的系数。
- 求 $(2-x)^5$ 展开式中各项系数和。

B 应用层

- 5 名男生、4 名女生站成一排, 4 名女生必须相邻, 有多少种排法?
- 从 6 名男生、5 名女生中选 5 人, 要求至少 3 名女生, 有多少种选法?
- 求 $\left(x^2 + \frac{1}{x}\right)^6$ 展开式的常数项。
- 用 0, 1, 2, 3, 4 组成无重复数字的五位数, 其中偶数有多少个? (提示: 末位为 0、2、4 分三类, 且首位不能为 0)

C 拓展层

- 6 名学生分成三组, 一组 1 人、一组 2 人、一组 3 人, 共有多少种分法? 若分别去三个不同的社区呢?
- 求 $(1+x)(1+2x)^6$ 展开式中 x^3 的系数。

参考答案 (章末集中)

Try It 即时练习答案

- 1.1: $4+3+2=9$ 种。
- 1.2: 含 1 的有 $C_4^1=4$ 种, 不含 1 的有 $C_4^2=6$ 种, $4+6=10=C_5^2$, 不重不漏 ✓。
- 1.3: $3 \times 4 \times 2=24$ 种。
- 1.4: 每一位 3 种, $3 \times 3 \times 3=27$ 个。
- 1.5: 分类: “直达” 3 种, “经 B” $2 \times 3=6$ 种, 共 9 种。
- 1.6: $2^5=32$ 个。
- 1.7: $3 \times 2 \times 2=12$ 种。
- 1.8: $A_7^3=7 \times 6 \times 5=210$; $A_8^2=8 \times 7=56$; $A_3^3=3!=6$ 。

- 1.9: 两人捆成块: 块与其余 3 人共 4 个元素全排 $4! = 24$, 块内 $2! = 2$, 共 48 种。
- 1.10: 甲不排头: 甲先占非排头的 4 个位置 (4 种), 其余 4 人全排 $4! = 24$, 共 $4 \times 24 = 96$ 种 (或用补集 $5! - 4! = 120 - 24 = 96$)。
- 1.11: 2 女捆块: 块与 4 男共 5 个元素全排 $5! = 120$, 块内 $2! = 2$, 共 240 种。
- 1.12: 先排 3 男 $3! = 6$, 形成 4 空插 3 女 $A_4^3 = 24$, 共 $6 \times 24 = 144$ 种。
- 1.13: $\frac{7!}{2!} = \frac{5040}{2} = 2520$ 种。
- 1.14: $C_6^2 = 15$, $C_9^2 = 36$, $C_{10}^3 = 120$ 。
- 1.15: (1) $A_6^4 = 6 \times 5 \times 4 \times 3 = 360$; (2) $C_6^4 = C_6^2 = 15$ 。
- 1.16: 补集: 总数 $C_{10}^3 = 120$, 全男生 $C_6^3 = 20$, $120 - 20 = 100$ 种。
- 1.17: $\frac{C_4^2 C_2^2}{2!} = \frac{6 \times 1}{2} = 3$ 种; 分给两个班级: $C_4^2 C_2^2 = 6$ 种。
- 1.18: $C_8^2 - 8 = 28 - 8 = 20$ 条。
- 1.19: $C_5^2 = 10$ 。
- 1.20: 通项 $C_6^r x^{6-2r}$, 常数项 $6 - 2r = 0 \rightarrow r = 3$, $C_6^3 = 20$ 。
- 1.21: $T_4 = C_7^3 \cdot 1^4 \cdot (-2x)^3 = 35 \times (-8) = -280$ 。
- 1.22: 令 $x = 1$: $(1 - 3)^6 = (-2)^6 = 64$ 。
- 1.23: $n = 9$ 为奇数, 中间两项 C_9^4 、 C_9^5 并列最大, 共 2 项。

各节练习答案

- 1.1 节: 1. $4 + 3 + 2 = 9$ 种。2. $3 \times 4 \times 2 = 24$ 种。3. 首位不能为 0, 有 4 种; 其余 4 位全排 $4! = 24$, 共 $4 \times 24 = 96$ 个。4. 分类: 直达 2 种; 经乙地 $4 \times 3 = 12$ 种; 共 14 种。5. $2^n = 128 = 2^7 \rightarrow n = 7$ 个元素。
- 1.2 节: 1. $A_5^4 = 120$, $A_7^2 = 42$, $A_4^2 = 12$ 。2. 甲先占非排头的 6 个位置 (6 种), 其余 6 人全排 $6! = 720$, 共 $6 \times 720 = 4320$ 种。3. 甲、乙捆块: 块与丙共 2 个元素全排 $2! = 2$, 块内 $2! = 2$, 共 4 种。4. 补集: $5! - 2 \times 4! = 120 - 48 = 72$ 种 (或插空: 先排其余 3 人 $3! = 6$, 4 空插 2 人 $A_4^2 = 12$, 共 72 种)。5. $\frac{5!}{2!} = 60$ 种。
- 1.3 节: 1. $C_7^3 = 35$, $C_9^2 = 36$, $C_6^4 = C_6^2 = 15$ 。2. (1) $C_{10}^4 = 210$; (2) $C_9^3 = 84$; (3) 甲、乙至多选一人 = 都不选 $C_8^4 = 70$, 或恰选 1 人 $C_2^1 C_8^3 = 2 \times 56 = 112$, 共 $70 + 112 = 182$ (或总数减去两人都选: $210 - C_8^2 = 210 - 28 = 182$)。3. $\frac{C_6^3 C_3^3}{2!} = \frac{20 \times 1}{2} = 10$ 种。4. $C_7^2 - 7 = 21 - 7 = 14$ 条。5. 男女生都有 = 总数 $C_7^3 = 35$ 减去全男 $C_4^3 = 4$ 、全女 $C_3^3 = 1$: $35 - 5 = 30$ 种 (或 1 男 2 女 $C_4^1 C_3^2 = 12$, 2 男 1 女 $C_4^2 C_3^1 = 18$, 共 30)。

- **1.4 节:** 1. $C_7^2 = 21$ 。2. 通项 $T_{r+1} = C_5^r x^{5-r} \cdot (-2)^r x^{-r} = C_5^r (-2)^r x^{5-2r}$; 令 $5 - 2r = 0$, $r = 2.5$ 不是整数 $\rightarrow$ **常数项不存在**。3. 通项 $C_9^r x^{18-2r} x^{-r} = C_9^r x^{18-3r}$, 令 $18 - 3r = 9 \rightarrow r = 3$, $C_9^3 = 84$ 。4. 令 $x = 1$: $(1-1)^8 = 0$ 。5. $n = 10$ 为偶数, 中间一项 $C_{10}^5 x^5 = 252x^5$ 。

分层练习答案

A1: 任取 1 本: $5 + 6 = 11$ 种; 各取 1 本: $5 \times 6 = 30$ 种。

A2: $A_6^3 = 6 \times 5 \times 4 = 120$; $C_7^4 = C_7^3 = 35$; $C_{10}^2 = 45$ 。

A3: $4! = 24$ 个; 末位是 4: 其余 3 位全排 $3! = 6$ 个。

A4: 甲固定排头, 其余 3 人全排: $3! = 6$ 种。

A5: $C_6^2 = 15$ 种。

A6: $C_5^3 = 10$ 。

A7: 令 $x = 1$: $(2-1)^5 = 1$ 。

B8: 4 女捆成块, 块与 5 男共 6 个元素全排 $6! = 720$, 块内 $4! = 24$, 共 $720 \times 24 = 17280$ 种。

B9: 至少 3 女分三类: 3 女 2 男 $C_5^3 C_6^2 = 10 \times 15 = 150$; 4 女 1 男 $C_5^4 C_6^1 = 5 \times 6 = 30$; 5 女 $C_5^5 = 1$; 共 $150 + 30 + 1 = 181$ 种。

B10: 通项 $C_6^r x^{12-2r} x^{-r} = C_6^r x^{12-3r}$, 令 $12 - 3r = 0 \rightarrow r = 4$, $C_6^4 = C_6^2 = 15$ 。

B11: 分三类: 末位 0: 其余 4 位全排 $4! = 24$; 末位 2: 首位不能为 0, 有 3 种 (1、3、4), 中间 3 位 $3! = 6$, 共 $3 \times 6 = 18$; 末位 4: 同理 18; 共 $24 + 18 + 18 = 60$ 个。(检验: 总数 $4 \times 4! = 96$, 奇数末位 1 或 3 各 18 个共 36, $96 - 36 = 60$ ✓)

C12: 三组人数 1、2、3 互不相同, 组与组可区分, 无需除重: $C_6^1 C_5^2 C_3^3 = 6 \times 10 \times 1 = 60$ 种; 分给三个不同社区: $60 \times 3! = 360$ 种。

C13: x^3 的系数来自两项之和: $1 \times C_6^3 \cdot 2^3 + 1 \times C_6^2 \cdot 2^2 = 20 \times 8 + 15 \times 4 = 160 + 60 = 220$ 。

想一想 (对抗) 答案

- **想一想 (对抗 M1):** 小明对。先选 2 人 ($C_5^2 = 10$) 这一步只完成了“选候选”, 还没完成“谁正谁副”, 所以 10 不是最后答案; 再接任命 2 种, $10 \times 2 = 20$ 与小明的 $5 \times 4 = 20$ 一致。“选 2 人”不是完成, 必须再分一步。
- **想一想 (对抗 M2):** 小华对。代表不分先后, “先甲后乙”与“先乙后甲”是同一种, 分步把它数成 2 次, 重复一倍: $20 \div 2 = 10 = C_5^2$ 。
- **想一想 (对抗 M3):** 漏掉了“首位为 0”这种不可能的情况——0 不能做首位, 所以首位只能是 1、2、3, 他的 3 类分类其实是**正确的** (没有“首位为 0 的四位数”存在); 真正要防的坑是: 有人把分类写成“首位为 0、1、2、3”四类再直接相加, 那就把不存在的“首位为 0”也数进去了。本题正确做法: 首位 3 种 (1、2、3), 其余 3 位 $3! = 6$, 共 $3 \times 6 = 18$ 个。

- **想一想 (对抗 M5):** 小刚错。块内 $3! = 6$ 种中, 有 4 种让 A、B 挨在一起 (如 A B C、B A C、C A B、C B A), 违背 “A、B 不相邻”。三人相邻且 A、B 不相邻 $\rightarrow$ 只能 C 夹在中间: $A C B$ 或 $B C A$, 块内 2 种; 块与 D、E 共 3 个对象全排 $3! = 6$ 。共 $2 \times 6 = 12$ 种。(若误用 $3!$ 会算出 36, 多出的 24 种都是 A、B 挨着的。)
- **想一想 (对抗 M6):** 小明对。先分组 $\frac{C_5^2 C_3^2 C_1^1}{2!} = 15$ (两组 2 人无区别所以除 $2!$), 再分给 3 个有名字的场馆 $\times 3! = 90$ 。
- **想一想 (对抗 M12):** 不一定在正中间。 2^r 会把系数峰值往右推。最稳妥: 比较相邻项系数比 $\frac{a_{r+1}}{a_r} = \frac{n-r}{r+1} \cdot k$ 与 1 的大小, 找到递增转递减的位置。

第二章 怎么用数字度量“运气”：概率

章首情境

2.1 随机事件与古典概型：给“可能”排队编号

学习目标

- 理解样本空间、样本点、随机事件的概念，会用集合语言表示事件；
- 理解并掌握互斥事件与对立事件的关系及概率性质；
- 掌握古典概型的两个前提与公式 $P(A) = \frac{n(A)}{n(\Omega)}$ ，会求古典概型概率。

前置复习 (Be Prepared)

- 掷一枚骰子，点数“大于 4”的结果有哪几个？
- 用 1, 2, 3, 4 四个数字能组成多少个无重复数字的四位数？
- 袋中有 2 红 3 蓝共 5 个球，任取 2 个球有几种取法？

2.1.1 样本空间与随机事件

从生活到数学

“掷骰子”是一次**随机试验**——结果不止一个，且事先不能确定。把所有可能的结果摆出来：1, 2, 3, 4, 5, 6 点。这 6 个结果组成一个集合，叫**样本空间**；其中每一个结果叫一个**样本点**。“点数大于 4”这个事件，就是样本空间的一个子集 $\{5, 6\}$ 。

样本空间与随机事件

在相同条件下可重复进行、结果具有随机性的试验叫做**随机试验**。把随机试验的每一个可能结果称为**样本点**，全体样本点的集合称为**样本空间**，通常记作 Ω 。样本空间 Ω 的任意子集称为**随机事件**，简称**事件**（事件一般用大写字母 $A, B, C, \dots$ 表示）；只含一个样本点的事件叫**基本事件**。 Ω 本身（必然发生）叫**必然事件**， $\emptyset$ （不可能发生）叫**不可能事件**。

例题 2.1

(样本空间与事件) 同时掷两枚质地均匀的硬币, 写出这个试验的样本空间, 并用集合表示事件 A : “恰好一枚正面”。**解**: 用“正、反”记录每枚硬币, 样本点用有序数对表示:

$$\Omega = \{(\text{正}, \text{正}), (\text{正}, \text{反}), (\text{反}, \text{正}), (\text{反}, \text{反})\}$$

共 4 个样本点。事件 $A = \{(\text{正}, \text{反}), (\text{反}, \text{正})\}$ 。(注意: 两枚硬币是**两枚不同的硬币**, $(\text{正}, \text{反})$ 与 $(\text{反}, \text{正})$ 是两个不同的样本点。) **Try It 2.1**: 掷一颗骰子, 用集合表示事件“点数为奇数”与事件“点数大于 4”, 并写出它们的交集与并集。

Q&A

- **问**: 样本空间和事件都用集合表示, 为什么?

答: 因为集合的交、并、补运算正好对应“同时发生”“至少一个发生”“不发生”——把概率问题翻译成集合问题, 就有一套现成的语言和运算规则。

2.1.2 事件的关系与运算

从生活到数学

事件既然可以看成集合, 集合的交、并、补自然搬到事件上。抽签抽到“1 号”和“2 号”不能同时发生 (互斥); “抽到偶数”和“抽到奇数”互为补集 (对立)。这一小节把事件之间的关系一次理清。

事件的关系与运算

- **并事件**: $A \cup B$ (A 发生或 B 发生);
- **交事件**: $A \cap B$ (A 与 B 同时发生);
- **互斥事件**: $A \cap B = \emptyset$ (不能同时发生);
- **对立事件**: $\bar{A}$ (A 不发生, 即 $\Omega \setminus A$)。

概率的基本性质

(1) $0 \leq P(A) \leq 1$; $P(\Omega) = 1$, $P(\emptyset) = 0$; (2) 若 A 、 B 互斥, 则 $P(A \cup B) = P(A) + P(B)$; (3) 一般地, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$; (4) 对立事件: $P(\bar{A}) = 1 - P(A)$ 。

例题 2.2

(事件关系与概率性质) 掷一颗骰子, 设 $A =$ “点数为偶数”, $B =$ “点数大于 3”。(1) 写出 $A \cup B$ 与 $A \cap B$; (2) 求 $P(A \cup B)$ 。**解**: $\Omega = \{1, 2, 3, 4, 5, 6\}$, $A = \{2, 4, 6\}$, $B = \{4, 5, 6\}$ 。(1) $A \cup B = \{2, 4, 5, 6\}$, $A \cap B = \{4, 6\}$ 。(2) A 与 B **不互斥** ($A \cap B = \{4, 6\} \neq \emptyset$)。

$\emptyset$), 必须用一般公式:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{3}{6} + \frac{3}{6} - \frac{2}{6} = \frac{4}{6} = \frac{2}{3}$$

答: $P(A \cup B) = \frac{2}{3}$ 。(若误用互斥加法 $P(A) + P(B) = 1$, 就把 $\{4, 6\}$ 这两个“同时满足”的样本点数了两次。) **Try It 2.2:** 掷一颗骰子, $A =$ “点数小于 3”, $B =$ “点数为奇数”。判断 A 与 B 是否互斥? 并求 $P(A \cup B)$ 。

样本空间与事件的 Venn 示意 (两圆相交图见排版稿, 此处用表格说明)

掷骰子时 $\Omega = \{1, 2, 3, 4, 5, 6\}$, 设 $A =$ 点数为偶数 $= \{2, 4, 6\}$, $B =$ 点数大于 3 $= \{4, 5, 6\}$:

关系	集合表示	掷骰子示例	概率公式
并事件	$A \cup B$	$\{2, 4, 5, 6\}$	$P(A \cup B) = P(A) + P(B) - P(A \cap B)$
交事件	$A \cap B$	$\{4, 6\}$	两个事件同时发生
互斥	$A \cap B = \emptyset$	$\{1\}$ 与 $\{2\}$	$P(A \cup B) = P(A) + P(B)$
对立	$\bar{A}$	$A = \{1, 3, 5\}, \bar{A} = \{2, 4, 6\}$	$P(\bar{A}) = 1 - P(A)$

若画 Venn 图: 样本空间 Ω 是一个大矩形 (或大圆), 事件 A, B 是两个圆。两圆**相交**表示 $A \cap B \neq \emptyset$ (不互斥); 两圆**不相交**表示互斥; 一个圆占满整个空间、另一个是它的补集, 表示对立。

2.1.3 古典概型: 等可能的格子怎么数

从生活到数学

掷骰子, 6 个面质地均匀, “1 点” 和 “6 点” 出现的可能性一样大。这种 “结果有限、每个结果等可能” 的试验, 概率就是 “数格子”: 想要的事件占几个格子, 除以总共有几个格子。

古典概型

如果随机试验满足: (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**, 那么该试验为**古典概型**。事件 A 的概率为

$$P(A) = \frac{n(A)}{n(\Omega)}$$

其中 $n(\Omega)$ 为样本点总数, $n(A)$ 为事件 A 所含样本点个数。

How To: 求古典概型概率

- 确认试验满足 “有限 + 等可能”;
- 用计数原理数出样本点总数 $n(\Omega)$ (有序对、排列、组合, 按需选用);
- 数出事件所含样本点个数 $n(A)$;
- 相除并约分。

例题 2.3

(掷骰子, 对抗 M1——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。**解:** 用有序数对 (a, b) 表示结果 (a 为第一枚点数, b 为第二枚), 共有 $6 \times 6 = 36$ 个样本点, 且等可能。和为 3 的样本点: $(1, 2)$ 、 $(2, 1)$, 共 2 个。

$$P = \frac{2}{36} = \frac{1}{18}$$

答: 点数之和为 3 的概率为 $\frac{1}{18}$ 。(若把样本空间写成“无序的和” $\{1, 2\}$ 一种, 得到 $\frac{1}{21}$ ——那是**不等可能**的计数: $(1, 2)$ 与 $(2, 1)$ 是两枚不同骰子的两个结果。古典概型的等可能, 必须对样本点而言。) **Try It 2.3:** 同时掷两枚骰子, 求点数之和为 7 的概率。(提示: 和为 7 的有 $(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)$ 共 6 个)

例题 2.4

(摸球, 组合计数) 袋中有 3 个红球、2 个蓝球, 大小质地完全相同。任取 2 个球, 求“恰好 1 个红球”的概率。**解:** 取球不看顺序, 样本点总数 $n(\Omega) = C_5^2 = 10$ 。“恰 1 红” = 1 红 1 蓝: $n(A) = C_3^1 \cdot C_2^1 = 6$ 。

$$P = \frac{6}{10} = \frac{3}{5}$$

答: 恰好 1 个红球的概率为 $\frac{3}{5}$ 。**Try It 2.4:** 接例题 2.4, 求“至少 1 个红球”的概率。(提示: 补集——“一个红球都没有”即两个都是蓝球)

例题 2.5

(“至少一个”转补集, 对抗 M4) 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。**解:** 正面枚举“中 1 次、2 次、3 次、4 次”很繁琐, 转反面: “一次都没中”。每次脱靶概率 $1 - 0.6 = 0.4$, 4 次都脱靶: 0.4^4 。

$$P(\text{至少中 1 次}) = 1 - 0.4^4 = 1 - 0.0256 = 0.9744$$

答: 至少命中 1 次的概率为 0.9744。(这是“至少一个”问题的万能钥匙: 先算“一个都没有”, 再用 1 减去。本题若正面枚举 4 种情况, 极易漏算。) **Try It 2.5:** 掷一枚骰子 3 次, 求“至少出现一次 6 点”的概率。

例题 2.6

(排队型, 排列计数) 4 名同学站成一排, 求甲站最左端的概率。**解:** 4 人全排 $n(\Omega) = 4! = 24$, 等可能。甲在最左端: 其余 3 人任意排, $n(A) = 3! = 6$ 。

$$P = \frac{6}{24} = \frac{1}{4}$$

答: 甲站最左端的概率为 $\frac{1}{4}$ 。**Try It 2.6:** 5 名同学站成一排, 求甲、乙两人相邻的概率。**想一想:** 两枚骰子“点数之和”有 2 ~ 12 共 11 种结果。有人说“点数之和为 7 的概率

是 $\frac{1}{11}$ ”，错在哪里？（提示：这 11 种结果**不等可能**——和为 7 有 6 个样本点，和为 2 只有 1 个。“等可能”是对样本点而言的，不是对“和的值”而言的。）**本节练习**（答案见章末）

- 抛一枚硬币 2 次，写出样本空间；求“恰好一次正面”的概率。
- 从 1, 2, 3, 4 中任取 2 个不同的数，求“取出的两数之和为偶数”的概率。
- 袋中 4 个白球、3 个红球，任取 2 球，求“至少 1 个白球”的概率。
- 掷两枚骰子，求点数之积为偶数的概率。
- 在 1 到 100 的整数中任取 1 个，求“它被 3 整除”的概率。

2.2 条件概率与全概率：信息来了，空间缩了

学习目标

- 理解条件概率的概念，掌握 $P(B|A) = \frac{P(A \cap B)}{P(A)}$ ；
- 理解乘法公式 $P(A \cap B) = P(A)P(B|A)$ ；
- 掌握全概率公式，会求“先分来源、再求和”的概率；
- 了解贝叶斯公式的“由果溯因”思想。

前置复习

- 掷一枚骰子，点数为奇数的概率是多少？
- 若已知掷出的点数为奇数，则“点数为 3”的概率是多少？（提示：样本空间已缩小为 $\{1, 3, 5\}$ ）
- 袋中 2 红 3 蓝，不放回连取 2 球，两次都取到红球的概率是多少？

2.2.1 条件概率：在缩小的样本空间里重新数

从生活到数学

袋中有 2 红 3 蓝共 5 个球，不放回连取两个。取完后有人告诉你：“第一个球是红的。”此时第二个球是红的概率是多少？——如果不知道这个信息，第二个球是红的概率是 $\frac{2}{5}$ ；但知道第一个是红的后，箱子里只剩 1 红 3 蓝，概率变成 $\frac{1}{4}$ 。**信息改变了概率**。这个“已知 A 发生，求 B 的概率”，就是条件概率。

条件概率

一般地，在事件 A 发生的条件下，事件 B 发生的概率称为**条件概率**，记作 $P(B|A)$ （其中 $P(A) > 0$ ）：

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

在古典概型下，等价地有

$$P(B|A) = \frac{n(A \cap B)}{n(A)}$$

How To：求条件概率

- 明确条件 A ：样本空间缩成 A ；
- 数“在 A 里， B 也发生”的部分： $n(A \cap B)$ 或 $P(A \cap B)$ ；
- 除以 A 的总量： $n(A)$ 或 $P(A)$ 。

条件概率“缩空间”示意（2 红 3 蓝共 5 球，不放回连取 2 个）

场景	样本空间	第二次取到红球的概率
未知信息	Ω （5 个球）	$P(B) = \frac{2}{5}$
已知首球为红	缩为 A （剩 1 红 3 蓝，4 个球）	$P(B A) = \frac{1}{4}$

信息一来，样本空间从 Ω **缩小到** A ——分子、分母同步在 A 里数，这就是条件概率的本质。

例题 2.7

（条件概率，对抗 M2/M12）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。**解**：方法一（缩空间）：已知第一次是红球，袋中剩 2 红 2 蓝共 4 球，

$$P(\text{二红} | \text{首红}) = \frac{2}{4} = \frac{1}{2}$$

方法二（公式）：设 $A =$ “首次红”， $A \cap B =$ “两次都红”。 $P(A) = \frac{3}{5}$ ， $P(A \cap B) = \frac{3}{5} \times \frac{2}{4} = \frac{3}{10}$ ，

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{3/10}{3/5} = \frac{1}{2}$$

答：第二次也摸到红球的概率为 $\frac{1}{2}$ 。（两种方法结果一致。分子用的是“两次都红” $P(A \cap B) = \frac{3}{10}$ ，而不是“第二次红” $P(B) = \frac{3}{5}$ ——后者忽略了“第一次必须是红球”这个前提。）**Try It 2.7**：接例题 2.7，若已知第一次摸到蓝球，求第二次摸到红球的概率。

例题 2.8

(公式运用, 对抗 M7) 已知 $P(A) = 0.4$, $P(A \cap B) = 0.1$, $P(B) = 0.2$ 。求 $P(B|A)$ 与 $P(A|B)$, 并比较二者。解:

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{0.1}{0.4} = 0.25, \quad P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.1}{0.2} = 0.5$$

答: $P(B|A) = 0.25$, $P(A|B) = 0.5$ 。($P(B|A)$ 与 $P(A|B)$ 是**两个不同的数**——条件不同, 分母就不同。把 $P(B|A)$ 与 $P(A \cap B)$ 混用, 是最容易犯的错。) **Try It 2.8**: 已知 $P(A) = 0.3$, $P(B) = 0.5$, $P(A \cup B) = 0.6$ 。求 $P(A \cap B)$ 与 $P(B|A)$ 。

Q&A

- 问: $P(B|A)$ 和 $P(A \cap B)$ 到底有什么区别?

答: $P(A \cap B)$ 是在**整个样本空间**里问 “A、B 同时发生” 占多大比例; $P(B|A)$ 是**把样本空间缩到 A 后**, 再问 “B 还占多大比例”。数值上 $P(B|A) = \frac{P(A \cap B)}{P(A)}$, 一般两者不相等。

- 问: 为什么 “已知 A 发生” 这个信息会改变概率?

答: 因为样本空间变了。好比全班 40 人, 随机抽 1 人是女生概率 $\frac{20}{40}$; 但若已知抽到的是 “篮球队成员” (篮球队恰好全是女生), 概率立刻变成 1——信息缩小了备选范围。

2.2.2 全概率公式与贝叶斯公式: 先分来源, 再求和

从生活到数学

某商场出售一种商品, 70% 来自甲厂 (次品率 1%), 30% 来自乙厂 (次品率 3%)。你随手拿一件, 它是不是次品? ——它可能来自甲厂, 也可能来自乙厂, 两个来源的 “次品率” 还不一样。这时不能只看某个厂, 而要把两个来源**各自算一遍再加起来**。这就是全概率公式的思路: “先按来源分类, 类内相乘、类间相加” ——它正是计数章 “分类加法” 在概率世界的投影。

全概率公式

设 $A_1, A_2, \dots, A_n$ 是样本空间 Ω 的一个**完备事件组** (两两互斥, 且 $A_1 \cup A_2 \cup \dots \cup A_n = \Omega$), 则对任一事件 B :

$$P(B) = P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + \dots + P(A_n)P(B|A_n)$$

贝叶斯公式

在同样的条件下, 若已知 B 已经发生, 求 “它来自 A_i ” 的概率:

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{P(A_1)P(B|A_1) + \dots + P(A_n)P(B|A_n)}$$

全概率公式树形图 (例题 2.9 的场景: 先分来源, 类内相乘、类间相加)

- 来源（完备事件组）
- A1 甲厂 (0.7)
- 次品 $B | A1$ (0.01): $0.7 \times 0.01 = 0.007$
- 合格 (0.99)
- A2 乙厂 (0.3)
- 次品 $B | A2$ (0.03): $0.3 \times 0.03 = 0.009$
- 合格 (0.97)

$P(B) = 0.007 + 0.009 = 0.016$ （两条“次品”分支相加，任何一条都不能漏）

例题 2.9

（全概率公式，对抗 M8）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。**解：**设 A_1 = “来自甲厂”， A_2 = “来自乙厂”， B = “次品”。 A_1 、 A_2 两两互斥且并集为 Ω ，构成完备事件组。

$$P(B) = P(A_1)P(B|A_1) + P(A_2)P(B|A_2) = 0.7 \times 0.01 + 0.3 \times 0.03 = 0.007 + 0.009 = 0.016$$

答：任取一件是次品的概率为 0.016，即 1.6%。（若漏掉乙厂这一项，会得到 0.007——明显偏小。**完备事件组的每一项都不能漏。**）**Try It 2.9：**某地区有甲、乙两个气象站：甲站预报准确的概率为 0.8，该地区 60% 的预报由甲站作出；乙站预报准确概率为 0.9。随机调取该地区一条预报记录，求它预报准确的概率。

例题 2.10

（贝叶斯公式，由果溯因）接例题 2.9：若任取一件发现它是次品，求它来自甲厂的概率。**解：**用贝叶斯公式：

$$P(A_1|B) = \frac{P(A_1)P(B|A_1)}{P(B)} = \frac{0.7 \times 0.01}{0.016} = \frac{0.007}{0.016} = \frac{7}{16}$$

答：次品来自甲厂的概率为 $\frac{7}{16} \approx 0.4375$ 。（直觉检验：甲厂供货多（70%）但次品率低（1%），乙厂供货少但次品率高（3%）——所以已知是次品时，它来自乙厂的概率 $\frac{9}{16}$ 其实并不低。**贝叶斯公式把“先验信息”和“新证据”合在了一起。**）**Try It 2.10：**接例题 2.10，若任取一件发现它是次品，求它来自乙厂的概率。（提示：可用 $1 - P(A_1|B)$ ，也可直接用贝叶斯公式）**想一想：**抽签公平性。3 人依次抽签（不放回），只有 1 个中奖签。第 1 人抽中的概率是 $\frac{1}{3}$ 。用全概率公式算一算第 2 人抽中的概率，并说说“抽签为什么公平”。**本节练习**（答案见章末）

- 袋中 2 白 3 黑共 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。
- 已知 $P(A) = 0.5$, $P(A \cap B) = 0.2$, 求 $P(B|A)$ 。
- 某厂有两台机床：甲机床加工了 60% 的零件，次品率 2%；乙机床加工 40%，次品率 4%。从中任取一件，求它是次品的概率。
- 接第 3 题：若取出的零件是次品，求它由甲机床加工的概率。
- 把一枚硬币连抛 3 次。已知三次中至少有一次正面，求“恰好两次正面”的概率。（提示：条件概率——在“至少一次正面”这个条件下）

2.3 事件的独立性与二项分布：互不干扰的重复

学习目标

- 理解事件相互独立的概念与判据 $P(A \cap B) = P(A)P(B)$ ；
- 能区分“独立”与“互斥”，知道二者不可兼得；
- 理解独立重复试验，掌握二项分布 $P(X = k) = C_n^k p^k (1 - p)^{n-k}$ ；
- 会求二项分布的概率及“至少一次”型问题。

前置复习

- 掷一颗骰子，“点数小于 3”与“点数为偶数”能同时发生吗？
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ 中， $P(A \cap B)$ 表示什么？
- 抛 3 次硬币，恰好 2 次正面有哪几种情况？

2.3.1 事件的相互独立性：信息不干扰

从生活到数学

甲、乙两人独立射击同一目标，命中率分别是 0.8、0.7。甲中不中，不影响乙中不中——这就是“独立”。可“互斥”呢？“甲中”和“乙中”其实可以同时发生，它们根本不互斥。独立与互斥，是两套完全不同的说法。

事件的相互独立性

如果事件 A 的发生不影响事件 B 发生的概率，即 $P(B|A) = P(B)$ ，则称事件 A 与 B 相互独立。等价地：

$$A \text{ 与 } B \text{ 独立} \iff P(A \cap B) = P(A)P(B)$$

How To: 判断两个事件是否独立

- 观察背景：放回抽样、独立试验、互不干扰的环节——通常独立；
- 或验证等式：算 $P(A \cap B)$ 与 $P(A)P(B)$ ，相等则独立；
- 不放回抽样**不独立**（第二次的概率随第一次变化）。

例题 2.11

（独立 vs 互斥辨析，对抗 M3/M9）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？**解：** $A = \{1, 3, 5\}$ ， $B = \{2, 4, 6\}$ ， $A \cap B = \emptyset \rightarrow$ **互斥**。 $P(A) = \frac{1}{2}$ ， $P(B) = \frac{1}{2}$ ， $P(A)P(B) = \frac{1}{4}$ ，而 $P(A \cap B) = 0 \neq \frac{1}{4} \rightarrow$ **不独立**。
答： A 与 B 互斥但不独立。（“一个数不可能既是奇数又是偶数”是互斥；但独立要的是**概率等式成立**——这里 $P(A \cap B)$ 与 $P(A)P(B)$ 差了 $\frac{1}{4}$ ，天壤之别。）**Try It 2.11：**掷一颗骰子， $A =$ “点数为奇数”， $B =$ “点数大于 4”。判断 A 与 B 是否独立。（提示： $A = \{1, 3, 5\}$ ， $B = \{5, 6\}$ ， $A \cap B = \{5\}$ ，算算 $P(A)P(B)$ 与 $P(A \cap B)$ 是否相等）

例题 2.12

（独立事件的交与并，对抗 M9）甲、乙两人独立地射击同一目标，命中率分别为 0.8、0.7。（1）两人都命中的概率；（2）至少一人命中的概率；（3）恰好一人命中的概率。**解：**设 $A =$ “甲中”， $B =$ “乙中”， A 、 B 独立。（1） $P(A \cap B) = P(A)P(B) = 0.8 \times 0.7 = 0.56$ 。（2） $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.8 + 0.7 - 0.56 = 0.94$ ，或用补集： $1 - 0.2 \times 0.3 = 0.94$ 。（3） $P(\text{恰一}) = 0.8 \times 0.3 + 0.2 \times 0.7 = 0.24 + 0.14 = 0.38$ 。（验证： $0.56 + 0.38 + 0.06 = 1$ ✓——两人都中、恰一中、都不中，三种情况不重不漏。）（注意（2）： A 、 B 独立但**不互斥**， $P(A \cup B)$ 必须**减去** $P(A \cap B)$ ；若误用互斥加法 $0.8 + 0.7 = 1.5$ ，概率都超过 1 了，一眼假。）**Try It 2.12：**甲、乙两人独立解题，甲解出概率 0.9，乙解出概率 0.8。求“至少一人解出”的概率。**想一想（对抗 M5）：**放回与不放回。袋中 2 红 3 蓝，连取 2 球。（1）有放回地取，两次都取到红球的概率是多少？（2）不放回地取，两次都取到红球的概率是多少？两次试验分别独立吗？（提示：有放回时第二次概率仍是 $\frac{2}{5}$ ，独立；不放回时第二次概率变成 $\frac{1}{4}$ ，**不独立**。同样是“两红”，一个独立、一个不独立，模型完全不同。）

2.3.2 独立重复试验与二项分布： n 次“只问成败”

从生活到数学

一次试验只有“成功/失败”两种结果（如抛硬币的正反、猜选择题的对错），在**相同条件下独立重复** n 次——这就是 n 重伯努利试验。抛 3 次硬币，“恰好 2 次正面”怎么算？先想“哪 2 次是正面”： C_3^2 种选法；再乘每次的概率： $\left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^1$ 。组合数 + 独立乘法，一相乘，就是二项分布。

二项分布

一般地，在 n 次独立重复试验中，每次“成功”的概率为 p ，设 X 为 n 次试验中成功的

次数，则

$$P(X = k) = C_n^k p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n$$

称 X 服从二项分布，记作 $X \sim B(n, p)$ 。

下面这张柱状图展示 $n = 10$ 、 $p = 0.5$ 时 $X \sim B(10, 0.5)$ 各取值 k 的概率（柱高正比于 $P(X = k) = \frac{C_{10}^k}{2^{10}}$ ）：

[figure 渲染失败: 二项分布 $X \sim B(10, 0.5)$ 的分布柱状图]

How To: 识别并计算二项分布概率

- 核对三条件：固定次数、两结果、独立且 p 恒定；
- 明确 n 、 p 、 k ；
- 代入 $P(X = k) = C_n^k p^k (1-p)^{n-k}$ ；
- “至少 k 次” 通常用补集（1- 少于 k 次的概率）。

例题 2.13

（二项分布概率）某射手射击命中率 0.8，各次射击独立。射击 5 次，求恰好命中 3 次的概率。**解：** 这是 $n = 5$ 、 $p = 0.8$ 的二项分布：

$$P(X = 3) = C_5^3 (0.8)^3 (0.2)^2 = 10 \times 0.512 \times 0.04 = 0.2048$$

答：恰好命中 3 次的概率为 0.2048。**Try It 2.13:** 抛一枚硬币 6 次，求恰好 3 次正面的概率。

例题 2.14

（猜选择题，至少/恰好）一份试卷有 10 道四选一的选择题，小明完全凭猜测作答。求他恰好猜对 3 道的概率。**解：** 每题猜对概率 $p = \frac{1}{4}$ ，独立作答， $X \sim B\left(10, \frac{1}{4}\right)$ ：

$$P(X = 3) = C_{10}^3 \left(\frac{1}{4}\right)^3 \left(\frac{3}{4}\right)^7 = 120 \times \frac{1}{64} \times \frac{2187}{16384} = \frac{262440}{1048576} \approx 0.2503$$

答：恰好猜对 3 道的概率约为 0.2503。**Try It 2.14:** 接例题 2.14，求他“至少猜对 1 道”的概率。（提示：补集——一题都没猜对）

例题 2.15

（模型辨析，对抗 M10/M11）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。**解：**（1）不放

回 → **不是**二项分布（各次不独立），用组合计数（超几何型）：

$$P = \frac{C_3^2 C_7^2}{C_{10}^4} = \frac{3 \times 21}{210} = \frac{63}{210} = \frac{3}{10}$$

(2) 有放回 → 每次次品率 $p = \frac{3}{10}$ 恒定且独立， $X \sim B\left(4, \frac{3}{10}\right)$ ：

$$P(X=2) = C_4^2 (0.3)^2 (0.7)^2 = 6 \times 0.09 \times 0.49 = 0.2646$$

答：(1) $\frac{3}{10}$ ；(2) 0.2646。（同样是“抽 4 件恰 2 件次品”，放回与不放回答案不同——**先判断模型，再代公式**。看到“不放回”直接套二项分布，必错。）**Try It 2.15**：袋中 5 个球，2 红 3 蓝。有放回地取 3 次，求“恰有 2 次取到红球”的概率。

Q&A

- **问**：二项分布里“至少一次成功”为什么总用补集？

答： $P(\text{至少一次}) = 1 - P(\text{零次}) = 1 - (1-p)^n$ 。正面枚举 $k = 1, 2, \dots, n$ 要加 n 项，补集只算一项，快且不易漏。

- **问**：独立事件和互斥事件什么时候会混淆？

答：都带“互不”的字眼。记住两句话：互斥是“**不能同时发生**”（集合上无交集），独立是“**概率上互不影响**”（乘法公式成立）。当 $P(A), P(B) > 0$ 时，互斥的事件必然不独立。

本节练习（答案见章末）

- 判断： A 与 B 互斥，且 $P(A) = 0.3$ ， $P(B) = 0.4$ ，则 $P(A \cup B) = 0.7$ 。这个说法对吗？
- 掷两枚骰子， $A =$ “第一枚为偶数”， $B =$ “第二枚为奇数”。判断 A 与 B 是否独立。
- 某人射击命中率 0.7，独立射击 4 次，求恰好命中 2 次的概率。
- 抛一枚骰子 5 次，求“至少出现一次 6 点”的概率。
- 一箱 8 件产品含 2 件次品，不放回抽 3 件，求恰好 1 件次品的概率。

2.4 数字特征与正态分布初步：分布的“平均数”与“钟形曲线”

学习目标

- 理解离散型随机变量及其分布列，掌握分布列的性质；

- 掌握期望与方差的定义及二项分布的 $E(X) = np$ 、 $D(X) = np(1-p)$;
- 了解正态分布曲线及 3σ 原则, 会用对称性求概率。

前置复习

- 掷一颗骰子, 点数的平均数是多少?
- 求 C_{10}^2 与 C_{10}^8 的关系。
- 二项分布 $X \sim B(n, p)$ 的概率公式是什么?

2.4.1 离散型随机变量的期望与方差

从生活到数学

一次抛硬币试验, 结果可以用数字 X 表示: 正面 $X = 1$, 反面 $X = 0$ 。这个 X 的取值不确定, 但每个取值都有确定的概率——它就是**离散型随机变量**。把所有取值和对应概率排成一张表, 就是**分布列**。有了分布列, 就能算“长期平均”(期望)和“波动大小”(方差)。

离散型随机变量的期望与方差

若随机变量 X 的分布列为 $P(X = x_i) = p_i$ ($i = 1, 2, \dots, n$), 则

$$E(X) = x_1p_1 + x_2p_2 + \cdots + x_np_n$$

称为 X 的**均值 (期望)**;

$$D(X) = \sum_{i=1}^n (x_i - E(X))^2 p_i$$

称为 X 的**方差**, $\sqrt{D(X)}$ 称为**标准差**。

How To: 求期望与方差

- 写出 X 的所有可能取值与概率 (分布列);
- 期望: 取值 $\times$ 概率, 全部相加;
- 方差: 先算期望, 再算 $(x_i - E)^2 \times p_i$ 之和;
- 若是二项分布, 直接套 $E = np$ 、 $D = np(1-p)$ 。

例题 2.16

(分布列与期望) 抛一枚硬币: 正面得 1 分, 反面得 0 分。设 X 为一次抛掷得分。写出 X 的分布列, 并求 $E(X)$ 与 $D(X)$ 。**解:** 分布列: $P(X = 1) = \frac{1}{2}$, $P(X = 0) = \frac{1}{2}$ 。

$$E(X) = 1 \times \frac{1}{2} + 0 \times \frac{1}{2} = \frac{1}{2}$$

$$D(X) = \left(1 - \frac{1}{2}\right)^2 \times \frac{1}{2} + \left(0 - \frac{1}{2}\right)^2 \times \frac{1}{2} = \frac{1}{4} \times \frac{1}{2} + \frac{1}{4} \times \frac{1}{2} = \frac{1}{4}$$

答: $E(X) = \frac{1}{2}$, $D(X) = \frac{1}{4}$ 。(这也验证了两点分布: $E = p$, $D = p(1 - p)$ 。) **Try It 2.16:** 掷一颗骰子, 设 X 为点数。求 $E(X)$ 。

例题 2.17

(二项分布的数字特征, 对抗 M6) 某射击手命中率 0.2, 独立射击 10 次, 设 X 为命中次数。求 $E(X)$ 与 $D(X)$ 。**解:** $X \sim B(10, 0.2)$:

$$E(X) = np = 10 \times 0.2 = 2, \quad D(X) = np(1 - p) = 10 \times 0.2 \times 0.8 = 1.6$$

答: $E(X) = 2$, $D(X) = 1.6$ 。(直觉校验: 每次命中概率 0.2, 射 10 次“平均中 2 次”——期望必须带 np ; 方差必须带 $(1 - p)$: $D = np(1 - p) = 1.6$, **不是** $np = 2$ 。) **Try It 2.17:** $X \sim B(100, 0.5)$, 求 $E(X)$ 与 $D(X)$ 。

例题 2.18

(由数字特征反解参数) 设 $X \sim B(n, p)$, 且 $E(X) = 12$, $D(X) = 4.8$, 求 n 、 p 。**解:**

$$\begin{cases} np = 12 \\ np(1 - p) = 4.8 \end{cases}$$
 。代入第一式: $12(1 - p) = 4.8$, $1 - p = 0.4$, $p = 0.6$; 于是 $n = \frac{12}{0.6} = 20$ 。答: $n = 20$, $p = 0.6$ 。(验证: $E = 20 \times 0.6 = 12$, $D = 20 \times 0.6 \times 0.4 = 4.8$ ✓) **Try It 2.18:** 设 $X \sim B(n, p)$, $E(X) = 8$, $D(X) = 4.8$, 求 n 、 p 。**想一想:** 抽奖活动: 抽一次花费 2 元, 中奖概率 0.1, 中奖可得 15 元。设 X 为一次抽奖的净收益 (中奖 $X = 13$, 不中 $X = -2$)。求 $E(X)$, 并判断这个抽奖“值不值得玩”。

2.4.2 正态分布初步: 无处不在的钟形曲线

从生活到数学

你量过全班同学的身高吗? 大多数人身高集中在平均值附近, 特别高和特别矮的都很少, 把频数画出来, 是一条两边低、中间高的**钟形曲线**。身高、考试分数、机器生产的零件尺寸……无数“大量独立微小因素叠加”的量, 都服从这条曲线。它就是正态分布。

正态分布

若随机变量 X 的概率密度曲线为

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

则称 X 服从**正态分布**, 记作 $X \sim N(\mu, \sigma^2)$, 其中 μ 是均值, σ 是标准差。曲线关于直线 $x = \mu$ 对称, 在 $x = \mu$ 处最高; σ 越大曲线越“矮胖”, σ 越小越“高瘦”。

3 σ 原则 (正态分布的概率)

$$P(\mu - \sigma < X < \mu + \sigma) \approx 0.6827, \quad P(\mu - 2\sigma < X < \mu + 2\sigma) \approx 0.9545, \quad P(\mu - 3\sigma < X < \mu + 3\sigma) \approx 0.9973$$

下图是标准正态分布 $N(0,1)$ 的密度曲线，并标出 $\mu \pm \sigma$ 、 $\mu \pm 2\sigma$ 、 $\mu \pm 3\sigma$ 的位置：

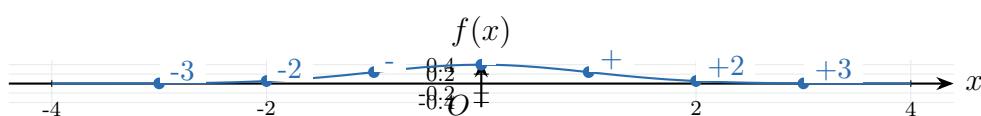


图 2.1: 标准正态分布曲线 $N(0,1)$ 与 3σ 标注

How To: 利用对称性求正态概率

- 找出 μ 、 σ ；
- 把所求区间“翻译”成关于 μ 对称的区间或一侧区间；
- 用对称性（两侧各半）+ 3σ 原则拼出答案。

例题 2.19

（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 $N(10, 0.04)$ ，即 $\mu = 10$ ， $\sigma = 0.2$ 。（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。**解：** $\mu = 10$ ， $\sigma = 0.2$ 。（1） $9.8 = 10 - 0.2 = \mu - \sigma$ ， $10.2 = 10 + 0.2 = \mu + \sigma$ ，故

$$P(9.8 < X < 10.2) = P(\mu - \sigma < X < \mu + \sigma) \approx 0.6827$$

（2）曲线关于 $x = 10$ 对称，小于 μ 与大于 μ 的面积各占一半：

$$P(X < 10) = 0.5$$

答：（1）约 0.6827；（2）0.5。**Try It 2.19:** 接例题 2.19，求 $P(9.6 < X < 10.4)$ 与 $P(X > 10.4)$ 的近似值。

例题 2.20

（对称性比较）设 $X \sim N(2, 9)$ ($\mu = 2$ ， $\sigma = 3$)。比较 $P(X < 0)$ 与 $P(X > 4)$ 的大小。

解：点 0 与 4 关于 $\mu = 2$ 对称 ($\frac{0+4}{2} = 2$)。正态曲线关于 $x = 2$ 对称，故

$$P(X < 0) = P(X > 4)$$

答：二者相等。（利用对称性，无需查表就能比较两侧尾部概率的大小。）**Try It 2.20:** 设 $X \sim N(1, 4)$ ($\sigma = 2$)，比较 $P(X < -1)$ 与 $P(X > 3)$ 的大小。

Q&A

- **问：**正态分布里的 μ 、 σ 与期望、方差有什么关系？

答: μ 就是期望 $E(X)$, σ^2 就是方差 $D(X)$ 。正态分布是连续型分布, 但“均值是重心、方差是离散度”的思想与离散型完全一致。

- 问: σ 越大, 曲线越矮胖还是越高瘦?

答: 越矮胖——数据越分散, 中间峰值越低, 两端尾巴越厚。

本节练习 (答案见章末)

- 设 $X \sim B(5, 0.4)$, 求 $E(X)$ 、 $D(X)$ 。
- 抛 3 次硬币, 设 X 为正面次数。写出 X 的分布列, 并求 $E(X)$ 、 $D(X)$ 。
- 某班学生身高 $X \sim N(170, 25)$ ($\sigma = 5$)。求身高在 $[165, 175]$ 内的概率近似值。
- 设 $X \sim N(0, 1)$ (标准正态, $\mu = 0$, $\sigma = 1$), 求 $P(X > 2)$ 的近似值。
- 摸奖箱中有 5 个白球、1 个红球, 有放回地摸 4 次, 设 X 为摸到红球的次数。求 $E(X)$ 、 $D(X)$ 。

本章回顾

概念地图

· 随机试验 $\rightarrow$ 样本空间 Ω (全体样本点) $\rightarrow$ 事件 (Ω 的子集) · 事件关系: $A \cup B$ (或) · $A \cap B$ (且) · 互斥 $A \cap B = \emptyset$ · 对立 $P(\bar{A}) = 1 - P(A)$ · 古典概型: 有限 + 等可能 $\rightarrow P(A) = \frac{n(A)}{n(\Omega)}$ (分子分母用计数原理数) · 概率性质: $0 \leq P(A) \leq 1$ · $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ · 条件概率: $P(B|A) = \frac{P(A \cap B)}{P(A)}$ (把样本空间缩到 A) · 乘法公式: $P(A \cap B) = P(A)P(B|A)$ · 全概率: $P(B) = \sum_i P(A_i)P(B|A_i)$ (完备事件组, 先分类再求和) · 贝叶斯: $P(A_i|B) = \frac{P(A_i)P(B|A_i)}{P(B)}$ (由果溯因) · 独立: $P(A \cap B) = P(A)P(B)$ (互斥与独立, $P > 0$ 时不可兼得) · 二项分布: $X \sim B(n, p)$, $P(X = k) = C_n^k p^k (1-p)^{n-k}$ · 数字特征: $E(X) = np$, $D(X) = np(1-p)$; 标准差 $\sqrt{D(X)}$ · 正态分布: $N(\mu, \sigma^2)$, 关于 $x = \mu$ 对称; 3σ 原则 0.6827 / 0.9545 / 0.9973

关键公式速查

内容	公式
古典概型	$P(A) = \frac{n(A)}{n(\Omega)}$ (有限 + 等可能)
并事件概率	$P(A \cup B) = P(A) + P(B) - P(A \cap B)$
对立事件	$P(\bar{A}) = 1 - P(A)$
条件概率	$P(B A) = \frac{P(A \cap B)}{P(A)}$
乘法公式	$P(A \cap B) = P(A)P(B A)$
全概率	$P(B) = \sum_i P(A_i)P(B A_i)$
独立判据	$P(A \cap B) = P(A)P(B)$
二项分布	$P(X = k) = C_n^k p^k (1-p)^{n-k}$
期望 / 方差 (二项)	$E(X) = np, D(X) = np(1-p)$
正态 3σ	$[\mu \pm \sigma]: 0.6827; [\mu \pm 2\sigma]: 0.9545; [\mu \pm 3\sigma]: 0.9973$

分层练习

A 基础层

- 掷一颗骰子，求点数不小于 4 的概率。
- 从 5 名男生、4 名女生中任选 2 人，求“恰 1 男 1 女”的概率。
- 抛 3 枚硬币，求“至少 2 枚正面”的概率。
- 已知 $P(A) = 0.5$, $P(B) = 0.4$, $P(A \cap B) = 0.2$, 求 $P(B|A)$ 。
- 设 $X \sim B(6, 0.3)$, 求 $P(X = 2)$ 。
- 设 $X \sim B(12, 0.25)$, 求 $E(X)$ 、 $D(X)$ 。
- 设 $X \sim N(100, 100)$ ($\sigma = 10$), 求 $P(90 < X < 110)$ 的近似值。
- 抛两枚骰子，求“点数之和为偶数”的概率。

B 应用层

- 某工厂生产一批产品，次品率为 4%，质检员独立抽检 5 件。求“至少 1 件次品”的概率。
- 一箱 12 件产品含 3 件次品，不放回抽 2 件，求“至少 1 件次品”的概率。
- 甲、乙两射手独立射击同一目标，命中率分别为 0.8、0.9。求“恰有 1 人命中”的概率。
- 设 $X \sim B(n, p)$, $E(X) = 4$, $D(X) = 3.2$, 求 n 、 p 。

C 拓展层

- 某地区人群患某种疾病的概率为 0.1%。检测试剂：患病者检出阳性的概率为 99%，未患病者误报阳性的概率为 2%。(1) 随机抽取一人检测为阳性的概率；(2) 已知检测为阳性，此人真正患病的概率。(结果保留四位小数)

- 设 $X \sim N(\mu, \sigma^2)$, 且 $P(X > 3) = P(X < 1)$, 求 μ 。

参考答案 (章末集中)

Try It 即时练习答案

- 2.1: 奇数 $A = \{1, 3, 5\}$; 大于 4 $B = \{5, 6\}$; $A \cap B = \{5\}$; $A \cup B = \{1, 3, 5, 6\}$ 。
- 2.2: $A = \{1, 2\}$, $B = \{1, 3, 5\}$, $A \cap B = \{1\} \neq \emptyset \rightarrow$ **不互斥**; $P(A \cup B) = \frac{5}{6}$ 。
- 2.3: 和为 7 的有 6 个样本点, $P = \frac{6}{36} = \frac{1}{6}$ 。
- 2.4: 补集: $1 - \frac{C_2^2}{C_5^2} = 1 - \frac{1}{10} = \frac{9}{10}$ 。
- 2.5: $1 - \left(\frac{5}{6}\right)^3 = 1 - \frac{125}{216} = \frac{91}{216}$ 。
- 2.6: 相邻排法 $2! \times 4! = 48$, 总数 $5! = 120$, $P = \frac{48}{120} = \frac{2}{5}$ 。
- 2.7: 已知首次为蓝, 袋中剩 3 红 1 蓝, $P = \frac{3}{4}$ 。
- 2.8: $P(A \cap B) = P(A) + P(B) - P(A \cup B) = 0.3 + 0.5 - 0.6 = 0.2$; $P(B|A) = \frac{0.2}{0.3} = \frac{2}{3}$ 。
- 2.9: $0.6 \times 0.8 + 0.4 \times 0.9 = 0.48 + 0.36 = 0.84$ 。
- 2.10: $1 - \frac{7}{16} = \frac{9}{16}$ (或 $\frac{0.3 \times 0.03}{0.016} = \frac{0.009}{0.016} = \frac{9}{16}$)。
- 2.11: $A = \{1, 3, 5\}$, $B = \{5, 6\}$, $P(A) = \frac{1}{2}$, $P(B) = \frac{1}{3}$, $P(A \cap B) = \frac{1}{6}$, 而 $P(A)P(B) = \frac{1}{2} \times \frac{1}{3} = \frac{1}{6}$, 相等 $\rightarrow$ **独立**。(独立的两个事件可以同时发生, 也可以不同时发生。)
- 2.12: $1 - 0.1 \times 0.2 = 0.98$ 。
- 2.13: $C_6^3 \left(\frac{1}{2}\right)^6 = \frac{20}{64} = \frac{5}{16}$ 。
- 2.14: $1 - \left(\frac{3}{4}\right)^{10} = 1 - \frac{59049}{1048576} = \frac{989527}{1048576} \approx 0.9437$ 。
- 2.15: $C_3^2 \left(\frac{2}{5}\right)^2 \left(\frac{3}{5}\right) = 3 \times \frac{4}{25} \times \frac{3}{5} = \frac{36}{125}$ 。
- 2.16: $E(X) = \frac{1+2+3+4+5+6}{6} = 3.5$ 。
- 2.17: $E = 100 \times 0.5 = 50$; $D = 100 \times 0.5 \times 0.5 = 25$ 。
- 2.18: $8(1-p) = 4.8 \rightarrow 1-p = 0.6$, $p = 0.4$; $n = \frac{8}{0.4} = 20$ 。

- **2.19:** $9.6 = \mu - 2\sigma$, $10.4 = \mu + 2\sigma \rightarrow P \approx 0.9545$; $P(X > 10.4) \approx \frac{1 - 0.9545}{2} = 0.02275$ 。
- **2.20:** -1 与 3 关于 $\mu = 1$ 对称 $\rightarrow P(X < -1) = P(X > 3)$, 相等。

各节练习答案

- **2.1 节:** 1. $\Omega = \{(\text{正}, \text{正}), (\text{正}, \text{反}), (\text{反}, \text{正}), (\text{反}, \text{反})\}$; “恰好一次正面” = $\{(\text{正}, \text{反}), (\text{反}, \text{正})\}$, $P = \frac{2}{4} = \frac{1}{2}$ 。2. 两数和为偶数 = 同奇或同偶: $C_2^2 + C_2^2 = 1 + 1 = 2$ 种, 总数 $C_4^2 = 6$, $P = \frac{2}{6} = \frac{1}{3}$ 。3. 补集: $1 - \frac{C_3^2}{C_7^2} = 1 - \frac{3}{21} = \frac{6}{7}$ 。4. 补集: 积为奇数 (两枚都奇数) 有 $3 \times 3 = 9$ 种, $P = 1 - \frac{9}{36} = \frac{3}{4}$ 。5. $1 \sim 100$ 中 3 的倍数共 33 个, $P = \frac{33}{100}$ 。
- **2.2 节:** 1. 已知第一个是白球, 袋中剩 1 白 3 黑, $P = \frac{3}{4}$ 。2. $P(B|A) = \frac{0.2}{0.5} = 0.4$ 。3. $P = 0.6 \times 0.02 + 0.4 \times 0.04 = 0.012 + 0.016 = 0.028$ 。4. $P = \frac{0.012}{0.028} = \frac{3}{7}$ 。5. 设 $A =$ “至少一次正面”, $B =$ “恰好两次正面”。 $P(A) = 1 - \left(\frac{1}{2}\right)^3 = \frac{7}{8}$, $P(B) = C_3^2 \left(\frac{1}{2}\right)^3 = \frac{3}{8}$, 且 $B \subseteq A$, $P(B|A) = \frac{P(B)}{P(A)} = \frac{3/8}{7/8} = \frac{3}{7}$ 。
- **2.3 节:** 1. 对——互斥时 $P(A \cup B) = P(A) + P(B) = 0.3 + 0.4 = 0.7$ 。2. 独立—— $P(A) = \frac{1}{2}$, $P(B) = \frac{1}{2}$, $P(A \cap B) = \frac{3 \times 3}{36} = \frac{1}{4}$, $P(A)P(B) = \frac{1}{4}$, 相等。3. $C_4^2(0.7)^2(0.3)^2 = 6 \times 0.49 \times 0.09 = 0.2646$ 。4. $1 - \left(\frac{5}{6}\right)^5 = 1 - \frac{3125}{7776} = \frac{4651}{7776} \approx 0.5981$ 。5. $\frac{C_2^1 C_6^2}{C_8^3} = \frac{2 \times 15}{56} = \frac{30}{56} = \frac{15}{28}$ 。
- **2.4 节:** 1. $E = 5 \times 0.4 = 2$, $D = 5 \times 0.4 \times 0.6 = 1.2$ 。2. $P(X = k) = C_3^k \left(\frac{1}{2}\right)^3$ ($k = 0, 1, 2, 3$): $\frac{1}{8}, \frac{3}{8}, \frac{3}{8}, \frac{1}{8}$; $E = 3 \times \frac{1}{2} = 1.5$, $D = 3 \times \frac{1}{2} \times \frac{1}{2} = 0.75$ 。3. $[165, 175] = [\mu - \sigma, \mu + \sigma] \rightarrow P \approx 0.6827$ 。4. $P(X > 2) = \frac{1 - P(-2 < X < 2)}{2} \approx \frac{1 - 0.9545}{2} = 0.02275$ 。5. $X \sim B\left(4, \frac{1}{6}\right)$: $E = \frac{4}{6} = \frac{2}{3}$, $D = 4 \times \frac{1}{6} \times \frac{5}{6} = \frac{20}{36} = \frac{5}{9}$ 。

分层练习答案

A1: $\{4, 5, 6\}$, $P = \frac{3}{6} = \frac{1}{2}$ 。

A2: $\frac{C_5^1 C_4^1}{C_9^2} = \frac{20}{36} = \frac{5}{9}$ 。

A3: 至少 2 正 = $C_3^2 + C_3^3 = 3 + 1 = 4$ 种, $P = \frac{4}{8} = \frac{1}{2}$ 。

A4: $P(B|A) = \frac{0.2}{0.5} = 0.4$ 。

A5: $C_6^2(0.3)^2(0.7)^4 = 15 \times 0.09 \times 0.2401 = 0.324135 \approx 0.3241$ 。

A6: $E = 12 \times 0.25 = 3$; $D = 12 \times 0.25 \times 0.75 = 2.25$ 。

A7: $[90, 110] = [\mu - \sigma, \mu + \sigma]$, $P \approx 0.6827$ 。

A8: 和偶数 = 同奇同偶: $\frac{3 \times 3 + 3 \times 3}{36} = \frac{18}{36} = \frac{1}{2}$ 。

B9: $1 - 0.96^5 = 1 - 0.8153726976 \approx 0.1846$ 。

B10: $1 - \frac{C_9^2}{C_{12}^2} = 1 - \frac{36}{66} = \frac{30}{66} = \frac{5}{11}$ 。

B11: $P(\text{甲中乙不中}) + P(\text{甲不中乙中}) = 0.8 \times 0.1 + 0.2 \times 0.9 = 0.08 + 0.18 = 0.26$ 。

B12: $4(1 - p) = 3.2 \rightarrow 1 - p = 0.8$, $p = 0.2$, $n = \frac{4}{0.2} = 20$ 。

C13: (1) $P(\text{阳性}) = 0.001 \times 0.99 + 0.999 \times 0.02 = 0.00099 + 0.01998 = 0.02097$; (2) $P(\text{患病}|\text{阳性}) = \frac{0.00099}{0.02097} \approx 0.0472$ 。(启示: 阳性率 2.097%, 但真正患病只有约 4.7%——**低患病率下, 检测阳性多数是误报**, 这正是贝叶斯公式教给我们的直觉。)

C14: 由对称性, $P(X > 3) = P(X < 1)$ 意味着 1 与 3 关于 μ 对称: $\mu = \frac{1+3}{2} = 2$ 。

想一想 (对抗) 答案

- **想一想 (对抗 M1):** 错。11 种“和值”并不等可能——和为 7 有 (1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1) 共 6 个样本点, 和为 2 只有 (1, 1) 这 1 个。“等可能”必须对样本点而言; 用“和值”做样本空间, 等可能前提被破坏。
- **想一想 (对抗 M5):** (1) 有放回: $\frac{2}{5} \times \frac{2}{5} = \frac{4}{25}$, 两次**独立**; (2) 不放回: $\frac{2}{5} \times \frac{1}{4} = \frac{1}{10}$, 第二次概率变了, **不独立**。放回与不放回是两个不同模型, 必须先判断清楚再算。
- **想一想 (抽签公平):** 设 $B =$ “第 2 人抽中”, $A_1 =$ “第 1 人抽中”, $A_2 =$ “第 1 人未抽中”。 $P(B) = P(A_1)P(B|A_1) + P(A_2)P(B|A_2) = \frac{1}{3} \times 0 + \frac{2}{3} \times \frac{1}{2} = \frac{1}{3}$ 。第 2 人抽中概率也是 $\frac{1}{3}$ ——**后抽不吃亏, 抽签公平**。
- **想一想 (抽奖决策):** $E(X) = 13 \times 0.1 + (-2) \times 0.9 = 1.3 - 1.8 = -0.5$ 。长期平均每抽一次亏 0.5 元——**期望是决策的尺子: 负期望的抽奖不值得玩**。

第三章 从数据里看世界：统计

章首情境

3.1 抽样：从“全查”到“抽查”

3.1.1 总体与样本：为什么不全查

学习目标

- 能说清总体、个体、样本、样本量的含义；
- 能比较普查与抽样调查的利弊，会按情境选合适的方式；
- 理解样本代表性是统计推断的生命线。

前置复习 (Be Prepared)

- 要检查一箱 100 个灯泡是否合格，最“保险”的做法是什么？最“省钱”的做法是什么？
- 如果全校只有 40 人，你愿意全查还是抽查？如果全校有 40000 人呢？

新知：总体、样本与两种调查

把调查对象的**全体**叫做**总体**，其中每一个调查对象叫做**个体**；从总体中抽取的一部分个体叫做**样本**，样本中包含的个体数叫做**样本量**。

普查（全面调查）对每一个个体都查，结果最准，但费时费力费钱，还可能破坏对象（比如检测灯泡寿命，测完灯泡就废了）。**抽样调查**只查一部分，省时省力，但结果可能“偏”——偏不偏，全看样本有没有**代表性**。

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) 某工厂要检验一批罐头是否密封良好；(2) 学校要统计全校学生本学期的获奖情况；(3) 医生要检查一批疫苗的保存有效期（需开瓶检测）。

解：(1) 抽样调查——罐头数量大，全检耗时长，且抽查部分即可判断整批质量。(2) 普查——获奖名单本就齐全，直接统计又快又全。(3) 抽样调查——开瓶检测会破坏疫苗，只能抽样，用部分样本的检测结果估计整批情况。关键判断：**查了会不会坏、查完用不用得上、查全代价高不高**——三个条件触发任何一个，就偏向抽样。

Try It 3.1：某市要调查全市居民的平均用水量。有人说“查我们小区 500 户就够了”；有人说“应该在全城 20 个区各随机抽 25 户”。你同意谁？为什么？

想一想 (对抗 M11)：小宇说“样本越多越准，我昨天在校门口拦了 500 个人，比分层抽样抽的 50 个人靠谱多了”。你同意吗？如果这 500 人全是去补课的（晚上 9 点以后才离开学校），他们能代表全校同学吗？——**样本量再大，代表不了全体也是白搭；统计的第一原则不是“多”，而是“代表”。**

3.1.2 简单随机抽样：抓阄的科学版

学习目标

- 理解简单随机抽样的定义与特点（等概率、不放回）；
- 掌握抽签法与随机数法两种操作；
- 能判断一个抽样过程是否属于简单随机抽样。

前置复习：掷一枚均匀硬币，“正面”出现的可能性是多少？为什么？

新知：简单随机抽样

一般地，设一个总体含有 N 个个体，从中**逐个不放回**地抽取 n 个个体作为样本 ($n \leq N$)，如果每次抽取时，总体内**每个个体被抽到的概率都相等**，那么这样的抽样方法叫做**简单随机抽样**。

操作工具有两种：**抽签法**（编号 → 搅匀 → 逐个抽，适合总体较小时）和**随机数法**（编号 → 用随机数表或计算器产生随机数 → 选号，适合总体较大时）。

How To：抽签法的四步

- **编号**：给总体中每个个体一个号码 ($1 \sim N$)；
- **制签**：把号码写在大小形状相同的签上；
- **搅匀**：充分混合，保证“每一签被抽到的机会均等”；
- **抽取**：不放回地逐个抽取 n 个签。

例题 3.2

(判断简单随机抽样)

判断下列抽样是否属于简单随机抽样：

(1) 从 50 件产品中任意抽出 10 件检查；(2) 把 1000 名学生编号后，用随机数表选出 80 个号码；(3) 体育老师按“1、3、5、7…”的规律从跑道上选出 20 名学生。

解：(1) 是——每件产品被抽到的概率相等，逐个不放回。(2) 是——随机数表保证等概率。(3) 不是——按固定规律挑选，排在奇数位置的学生注定被选中、偶数位置的学生注定落选，概率不相等。判断标准就一条：**是不是“每个个体机会均等 + 随机决定选谁”**。凡是有“规律”“习惯”“图方便”掺进去的，都不是。

Try It 3.2：从 200 名学生中抽取 20 名进行体质测试。有人建议“按学号 1、2、3、…、20 直接取前 20 名”，这合理吗？为什么？

Q&A

- **问：**为什么要“不放回”？

答：放回的话同一个体可能被抽中多次，样本里的个体不独立，且变相让部分个体“多了一次中签机会”；不放回才能保证每个个体至多被抽中一次、机会均等。

- **问：**简单随机抽样一定“准”吗？

答：不一定。它只能保证**过程公平**，不能保证**结果恰好代表总体**——运气不好也可能抽到一群极端值。所以“过程随机”和“结果准确”是两件事，后一章的统计推断会在“怎么接受这种偶然偏差”上继续讨论。

3.1.3 分层抽样与系统抽样：更聪明的抽法

学习目标

- 理解分层抽样的定义与使用场合；
- 会用公式 $n_i = \frac{N_i}{N} \times n$ 求每层应抽的人数；
- 了解系统抽样的操作，能求抽样间隔与首个体；
- 会结合情境设计合理的抽样方案。

前置复习： $\frac{2}{5} \times 100 = ?$ $\frac{3}{10} \times 80 = ?$

新知：分层抽样

一般地，当总体由**差异明显的几个部分组成**时，把总体分成**互不重叠的层**，然后按照**各层个体数所占的比例**，从各层独立地抽取一定数量的个体，将各层取出的个体合在一起作为样本，这种抽样方法叫做**分层抽样**。

若总体共有 N 个个体，某一层有 N_i 个个体，样本量为 n ，则该层应抽取的个体数为

$$n_i = \frac{N_i}{N} \times n$$

How To：分层抽样的三步

- **分层**：按调查目的把总体分成互不重叠的若干层（层内差异小、层间差异大最合适）；
- **定比**：计算每层人数占总体比例 $p_i = \frac{N_i}{N}$ ；
- **抽取**：每层按比例独立抽取 $n_i = p_i \times n$ 人，各层样本合并。

例题 3.3

（分层抽样求每层人数，对抗 M1、M2）

某中学高一年级共有学生 600 人，其中男生 360 人、女生 240 人。为了解全年级学生的睡眠情况，采用分层抽样抽取样本量为 60 人的样本。问男、女生各应抽取多少人？

解：按性别分层，男、女两层人数分别为 $N_1 = 360$ ， $N_2 = 240$ ，总体 $N = 600$ ，样本量 $n = 60$ 。男生： $n_1 = \frac{360}{600} \times 60 = 0.6 \times 60 = 36$ （人）；女生： $n_2 = \frac{240}{600} \times 60 = 0.4 \times 60 = 24$ （人）。

检验： $36 + 24 = 60$ ✓，且 $\frac{n_1}{n} = \frac{36}{60} = 0.6 = \frac{N_1}{N}$ ✓。答：男生抽 **36** 人、女生抽 **24** 人。

Try It 3.3（反求总体，对抗 M1）：某校按年级分层抽样，样本量为 80。其中高三抽出 20 人，而高三年级共有 400 人。求全校学生总数。（提示： $\frac{20}{80}$ 是多少？全校人数 = 高三人数 ÷ 该比例）

新知：系统抽样

当总体个体数较多且均匀排列时，可以把总体编号后，先随机确定第一个个体，再按固定间隔 k 依次抽取，这样的抽样方法叫做**系统抽样**（等距抽样），间隔 $k = \left\lfloor \frac{N}{n} \right\rfloor$ 。

例题 3.4

（系统抽样）

某学校有 500 名学生，编号为 1 ~ 500。现要从中抽取 50 名学生参加体检，采用系统抽样。

（1）抽样间隔 k 是多少？（2）若在第一段（1 ~ 10 号）中随机抽中了 6 号，写出样本中的前 4 个编号及最后一个编号。

解：（1） $k = \frac{500}{50} = 10$ 。（2）编号依次为 6, 16, 26, 36, …，第 m 个编号为 $6 + (m - 1) \times 10$ ，第 50 个编号为 $6 + 49 \times 10 = 496$ 。答：前 4 个编号是 6, 16, 26, 36，最后一个编号是 **496**。

Try It 3.4：总体 1000 个编号，样本量 100，第一段抽中 8 号。写出抽样间隔、第 3 个编号与最后一个编号。

想一想（对抗 M12）：某工厂流水线每生产 10 件产品就有 1 件次品（次品恰好固定在第 10、20、30、…件）。质检员用系统抽样，间隔 $k = 10$ ，随机起点恰好取到 1。请你算算：他抽到的第 1、2、3 件产品分别是第几件？会是次品吗？——当总体存在周期性规律时，系统抽样的起点必须随机，否则整段样本都会被“带偏”。

本节练习（答案见章末）

- 判断下列抽样方式：从 1000 名学生中随机抽取 100 名进行视力检查（ ）；按年级分层抽取 100 名（ ）；按学号等距抽取 100 名（ ）。
- 某校高一、高二、高三人数分别为 300、240、180，采用分层抽样抽取样本量 60，各年级应各抽多少人？
- 总体 1000 个个体编号 1~1000，用系统抽样抽 100 个，间隔 k 是多少？若第一段抽中 4，写出最后一个编号。
- 调查“全校同学的平均身高”，有人说“直接在操场上拦 200 个人问”就行。这种做法有什么问题？怎样改才更合理？

3.2 频率分布直方图：面积才是频率

3.2.1 从数据到直方图

学习目标

- 掌握分组作频率分布表与频率分布直方图的步骤；
- 理解“小长方形面积 = 频率”、“面积和 = 1”；
- 会由直方图读出频数、频率、样本量。

前置复习：把 10 个数据 3, 1, 4, 1, 5, 9, 2, 6, 5, 3 从小到大排一排，数一数“小于 4 的有几个”。

新知：分组作直方图

把数据分组后，算出每组的频数（落在这组的个数）和频率（频数 ÷ 样本量），就得到**频率分布表**；再画出**频率分布直方图**。

How To：作频率分布直方图五步

- **求极差：**最大值 - 最小值；
- **定组距与组数：**组距 × 组数 ≈ 极差，组数一般取 5~12；
- **分组：**通常“左闭右开”，如 $[150, 155)$ ；
- **列表：**统计每组频数、计算频率；
- **画图：**横轴为数据，纵轴为 $\frac{\text{频率}}{\text{组距}}$ ，每个小区间上画一个矩形。

关键约定（面积即频率）：小长方形的面积 = 组距 × $\frac{\text{频率}}{\text{组距}}$ = 频率。因此所有小长方形的面积之和等于 1。

例题 3.5

(作直方图并读频率, 对抗 M3、M4)

某班 50 名同学的身高数据分组 (单位: cm) 及频数: $[150, 155): 4$, $[155, 160): 8$, $[160, 165): 14$, $[165, 170): 12$, $[170, 175): 8$, $[175, 180): 4$ 。

(1) 求各组的频率及组距; (2) 哪个区间人数最多? (3) 验证 “面积和 = 1”。

解: (1) 组距为 $155 - 150 = 5$ 。各组的频率 = 频数 $\div$ 50:

分组	频数	频率	纵轴高度 $\frac{\text{频率}}{\text{组距}}$
$[150, 155)$	4	0.08	0.016
$[155, 160)$	8	0.16	0.032
$[160, 165)$	14	0.28	0.056
$[165, 170)$	12	0.24	0.048
$[170, 175)$	8	0.16	0.032
$[175, 180)$	4	0.08	0.016

(2) 频数 14 最大, 所以区间 $[160, 165)$ 人数最多。(3) 面积和 = 组距 $\times$ (各纵轴高度之和) = $5 \times (0.016 + 0.032 + 0.056 + 0.048 + 0.032 + 0.016) = 5 \times 0.2 = 1 \checkmark$ (也可直接相加各频率: $0.08 + 0.16 + 0.28 + 0.24 + 0.16 + 0.08 = 1 \checkmark$)。

Try It 3.5: 某样本有 40 人, 体重分组 (单位: kg) 及频数为: $[40, 45): 4$, $[45, 50): 8$, $[50, 55): 12$, $[55, 60): 10$, $[60, 65): 6$ 。求组距、各组频率, 并回答: 体重不低于 55 kg 的频率与人数各是多少?

想一想 (对抗 M3、M4): 有两个直方图。第一个组距为 1, 某矩形高度为 0.30; 第二个组距为 4, 某矩形高度为 0.10。小华说 “第一个的矩形高, 所以频率大”, 对吗? ——算面积: 第一个频率 = $1 \times 0.30 = 0.30$, 第二个频率 = $4 \times 0.10 = 0.40$ 。高度矮不代表频率小, 组距不一样时绝不能拿 “高度” 直接比频率。

3.2.2 由直方图读统计量: 中位数、平均数、众数区间

学习目标

- 会用 “面积平分” 估计中位数;
- 会用 “组中值 $\times$ 频率” 估计平均数;
- 会找出众数所在的区间;
- 能说出平均数与中位数在直方图中的位置关系 (偏态时谁在 “长尾” 一边)。

前置复习: 数集 5, 6, 7, 8, 9 的中位数、平均数分别是多少? (提示: 中位数是排序后中间的数)

新知: 从直方图估计三个统计量

- **中位数**：把直方图的总面积（为 1）左右平分，**面积平分线所在位置**就是中位数的估计值；
- **平均数**：用每组的**组中值**（区间中点）乘该组频率再求和，即 $\bar{x} \approx \sum(\text{组中值}) \times (\text{频率})$ ；
- **众数区间**：频率（面积）最大的那个区间，即“最高的矩形”所在区间。

例题 3.6

（由直方图估计统计量）

利用例题 3.5 的 50 人身高数据，估计中位数、平均数与众数区间。

解：（先列累积频率表，再“折面积”）各频率依次为 0.08, 0.16, 0.28, 0.24, 0.16, 0.08，累积频率依次为 0.08, 0.24, 0.52, 0.76, 0.92, 1.00。

（1）中位数：总面积的一半是 0.5，累积到 $[160, 165)$ 前为 $0.24 < 0.5$ ，落进 $[160, 165)$ 后累积到 $0.52 > 0.5$ ，所以中位数在区间 $[160, 165)$ 内。从 160 往里再走

$$\frac{0.5 - 0.24}{0.28} \times 5 = \frac{0.26}{0.28} \times 5 \approx 4.6$$

所以中位数 $\approx 160 + 4.6 = 164.6$ (cm)。

（2）平均数：组中值分别为 152.5, 157.5, 162.5, 167.5, 172.5, 177.5，

$$\begin{aligned}\bar{x} &\approx 152.5 \times 0.08 + 157.5 \times 0.16 + 162.5 \times 0.28 + 167.5 \times 0.24 + 172.5 \times 0.16 + 177.5 \times 0.08 \\ &= 12.2 + 25.2 + 45.5 + 40.2 + 27.6 + 14.2 = 164.9\end{aligned}$$

（3）频率最大的是 0.28，对应区间 $[160, 165)$ ，即众数区间为 $[160, 165)$ 。

Try It 3.6：用 Try It 3.5 的 40 人体重数据，估计体重的中位数。（提示：累积频率 0.10, 0.30, 0.60, ..., 0.5 落在哪个区间？）

本节练习（答案见章末）

- 某样本 100 人，分组（组距 5）及纵轴高度依次为：0.02, 0.04, 0.06, 0.05, 0.03。求各组频率，并求“样本量 100 时，第 4 组的频数”。
- 一个直方图组距为 2，三个矩形的高度分别是 0.10, 0.15, 0.25。求三个区间的频率，并验证面积和。
- 已知直方图四个矩形的面积（频率）为 0.10, 0.25, 0.35, 0.15，求第五个矩形的频率。
- 用第 1 题数据估计中位数所在的区间。

3.3 百分位数：排名是怎么定的

3.3.1 百分位数的定义与计算

学习目标

- 理解第 p 百分位数的含义；
- 掌握“先排序、再定位”的计算算法（含 i 为整数与非整数两种情形）；
- 会用百分位数解读排名。

前置复习：把 3, 7, 1, 9, 5 从小到大排序，中间的数是几？

新知：第 p 百分位数

一组数据的第 p 百分位数是这样一个值，它使得这组数据中**小于或等于它的数据约占 $p\%$** 。它用于排名比较（“超过 90% 的同学”），与分数本身的数值无关。

How To：计算第 p 百分位数（排序 → 定位）

- **排序：**把 n 个数据从小到大排好；
- **算位：**计算 $i = \frac{p}{100} \times n$ ；
- **分情形：**
 - 若 i **不是整数**，取大于 i 的最小整数 j ，则**第 j 个数据**就是第 p 百分位数；
 - 若 i **是整数**，取**第 i 个与第 $i + 1$ 个数据的平均数**。

例题 3.7

（求百分位数，对抗 M5、M6）

某班 20 名同学一次数学测验的成绩如下（未排序）：

52, 61, 45, 70, 58, 83, 67, 49, 72, 88, 60, 75, 95, 63, 51, 78, 66, 82, 90, 54

求第 72 百分位数与第 80 百分位数。

解：第一步必须先**排序**：45, 49, 51, 52, 54, 58, 60, 61, 63, 66, 67, 70, 72, 75, 78, 82, 83, 88, 90, 95

(1) 第 72 百分位数： $i = \frac{72}{100} \times 20 = 14.4$ ，不是整数 → 取大于 14.4 的最小整数 15，即第 15 个数据，为 **78**。(2) 第 80 百分位数： $i = \frac{80}{100} \times 20 = 16$ ，**恰为整数** → 取第 16 个与第 17 个数据的平均数： $\frac{82 + 83}{2} = 82.5$ 。

Try It 3.7：用例题 3.7 的 20 个数据，求第 60 百分位数与第 90 百分位数。（提示： $0.6 \times 20 = 12$ 是整数； $0.9 \times 20 = 18$ 也是整数）

想一想（对抗 M5）：小丽没排序就算第 70 百分位数：数据 2, 8, 3, 9, 1, 7 有 6 个， $0.7 \times 6 = 4.2$ ，她直接数第 5 个（9），对吗？——**先排序**：1, 2, 3, 7, 8, 9，第 5 个是 8。乱序数据不排序就定位，结果差之千里。

3.3.2 四分位数与箱线图

学习目标

- 掌握下四分位数 Q_1 、中位数、上四分位数 Q_3 的求法；
- 会用“五数概括”（最小值、 Q_1 、中位数、 Q_3 、最大值）作箱线图；
- 会读箱线图判断分布、找离群点。

前置复习：一组数据排序后，第 50 百分位数叫什么？它和第 25、第 75 百分位数是什么关系？

新知：四分位数与五数概括

- 第 25 百分位数叫**下四分位数**，记 Q_1 ；
- 第 50 百分位数就是**中位数**；
- 第 75 百分位数叫**上四分位数**，记 Q_3 。

由最小值、 Q_1 、中位数、 Q_3 、最大值这五个数构成的概括叫**五数概括**，画成图就是**箱线图**：中间一个箱子从 Q_1 拉到 Q_3 ，箱子正中的线是中位数，左右两根“须”分别伸到最小值和最大值。

例题 3.8

（求五数概括）

用例题 3.7 的 20 个成绩数据，求五数概括，并判断 95 分是否像是离群点。

解：排序后的数据同例题 3.7。

- 最小值：45；最大值：95；
- Q_1 （第 25 百分位）： $i = 0.25 \times 20 = 5$ （整数） $\rightarrow \frac{54 + 58}{2} = 56$ ；
- 中位数（第 50 百分位）： $i = 0.5 \times 20 = 10$ （整数） $\rightarrow \frac{66 + 67}{2} = 66.5$ ；
- Q_3 （第 75 百分位）： $i = 0.75 \times 20 = 15$ （整数） $\rightarrow \frac{78 + 82}{2} = 80$ 。

五数概括为：45, 56, 66.5, 80, 95。箱子从 56 到 80，中位线在 66.5，箱子内集中了约一半的同学；95 明显远离 $Q_3 = 80$ ，可看作可能的**离群点**（还需结合背景判断）。

Try It 3.8：数据 3, 5, 7, 8, 10, 12, 15, 18, 20, 24（已排序），求 Q_1 、中位数、 Q_3 。（提示： $n = 10$, $0.25 \times 10 = 2.5$ 非整数）

本节练习（答案见章末）

- 未排序数据：3, 7, 1, 9, 5, 8, 2, 6, 4, 10。求第 60 百分位数。
- 数据 12, 15, 18, 20, 25（已排序）。求 Q_1 、中位数、 Q_3 及五数概括。
- 某次考试你的成绩处于第 90 百分位数。这句话说明你大约超过了多少比例的同学？如果你的成绩恰好是中位数，又说明什么？

3.4 中心与离散：数据稳不稳

3.4.1 平均数、中位数、众数：三个“中心”的三副面孔

学习目标

- 掌握平均数、中位数、众数的计算与含义；
- 理解极端值对三者不同的影响；
- 能按情境选对“中心”指标。

前置复习：求 2, 3, 3, 5, 7 的平均数、中位数、众数。

新知：三个中心指标

- **平均数** $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ ：数据的“算术中心”，**对极端值敏感**；
- **中位数**：排序后中间位置的数，**抗极端值**；
- **众数**：出现次数最多的数，只看“最常见”。

适用场景铁律：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）。

例题 3.9

（极端值的陷阱，对抗 M9、M10）

某公司 10 名员工，9 人月薪 3000 元，1 人月薪 30000 元。

（1）求平均月薪与月薪中位数；（2）公司招聘广告写“平均月薪 5700 元”，为何有误导性？该用什么指标描述“大多数员工”的收入？

解：（1）平均数 $= \frac{9 \times 3000 + 30000}{10} = \frac{27000 + 30000}{10} = 5700$ （元）；中位数：排序后第 5、6 位都是 3000，故中位数 $= 3000$ （元）。（2）平均月薪 5700 元是被 1 个 30000 元“拉高”的，10 人中有 9 人实际只有 3000 元——“平均”严重偏离了大多数人的现实。描述“大多数员工”的收入，应该用**中位数 3000 元**。

Try It 3.9：某班 30 人，29 人考了 60 分，1 人考了 100 分。求平均分与中位数。老师想对“大多数同学”说句话，该引用哪个数字？

想一想（对抗 M10）：奶茶店老板要决定下个月多进哪种口味的原料，该看顾客口味的平均数、中位数还是众数？全校期末总分排名，该用平均分还是中位数？——“最常见”归众数，“总贡献”归平均数，两件事别混用。

3.4.2 方差与标准差：离“中心”多远

学习目标

- 理解方差的含义（离均平方的平均）；
- 掌握方差与标准差的计算（含简算公式）；
- 掌握方差的两条性质（加 c 不变、乘 k 变 k^2 倍）；
- 能用方差比较两组数据的稳定性。

前置复习：求 1, 2, 3, 4, 5 的平均数。

新知：方差与标准差

设一组数据 $x_1, x_2, \dots, x_n$ ，平均数为 $\bar{x}$ ，则这组数据的**方差**为

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

标准差为方差的算术平方根： $s = \sqrt{s^2}$ 。

- 方差的含义：**数据离平均数的平均“距离”（平方后）**。方差越大，数据越分散、越不稳定；
- 标准差与数据**同单位**，更贴近“平均偏离多远”的直觉；
- 简算公式： $s^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2$ （先算平方和再减平均的平方）。

两条性质：所有数据都加上同一个常数 c ，方差**不变**；所有数据都乘以常数 k ，方差变为原来的 k^2 倍（标准差变为原来的 $|k|$ 倍）。

例题 3.10

（计算方差与标准差）

甲、乙两名运动员 5 次射击成绩（环）如下：

甲：5, 6, 7, 8, 9 乙：1, 5, 7, 9, 13

（1）求两人的平均成绩；（2）求两人的方差与标准差，判断谁更稳定。

解：（1）甲的平均数 $\bar{x}_{\text{甲}} = \frac{5+6+7+8+9}{5} = 7$ ；乙的平均数 $\bar{x}_{\text{乙}} = \frac{1+5+7+9+13}{5} = 7$ 。
两人平均水平相同。（2）甲方差：离均差为 -2, -1, 0, 1, 2，

$$s_{\text{甲}}^2 = \frac{4+1+0+1+4}{5} = \frac{10}{5} = 2$$

乙方差：离均差为 -6, -2, 0, 2, 6，

$$s_{\text{乙}}^2 = \frac{36+4+0+4+36}{5} = \frac{80}{5} = 16$$

标准差： $s_{\text{甲}} = \sqrt{2} \approx 1.41$ ， $s_{\text{乙}} = 4$ 。因为 $s_{\text{甲}} < s_{\text{乙}}$ ，所以甲的成绩**更稳定**。

Try It 3.10：求数据 1, 2, 3, 4, 5 的方差与标准差。

例题 3.11

(方差性质)

已知数据 2, 4, 6, 8, 10。

(1) 求平均数、方差、标准差；(2) 若每个数据都加上 3，方差变吗？(3) 若每个数据都乘以 2，新的方差、标准差是多少？

解：(1) $\bar{x} = 6$ 。离均差 -4, -2, 0, 2, 4,

$$s^2 = \frac{16 + 4 + 0 + 4 + 16}{5} = \frac{40}{5} = 8, \quad s = \sqrt{8} = 2\sqrt{2} \approx 2.83$$

(2) 都加 3 后，离均差完全不变 $\rightarrow$ 方差仍为 8 (数据的“离散程度”没变，只是整体平移了)。(3) 都乘 2：方差 $s'^2 = k^2 s^2 = 2^2 \times 8 = 32$ ，标准差 $s' = |2| \times s = 4\sqrt{2} \approx 5.66$ 。检验：新数据 4, 8, 12, 16, 20，平均数 12，离均差 -8, -4, 0, 4, 8，方差 $= \frac{64 + 16 + 0 + 16 + 64}{5} = \frac{160}{5} = 32 \checkmark$ 。

Try It 3.11: 已知一组数据的方差为 3。若把这组数据每个都乘 -2 再加 1，新数据的方差是多少？

想一想 (对抗 M8): 甲组有 100 个数据、方差为 5；乙组只有 10 个数据、方差也为 5。小文说“甲组 100 个数据，肯定比乙组更不稳定”。对吗？——**方差刻画的是“离平均的离散程度”，与数据个数、数据大小都无关。**两组方差相同，稳定程度就相同。

本节练习 (答案见章末)

- 求数据 3, 5, 7 与 1, 5, 9 的平均数与方差，判断哪组更稳定。
- 数据 4, 6, 8, 10, 12 每个都加 5，方差变不变？原方差是多少？
- 某公司 9 人月薪 3000 元、1 人月薪 30000 元。求平均月薪与月薪中位数，说明描述“大多数员工”该用哪个。
- 已知甲组数据的方差为 4，乙组数据是甲组每个数乘 3 再加 10 得到的，求乙组数据的方差。

本章回顾

概念地图

· 统计：用小样本看大总体 (样本代表性是生命线) · 抽样：普查 (全而贵) vs 抽样调查 (省而可能偏) · 简单随机抽样：逐个等概率 (抽签法 / 随机数法) · 分层抽样： $n_i = \frac{N_i}{N} \times n$ (层内差异小、层间差异大) · 系统抽样：随机起点 + 等距 k (防周期偏差) · 描述统计 · 频率分布直方图：小矩形面积 = 频率，面积和 = 1 · 直方图读统计量：中位数 (面积平分)、平均数 (组中值 $\times$ 频率)、众数区间 (最高矩形) · 百分位数：先排序 $\rightarrow i = \frac{p}{100}n \rightarrow$ 非整数取第 $[i]$ 个，整数取第 $i, i+1$ 平均 · 四分位数： Q_1 (第 25 百分位)、中位数 (第 50)、 Q_3 (第 75) $\rightarrow$ 箱线图五数概括 · 中心与离散：平均数 (总水平，怕极端值) · 中位数 (大多数) · 众

数（最常见）· 方差 $s^2 = \frac{1}{n} \sum (x_i - \bar{x})^2$ （稳不稳）；标准差 $s = \sqrt{s^2}$ （同单位）· 方差性质：加 c 不变，乘 k 乘 k^2

关键公式速查

内容	要点
分层抽样	$n_i = \frac{N_i}{N} \times n$ （人数比，不是层数比）
系统抽样	间隔 $k = \left\lfloor \frac{N}{n} \right\rfloor$ ，首个体随机
直方图	频率 = 组距 $\times$ 高度（面积）；面积和 = 1
直方图估计	中位数：面积平分线；平均数： $\sum(\text{组中值} \times \text{频率})$
百分位数	排序 $\rightarrow i = \frac{p}{100}n \rightarrow$ 非整数取第 $[i]$ 个；整数取第 $i, i+1$ 平均
平均数	$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$
方差	$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n} \sum x_i^2 - \bar{x}^2$
标准差	$s = \sqrt{s^2}$
方差性质	加 c 不变；乘 k 方差乘 k^2 、标准差乘 $ k $
指标选择	总水平用平均、大多数用中位、最常见用众数

分层练习

A 基础层

- 判断：从 500 名学生中随机抽取 50 名进行体质测试，属于哪种抽样方法？说明理由。
- 某校高一 600 人、高二 500 人、高三 400 人，采用分层抽样抽取样本量 90。各年级应各抽多少人？
- 某样本 40 人体重分组 (kg, 组距 5)：[40, 45)：4, [45, 50)：8, [50, 55)：12, [55, 60)：10, [60, 65)：6。求各组频率；估计中位数；求体重不低于 55 kg 的人数（估计）。
- 已排序的 15 个数据：10, 12, 15, 18, 20, 21, 23, 25, 26, 28, 30, 32, 35, 38, 40。求第 80 百分位数。
- 求数据 2, 4, 6, 8, 10 的方差与标准差。
- 某班 30 人，29 人考了 60 分，1 人考了 100 分。求平均分与中位数，说明“平均分”为何有误导性。
- 总体 1000 个个体编号 1 ~ 1000，用系统抽样抽 100 个，第一段抽中 4。写出样本中的最后一个编号。
- 甲组数据方差为 4，乙组数据是甲组每个数乘 2 再加 10 得到的。求乙组数据的方差。

B 应用层

- 某校按年级分层抽样，样本量为 160，其中高二抽出 60 人，高二级共有 450 人。求全校人数。
- 用 A3 的 40 人体重数据（累积频率 \$0.10, 0.30, 0.60, 0.85, 1.00\$）估计第 80 百分位数。
- 甲组 5 个数据方差为 4；乙组数据由甲组每个数加 10 后乘 2 得到，求乙组方差。
- 甲、乙两班数学成绩：甲班平均 85 分、方差 16；乙班平均 82 分、方差 4。评价两个班“谁学得更好、更稳定”。

C 拓展层

- 右偏数据 1, 2, 3, 4, 5, 6, 7, 8, 9, 50。求平均数与中位数，并说明为什么会出现“平均远大于中位数”。
- 数据 2, 5, 6, 7, 8, 9, 10, 11, 12, 30。求五数概括，指出可能的离群点，并解释它如何影响平均数与中位数。

参考答案（章末集中）

Try It 即时练习答案

- **3.1:** 同意第二种。第一种只查“我们小区”，全体 500 户无法代表全市（可能全是年轻人或全是老年人）；第二种按区分层、每层随机抽，兼顾了各区差异，样本更有代表性。**样本量小但代表性强的样本，胜过量大但偏的样本。**
- **3.2:** 不合理。只取前 20 名，学号靠前的同学被“锁定”中签、靠后的必定落选，概率不相等，不是随机抽样。应编号后用抽签法或随机数法抽取。
- **3.3:** $\frac{20}{80} = \frac{1}{4}$ ，即高三占全校的 $\frac{1}{4}$ ，全校人数 $= 400 \div \frac{1}{4} = 1600$ 。
- **3.4:** $k = \frac{1000}{100} = 10$ ；第 3 个编号 $= 8 + 2 \times 10 = 28$ ；最后一个编号 $= 8 + 99 \times 10 = 998$ 。
- **3.5:** 组距 5；频率依次为 \$0.10, 0.20, 0.30, 0.25, 0.15\$；不低于 55kg 的频率 $= 0.25 + 0.15 = 0.40$ ，人数 $= 40 \times 0.40 = 16$ 。
- **3.6:** 累积频率 \$0.10, 0.30, 0.60\$，0.5 落在 [50, 55) 内，中位数 $\approx 50 + \frac{0.5 - 0.30}{0.30} \times 5 = 50 + \frac{0.2}{0.3} \times 5 \approx 53.3$ (kg)。
- **3.7:** 第 60 百分位： $0.6 \times 20 = 12$ (整数) $\rightarrow \frac{70 + 72}{2} = 71$ ；第 90 百分位： $0.9 \times 20 = 18$ (整数) $\rightarrow \frac{88 + 90}{2} = 89$ 。
- **3.8:** Q_1 : $0.25 \times 10 = 2.5 \rightarrow$ 第 3 个，为 7；中位数： $0.5 \times 10 = 5$ (整数) $\rightarrow \frac{8 + 10}{2} = 9$ ； Q_3 : $0.75 \times 10 = 7.5 \rightarrow$ 第 8 个，为 18。

- **3.9:** 平均分 $= \frac{29 \times 60 + 100}{30} = \frac{1840}{30} \approx 61.3$; 中位数 = 60。对“大多数同学”说话应引用中位数 60 (29 人都考了 60 分)。
- **3.10:** $\bar{x} = 3$; 方差 $= \frac{4+1+0+1+4}{5} = 2$; 标准差 $\sqrt{2} \approx 1.41$ 。
- **3.11:** 乘 -2 : 方差乘 $(-2)^2 = 4 \rightarrow$ 新方差 $= 3 \times 4 = 12$; 加 1 不改变方差。故新方差为 12。

各节练习答案

- **3.1 节:** 1. 简单随机抽样、分层抽样、系统抽样。2. 高一 $300/720 \times 60 = 25$, 高二 $240/720 \times 60 = 20$, 高三 $180/720 \times 60 = 15$ (和为 60 $\checkmark$)。3. $k = 10$, 最后编号 $4 + 99 \times 10 = 994$ 。4. 操场上拦人只代表“当时在操场活动”的学生 (偏样), 且未随机。应编号后分层 (如按年级) 随机抽取。
- **3.2 节:** 1. 频率 = 高度 $\times$ 组距: 0.10, 0.20, 0.30, 0.25, 0.15; 第 4 组频率 0.25, 频数 $= 100 \times 0.25 = 25$ 。2. 频率 = 0.20, 0.30, 0.50, 面积和 $= 0.2 + 0.3 + 0.5 = 1 \checkmark$ 。3. 第五个频率 $= 1 - (0.10 + 0.25 + 0.35 + 0.15) = 0.15$ 。4. 累积频率 \$0.10, 0.30, 0.60, 0.5\$ 落在第 3 组 (对应区间 [50, 55)\$, 若以 A 数据计即第 3 个分组)。
- **3.3 节:** 1. 排序 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, $0.6 \times 10 = 6$ (整数) $\rightarrow \frac{6+7}{2} = 6.5$ 。2. $Q_1: 0.25 \times 5 = 1.25 \rightarrow$ 第 2 个, 15; 中位数: $0.5 \times 5 = 2.5 \rightarrow$ 第 3 个, 18; $Q_3: 0.75 \times 5 = 3.75 \rightarrow$ 第 4 个, 20。五数概括 12, 15, 18, 20, 25。3. 大约超过 90% 的同学; 若是中位数, 则约一半同学在你前面、一半在你后面。
- **3.4 节:** 1. 3, 5, 7: 平均 5, 方差 $= \frac{4+0+4}{3} = \frac{8}{3}$; 1, 5, 9: 平均 5, 方差 $= \frac{16+0+16}{3} = \frac{32}{3}$ 。第一组更稳定。2. 加 5 方差不变; 原方差: 平均 8, 离均 $-4, -2, 0, 2, 4$, 方差 $= \frac{16+4+0+4+16}{5} = 8$ 。3. 平均 $= \frac{9 \times 3000 + 30000}{10} = 5700$ (元), 中位数 3000 元; 描述“大多数员工”用中位数。4. 乘 3 $\rightarrow$ 方差 $= 4 \times 9 = 36$; 加 10 不变。故乙组方差为 36。

A1: 简单随机抽样——编号后随机抽取, 每个学生被抽到的机会均等。

A2: 高一 $600/1500 \times 90 = 36$; 高二 $500/1500 \times 90 = 30$; 高三 $400/1500 \times 90 = 24$ (和 90 $\checkmark$)。

A3: 频率 \$0.10, 0.20, 0.30, 0.25, 0.15\$; 中位数: 累积 \$0.10, 0.30, 0.60, 0.5\$ 落 [50, 55), $\approx 50 + \frac{0.2}{0.3} \times 5 \approx 53.3$ (kg); 不低于 55 kg: $40 \times 0.40 = 16$ (人)。

A4: $0.8 \times 15 = 12$ (整数) $\rightarrow \frac{32+35}{2} = 33.5$ 。

A5: $\bar{x} = 6$; 方差 $= \frac{16+4+0+4+16}{5} = 8$; 标准差 $2\sqrt{2} \approx 2.83$ 。

A6: 平均 ≈ 61.3 分、中位数 60 分。一个 100 分把平均拉高, 而 29 人的真实水平是 60 分, 故用中位数更真实。

A7: $4 + 99 \times 10 = 994$ 。

A8: 乘 2 $\rightarrow$ 方差 $4 \times 4 = 16$; 加 10 不变。故为 **16**。

B9: $\frac{60}{160} = \frac{3}{8}$, 全校 $= 450 \div \frac{3}{8} = 450 \times \frac{8}{3} = 1200$ 。

B10: 第 80 百分位: 累积频率 $0.60 < 0.8 < 0.85$, 落 $[55, 60)$, $\approx 55 + \frac{0.8 - 0.60}{0.25} \times 5 = 55 + 4 = 59$ (kg)。

B11: 加 10 不变, 乘 2 $\rightarrow$ 方差 $= 4 \times 4 = 16$ 。

B12: 甲班平均分更高; 但乙班方差小得多, 说明乙班成绩更稳定、两极分化小。评价不能只看平均分, 要“平均 + 方差”一起看。

C13: 平均 $= \frac{45 + 50}{10} = 9.5$, 中位数 $= \frac{5 + 6}{2} = 5.5$ 。一个 50 把平均数拉到 9.5, 而一半数据不超过 5.5——右偏 (长尾在右) 时平均数被极端值拉高、中位数抗极端值, 故 $\bar{x} > M_e$ 。

C14: 排序后 2, 5, 6, 7, 8, 9, 10, 11, 12, 30。 $Q_1: 0.25 \times 10 = 2.5 \rightarrow$ 第 3 个, 6; 中位数: $\frac{8 + 9}{2} = 8.5$; $Q_3: 0.75 \times 10 = 7.5 \rightarrow$ 第 8 个, 11。五数概括: 2, 6, 8.5, 11, 30。30 明显远离 $Q_3 = 11$, 是可能的离群点; 它把平均数拉到 $\bar{x} = 10$, 却几乎不影响中位数 8.5。

第四章 两个变量是“伙伴”还是“路人”： 成对数据的统计分析

章首情境

4.1 相关关系：是“伙伴”还是“路人”

4.1.1 相关关系与函数关系

学习目标

- 能区分函数关系、相关关系与无关系；
- 理解相关关系“有趋势但不确定”的特点；
- 理解“相关 $\neq$ 因果”并会找背后的共同原因。

前置复习：正方形的边长 a 与面积 S 的关系是什么？给定边长能确定面积吗？

新知：三种关系

两个变量之间的关系大致分三类：

- **函数关系：**一个变量完全确定另一个变量，如电压 U 、电流 I 、电阻 R 满足 $U = IR$ (R 固定时 U 由 I 完全确定)；
- **相关关系：**一个变量变化时，另一个变量有**变化趋势**，但同一个 x 可以对应多个 y ，如身高与体重；
- **无关系：**两变量看不出任何相互影响，如今天的气温与明天的彩票开奖号码。

例题 4.1

(判断关系类型，对抗 M1)

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) 电压一定时，电路中的电流与电阻；(2) 某班同学的身高与体重；(3) 某市七月的冰淇淋销量与溺水人数；(4) 某天凌晨的气温与当天彩票开奖号码。

解：(1) 函数关系——欧姆定律 $I = \frac{U}{R}$ ，给定电阻，电流被完全确定。(2) 相关关系——身高高的人体重通常也重（趋势），但同一身高对应不同体重（不确定）。(3) 相关关系——两者同步上涨，但**没有谁导致谁**：背后是共同原因“夏天到了”（天热大家吃冰淇淋，天热大家爱下水游泳）。(4) 无关系——气温与彩票开奖互不影响。

Try It 4.1: 判断：圆的半径与面积（ ）；学习时长与数学成绩（ ）；你每天的心情与你同桌的鞋码（ ）。

想一想（呼应第 3 章）：第 3 章说过“样本代表性是统计推断的生命线”。现在想：“身高中等的人体重不一定中等”——这正是相关关系的不确定性。如果样本没有代表性（比如只统计了篮球队队员），你看到的“身高—体重”关系还真实吗？——**关系是真是假，先看样本不可信。**

4.1.2 散点图：先看图再算数

学习目标

- 会画散点图；
- 会从散点图读出方向、形状、强度、离群点；
- 记住“任何回归计算前先画散点图”的铁律。

前置复习：平面直角坐标系里，点 (2, 3) 怎么找？横轴竖轴各表示什么？

新知：散点图与读图四问

把每对数据 (x_i, y_i) 作为平面直角坐标系中的点描出来，得到的图叫**散点图**。读图看四件事：

- **方向：**点云从左下斜向右上 → **正相关**；从左上斜向右下 → **负相关**；横向铺开 → 无明显相关；
- **形状：**点是否大致排成一条直线？若是“直线”才有资格谈线性相关；
- **强度：**点离一条直线近不近？越近，线性关系越强；
- **离群点：**有没有个别点偏离整体云团？它可能藏着故事（记错数据？特殊人群?）。

铁律：任何回归计算之前，先画散点图。数据明明是弯的，还硬套一条直线，“相关系数”再漂亮也是白搭。

例题 4.2

(读散点图)

某班 8 名同学的数学成绩 x (分) 与物理成绩 y (分) 的散点图 (示意) 中, 点的云团从左下斜向右上, 且点都紧挨着一条直线; 只有一名同学的点明显“掉队”——数学 92 分而物理只有 40 分。

(1) 指出这组数据的相关方向; (2) 指出其大致强度; (3) 那个“掉队”的点叫什么? 可能是什么原因?

解: (1) 点云从左下到右上 → **正相关**, 数学成绩高的同学物理成绩通常也高。(2) 点紧挨直线 → **线性相关较强** (r 应接近 1)。(3) 这个“掉队”的点叫**离群点**。可能原因: 该同学物理考试当天生病发挥失常、或数据录入错误——需要核对, 不能直接删掉。

Try It 4.2: 某研究统计了 10 名高三学生“每天睡眠小时数”与“次日课堂犯困次数”, 散点图的点云从左上斜向右下。判断这组数据的相关方向, 并说明“睡得多所以犯困少”这个说法成立吗? (提示: 只凭相关, 能不能断言因果?)

想一想 (对抗 M2): 有一组数据满足 $y = (x - 3)^2$ (抛物线), x 取 1, 2, 3, 4, 5。它的散点图是标准的 U 形。这时如果硬算线性相关系数, 会得到 $r \approx 0$ 。小强说: “ $r \approx 0$, 所以这两个变量没关系。”对吗? —— r 只刻画“**线性**”相关; U 形关系是“**曲线相关**”, r 根本量不出来。 $|r|$ 接近 0 只能说“**没有线性相关**”, 不能说“**没有关系**”。

4.1.3 相关系数 r : 方向与强度

学习目标

- 掌握相关系数 r 的计算公式;
- 理解 r 的性质: $-1 \leq r \leq 1$, 符号定方向, $|r|$ 定强度;
- 会用 r 判断线性相关的强弱。

前置复习: 计算 $(1 - 2)^2$ 、 $(3 - 3)^2$ 、 $(5 - 4)^2$ 各等于多少。

新知: 相关系数

设 n 组数据 $(x_1, y_1), \dots, (x_n, y_n)$, 样本均值分别为 $\bar{x}$ 、 $\bar{y}$, 则**相关系数**

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

性质:

- $-1 \leq r \leq 1$ 恒成立;
- $r > 0$: 正相关; $r < 0$: 负相关——**符号决定方向**;
- $|r|$ 越接近 1, 线性关系越强; 越接近 0 越弱—— $|r|$ **决定强度**;
- $r = \pm 1$: 所有点恰好一条直线上; $r = 0$: 无线性相关 (仍可能有曲线相关)。

经验约定： $|r| \geq 0.75$ 强相关； $0.3 \leq |r| < 0.75$ 中等相关； $|r| < 0.3$ 弱相关（近似无线性相关）。

例题 4.3

（计算相关系数）

4 名同学的数学周测成绩 x 与物理周测成绩 y 如下：

x	1	2	3	4
y	2	3	5	4

求这组数据的相关系数 r 。

解： $\bar{x} = \frac{1+2+3+4}{4} = 2.5$ ， $\bar{y} = \frac{2+3+5+4}{4} = 3.5$ 。

$x_i - \bar{x}$	-1.5	-0.5	0.5	1.5
$y_i - \bar{y}$	-1.5	-0.5	1.5	0.5
乘积	2.25	0.25	0.75	0.75

$\sum (x_i - \bar{x})(y_i - \bar{y}) = 2.25 + 0.25 + 0.75 + 0.75 = 4$ ； $\sum (x_i - \bar{x})^2 = 2.25 + 0.25 + 0.25 + 2.25 = 5$ ； $\sum (y_i - \bar{y})^2 = 2.25 + 0.25 + 2.25 + 0.25 = 5$ 。

$$r = \frac{4}{\sqrt{5} \times \sqrt{5}} = \frac{4}{5} = 0.8$$

答： $r = 0.8$ ，属于**强正相关**（ $|r| \geq 0.75$ ）。

Try It 4.3：数据 $(1, 3), (2, 5), (3, 7), (4, 9)$ 。求 r 。（提示：这组数据恰好在直线 $y = 2x + 1$ 上）

想一想（对抗 M3）：甲研究得 $r = 0.9$ ，乙研究得 $r = -0.95$ 。小刚说“甲的相关更强，因为 0.9 比 -0.95 大”。对吗？——**比强度要比 $|r|$** ： $|-0.95| = 0.95 > 0.9$ ，乙的线性关系更强。负相关不是“坏”相关，只是方向相反。

本节练习（答案见章末）

- 判断下列各组变量的关系类型：圆的半径与周长（ ）；每天的学习时长与数学成绩（ ）；气温与冷饮销量（ ）。
- 某研究显示“阅读量高的学生数学成绩也高”。能据此说“多读书能提高数学成绩”吗？说明理由。
- 三幅散点图：甲，点云从左下到右上且很密集；乙，点云横向铺开；丙，点云从左上到右下且很稀疏。分别判断相关方向与大致强度。
- 某散点图的点几乎都贴成一条向右上倾斜的直线。 r 大约接近多少？

4.2 回归：用一条线做预测

4.2.1 回归直线与最小二乘

学习目标

- 理解回归直线 $\hat{y} = bx + a$ 的含义 ($\hat{y}$ 是预测值);
- 掌握最小二乘的思想与 b 、 a 的计算公式;
- 掌握核心性质：回归直线必过样本中心 $(\bar{x}, \bar{y})$ 。

前置复习：一次函数 $y = kx + m$ 的图象是什么？ k 表示什么？

新知：回归直线与最小二乘

设样本点为 (x_i, y_i) ，我们希望找一条直线 $\hat{y} = bx + a$ 来“预测” y 。点 i 的残差为

$$e_i = y_i - \hat{y}_i = y_i - (bx_i + a)$$

它是观测值减去预测值的纵向距离。**最小二乘法**就是选 a 、 b ，使残差平方和

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - bx_i - a)^2$$

最小。此时

$$b = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad a = \bar{y} - b\bar{x}$$

核心性质：回归直线必过样本中心点 $(\bar{x}, \bar{y})$ ——把 $x = \bar{x}$ 代入得 $\hat{y} = b\bar{x} + a = \bar{y}$ ✓。

How To：求回归方程五步

- 算 $\bar{x}$ 、 $\bar{y}$;
- 列每对 $(x_i - \bar{x})$ 、 $(y_i - \bar{y})$;
- 求分子 $S_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y})$ 与分母 $S_{xx} = \sum (x_i - \bar{x})^2$;
- 算 $b = \frac{S_{xy}}{S_{xx}}$ ，再算 $a = \bar{y} - b\bar{x}$;
- 写出 $\hat{y} = bx + a$ ，并用**样本中心检验**是否在线上。

例题 4.4

(求回归方程并预测)

某商场统计了 5 个月的广告费 x (万元) 与销售额 y (万元):

x	2	3	4	5	6
y	28	33	39	42	48

(1) 求回归方程 $\hat{y} = bx + a$; (2) 预测广告费为 7 万元时的销售额。

解: (1) $\bar{x} = \frac{2+3+4+5+6}{5} = 4$, $\bar{y} = \frac{28+33+39+42+48}{5} = 38$ 。

$x_i - \bar{x}$	-2	-1	0	1	2
$y_i - \bar{y}$	-10	-5	1	4	10
乘积	20	5	0	4	20
$(x_i - \bar{x})^2$	4	1	0	1	4

$$S_{xy} = 20 + 5 + 0 + 4 + 20 = 49, S_{xx} = 4 + 1 + 0 + 1 + 4 = 10,$$

$$b = \frac{49}{10} = 4.9, \quad a = 38 - 4.9 \times 4 = 38 - 19.6 = 18.4$$

回归方程为 $\hat{y} = 4.9x + 18.4$ 。检验: 把中心 $(4, 38)$ 代入, $4.9 \times 4 + 18.4 = 19.6 + 18.4 = 38 \checkmark$, 直线过样本中心。

(2) 当 $x = 7$ 时, $\hat{y} = 4.9 \times 7 + 18.4 = 34.3 + 18.4 = 52.7$ (万元)。答: 预测销售额约为 **52.7 万元**。

Try It 4.4: 数据 $(1, 1), (2, 3), (3, 4), (4, 6)$ 。求回归方程, 并预测 $x = 5$ 时的 $\hat{y}$ 。(提示: $\bar{x} = 2.5$, $\bar{y} = 3.5$)

想一想 (对抗 M4): 例题 4.4 的回归线 $\hat{y} = 4.9x + 18.4$ 过原点吗? 过任何一个原始数据点吗? ——不过原点 ($x = 0$ 时 $\hat{y} = 18.4 \neq 0$), 也不一定过某个数据点 ($(2, 28)$ 处预测值是 28.2)。**回归直线唯一保证经过的是样本中心 $(\bar{x}, \bar{y})$ 。**初中“正比例函数过原点”的印象在这里不适用。

4.2.2 残差与外推

学习目标

- 会计算残差 $e_i = y_i - \hat{y}_i$ 并判断哪个点拟合最差;
- 理解残差是“纵向距离”, 不是“错误”;
- 理解内插可预测、外推不可靠。

前置复习: 把 $x = 5$ 代入 $\hat{y} = 4.9x + 18.4$, 得到多少? 若实际观测值是 42, 预测误差是多少?

新知: 残差与外推

- **残差 $e_i = y_i - \hat{y}_i$:** 点在回归线上方残差为正 (线“低估”了它), 下方为负。 $|e_i|$ 越大, 这个点偏离回归线越远, 即该点**拟合越差**。
- **内插与外推:** 回归直线只对**样本数据 x 的取值范围内**做预测才可靠 (内插); 把 x 取到范围之外去预测 (外推), 因为没有数据支撑, 很可能离谱。

例题 4.5

(残差分析, 对抗 M11)

利用例题 4.4 的回归方程 $\hat{y} = 4.9x + 18.4$:

(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 $x = 10$ 时的销售额, 合理吗?

解: (1) 预测值依次为 \$28.2, 33.1, 38, 42.9, 47.8\$ (如 $x=2$ 时 $\hat{y} = 4.9 \times 2 + 18.4 = 28.2$), 残差为

$$28 - 28.2 = -0.2, \quad 33 - 33.1 = -0.1, \quad 39 - 38 = 1, \quad 42 - 42.9 = -0.9, \quad 48 - 47.8 = 0.2$$

(2) $|e_i|$ 分别为 \$0.2, 0.1, 1, 0.9, 0.2\$, 最大的是 $|-0.9|$, 即 $x = 5$ 这个点拟合最差 (但也只偏了 0.9 万元, 整体拟合不错)。(3) $x = 10$ 已超出样本范围 $x \in [2, 6]$, 属**外推**, 没有数据支撑, 不合理。

Try It 4.5: 回归方程 $\hat{y} = 1.6x - 0.5$, 样本点 $(1, 1), (2, 3), (3, 4), (4, 6)$ 。求各点残差, 指出拟合最差的点。

想一想 (对抗 M12): 若某组数据的 $r = -0.8$, 它的回归直线斜向哪边? ——回归斜率 b 与 r 同号: $r < 0$ 则 $b < 0$, 直线从左到右下降。 r 的正负和 b 的正负不可能相反, 这正是“方向一致”的体现。

本节练习 (答案见章末)

- 已知回归直线过样本中心 $(3, 7)$, 斜率 $b = 2$ 。求 a 并写出回归方程。
- 样本数据中 $x \in [40, 60]$, 回归方程 $\hat{y} = 0.6x + 10$ 。用它对 $x = 100$ 做预测合理吗? 为什么?
- 回归方程 $\hat{y} = 2x + 1$, 观测点 $(1, 2), (2, 5), (3, 7), (4, 8)$ 。求各残差, 指出拟合最差的点。
- 某组数据 $r = -0.8$, 画出回归直线的大致方向 (上升/下降), 并说明理由。

4.3 独立性检验: 证据够不够

4.3.1 列联表与期望频数

学习目标

- 会完善 2×2 列联表 (三线合计校验);
- 会按公式求期望频数 $E = \frac{\text{行合计} \times \text{列合计}}{\text{总人数}}$;
- 理解期望频数代表“假设两变量无关”时应该出现的频数。

前置复习： $200 \times 110 \div 200 = ?$ $100 \times 90 \div 200 = ?$

新知：列联表与期望频数

把两个分类变量的数据按两个类别交叉排成的表叫**列联表**（ 2×2 列联表）。制作时**行合计、列合计、总合计必须三线对应**。

在“两变量无关”的假设（称为**原假设**）下，某一格的**期望频数**应为

$$E = \frac{\text{该格所在行的合计} \times \text{该格所在列的合计}}{\text{总人数}}$$

它的直观含义：如果锻炼与近视无关，那么“经常锻炼且近视”的比例就应该等于“经常锻炼的比例 $\times$ 近视的比例”（乘法原理），据此折算的人数就是期望值。

例题 4.6

（完善列联表与求期望，对抗 M8、M9）

某商场就“是否支持新的会员方案”对 100 名顾客作调查，得到不完整的 2×2 列联表（频数）：

	支持	反对	合计
方案 A	30	20	50
方案 B			
合计	70		100

（1）补全列联表；（2）求四格的期望频数。

解：（1）先看合计线：支持列合计 70，方案 A 支持 30，故方案 B 支持 $= 70 - 30 = 40$ ；方案 B 合计 $= 100 - 50 = 50$ ，故方案 B 反对 $= 50 - 40 = 10$ ；反对列合计 $= 20 + 10 = 30$ ，且 $30 + 70 = 100 \checkmark$ （三线校验通过）。

	支持	反对	合计
方案 A	30	20	50
方案 B	40	10	50
合计	70	30	100

（2）在原假设（“方案与态度无关”）下： $E(\text{方案 A, 支持}) = \frac{50 \times 70}{100} = 35$ ； $E(\text{方案 A, 反对}) = \frac{50 \times 30}{100} = 15$ ； $E(\text{方案 B, 支持}) = \frac{50 \times 70}{100} = 35$ ； $E(\text{方案 B, 反对}) = \frac{50 \times 30}{100} = 15$ 。

Try It 4.6：某校 70 名学生，其中男生 40 人、女生 30 人；会打篮球的共 35 人（其中男生 20 人）。完善列联表，并求“男生且会打篮球”这一格的期望频数。

4.3.2 χ^2 统计量与结论

学习目标

- 掌握 χ^2 统计量公式: $\chi^2 = \sum \frac{(\text{观测} - \text{期望})^2}{\text{期望}}$;
- 会用 χ^2 与临界值比较并下规范结论;
- 掌握比较方向: χ^2 大于临界值才拒绝“无关”。

前置复习: $\frac{(40 - 55)^2}{55}$ 等于多少? (提示: 先算 $40 - 55$)

新知: χ^2 统计量与比较规则

$$\chi^2 = \sum \frac{(\text{观测频数} - \text{期望频数})^2}{\text{期望频数}}$$

对 2×2 列联表, 四格全部代入求和即得 χ^2 。判断规则 (临界值表): 在显著性水平 $\alpha = 0.05$ 下临界值 $k = 3.841$; $\alpha = 0.01$ 下 $k = 6.635$ 。

- 若 $\chi^2 > k$: **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 ;
- 若 $\chi^2 \leq k$: **不拒绝原假设**, 没有充分证据认为两变量有关联。

例题 4.7

(计算 χ^2 并下结论, 对抗 M6)

为研究锻炼习惯与近视是否有关, 调查了 200 名学生, 得到列联表:

	近视	不近视	合计
经常锻炼	40	60	100
很少锻炼	70	30	100
合计	110	90	200

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

解: (1) $E(\text{经常, 近视}) = \frac{100 \times 110}{200} = 55$; $E(\text{经常, 不近视}) = \frac{100 \times 90}{200} = 45$; $E(\text{很少, 近视}) = 55$; $E(\text{很少, 不近视}) = 45$ 。(2)

$$\begin{aligned}\chi^2 &= \frac{(40 - 55)^2}{55} + \frac{(60 - 45)^2}{45} + \frac{(70 - 55)^2}{55} + \frac{(30 - 45)^2}{45} = \frac{225}{55} + \frac{225}{45} + \frac{225}{55} + \frac{225}{45} \\ &= 2 \times \frac{225}{55} + 2 \times \frac{225}{45} \approx 2 \times 4.09 + 2 \times 5 = 18.18\end{aligned}$$

(3) $\chi^2 \approx 18.18$ 。与 $\alpha = 0.05$ 的临界值 3.841 比: $18.18 > 3.841$, **拒绝原假设**; 与 $\alpha = 0.01$ 的临界值 6.635 比: $18.18 > 6.635$, 仍拒绝。结论 (规范表达): **在犯错误的概率不超过 0.01 的前提下, 认为锻炼习惯与近视状况有关联**——很少锻炼的学生中近视的比例明显更高。

Try It 4.7 (对抗 M6): 某调查算得 $\chi^2 = 5.02$, 临界值取 3.841。是否拒绝“两变量无关”? 用规范语言下结论。

想一想 (对抗 M7): 另一项调查算得 $\chi^2 = 2.1 < 3.841$ 。小亮说“所以这两个变量没有关系”。这么说对吗? ——不对。规范说法是: **没有充分证据认为两变量有关联**。“证据不足”不等于“证明了无关”, 就像“没找到外星人”不等于“证明没有外星人”。结论的力度要和证据的强度匹配。

本节练习 (答案见章末)

- 两家奶茶店各调查了 60 名、40 名顾客的满意度, 得列联表:

	满意	不满意	合计
甲店	45	15	60
乙店	25	15	40
合计	70	30	100

求期望频数, 计算 χ^2 , 并在 $\alpha = 0.05$ (临界值 3.841) 下下结论。

- 某调查算得 $\chi^2 = 5.02$, 临界值 3.841。是否拒绝“两变量无关”? 为什么?
- 某调查算得 $\chi^2 = 2.1 < 3.841$ 。用规范语言下结论 (注意: 不能写“无关”)。

本章回顾

概念地图

· 成对数据统计分析: 回答“两变量有没有关系、多强、能否预测” · 关系类型: 函数 (确定) · 相关 (趋势不确定) · 无关 · 相关 $\neq$ 因果 (共同原因可能藏在背后) · 散点图: 方向 / 形状 / 强度 / 离群点 (先看图再算数) · 相关系数 r : 符号定方向、 $|r|$ 定强度; $r = 0$ 只说明无**线性**相关 · 回归直线: $\hat{y} = bx + a$ (最小二乘: 残差平方和最小) · $b = \frac{S_{xy}}{S_{xx}}$, $a = \bar{y} - b\bar{x}$, 必过样本中心 $(\bar{x}, \bar{y})$ · 残差 $e_i = y_i - \hat{y}_i$ (纵向距离); 内插可靠、外推危险 · 独立性检验: 列联表 $\rightarrow$ 期望 $E = \frac{\text{行} \times \text{列}}{\text{总}}$ $\rightarrow \chi^2 = \sum \frac{(O - E)^2}{E}$ $\rightarrow$ 大则拒绝“无关” · 规范结论: 在犯错误概率不超过 α 的前提下认为有关联 / 没有充分证据

关键公式速查

内容	要点
相关系数	$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} \sqrt{\sum (y_i - \bar{y})^2}}$
回归斜率	$b = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$
回归截距	$a = \bar{y} - b\bar{x}$ (直线必过 $(\bar{x}, \bar{y})$)
回归方程	$\hat{y} = bx + a$; 只内插、不外推
残差	$e_i = y_i - \hat{y}_i$
期望频数	$E = \frac{\text{行合计} \times \text{列合计}}{\text{总人数}}$
χ^2	$\chi^2 = \sum \frac{(\text{观测} - \text{期望})^2}{\text{期望}}$; $\chi^2 > k$ 才拒绝“无关”
临界值	$\alpha = 0.05$ 时 $k = 3.841$; $\alpha = 0.01$ 时 $k = 6.635$

分层练习

A 基础层

- 判断下列各组变量属于函数关系、相关关系还是无关系：年龄与身高（ ）；电压与电流（电阻一定）（ ）；天气与彩票号码（ ）。
- 有三幅散点图的描述：甲，点云从左下到右上且很密集；乙，点云横向铺开；丙，点云从左上到右下且很稀疏。判断 r 大致为正值、负值还是接近 0。
- 甲研究 $r = 0.9$ ，乙研究 $r = -0.95$ 。谁的线性相关更强？为什么？
- 求数据 $(1, 1), (2, 3), (3, 4), (4, 6)$ 的回归方程，并预测 $x = 5$ 时的 $\hat{y}$ 。
- 已知回归直线过样本中心 $(4, 10)$ ，斜率 $b = 3$ 。求 a 。
- 完善列联表并校验合计：

	男	女	合计
会打篮球	20		35
不会			
合计	40	30	70

- 某调查 $\chi^2 = 4.9$ ，临界值 3.841。是否拒绝“两变量无关”？结论是什么？
- 某调查 $\chi^2 = 2.8 < 3.841$ 。用规范语言下结论。

B 应用层

- 某市统计了日最高气温 x ($^{\circ}\text{C}$) 与当日全市用电量 y (万度)：

x	10	15	20	25	30
y	4	6	9	11	15

求回归方程，并预测气温 28°C 时的用电量。

- 用第 9 题的回归方程计算各点残差，指出拟合最差的点。
- 某疫苗效果调查得到列联表（接种/未接种 $\times$ 感染/未感染，共 300 人）：接种者 200 人中感染 20 人；未接种者 100 人中感染 40 人。计算 χ^2 ，并在 $\alpha = 0.01$ （临界值 6.635）下下结论。
- 某项研究 $\chi^2 = 5.2$ 。在 $\alpha = 0.05$ 与 $\alpha = 0.01$ 下结论分别是什么？这说明了什么？

C 拓展层

- 有两组数据，用最小二乘法算出的回归方程完全相同，但第一组散点图是明显的直线带，第二组散点图是 U 形抛物线。有人认为“既然回归方程一样，就放心用第一组的结论预测第二组”。请说明这个想法错在哪里。
- 一项观察研究发现“每天喝咖啡的人心脏病发病率更低”，于是有人宣称“喝咖啡能预防心脏病”。请分析：这个结论能由观察数据直接得出吗？若要确认因果，应当怎么做？

参考答案（章末集中）

Try It 即时练习答案

- 4.1: 圆的半径与面积——函数关系 ($S = \pi r^2$)；学习时长与数学成绩——相关关系；心情与同桌鞋码——无关系。
- 4.2: 负相关（睡得多、犯困少）。但“睡得多所以犯困少”这个因果说法不能由相关直接断言——可能还有“作业量”“作息规律”等共同原因；要证明因果需控制变量做对照实验。
- 4.3: $\bar{x} = 2.5$, $\bar{y} = 6$; 分子 $= (-1.5)(-3) + (-0.5)(-1) + 0.5 \times 1 + 1.5 \times 3 = 4.5 + 0.5 + 0.5 + 4.5 = 10$; $\sum(x - \bar{x})^2 = 5$, $\sum(y - \bar{y})^2 = 9 + 1 + 1 + 9 = 20$; $r = \frac{10}{\sqrt{5 \times 20}} = \frac{10}{10} = 1$ 。数据恰好是一条直线上，完全线性相关。
- 4.4: $\bar{x} = 2.5$, $\bar{y} = 3.5$; $S_{xy} = (-1.5)(-2.5) + (-0.5)(-0.5) + 0.5 \times 0.5 + 1.5 \times 2.5 = 3.75 + 0.25 + 0.25 + 3.75 = 8$; $S_{xx} = 5$; $b = \frac{8}{5} = 1.6$, $a = 3.5 - 1.6 \times 2.5 = -0.5$ 。回归方程 $\hat{y} = 1.6x - 0.5$; $x = 5$ 时 $\hat{y} = 1.6 \times 5 - 0.5 = 7.5$ 。
- 4.5: 预测值依次为 \$1.1, 2.7, 4.3, 5.9\$，残差为 $-0.1, 0.3, -0.3, 0.1$ ； $|e|$ 最大为 0.3 ，拟合最差的是 $x = 2$ 与 $x = 3$ 两个点。
- 4.6: 补全后：会打篮球（男 20、女 15），不会（男 20、女 15），合计：男 40、女 30、总 70。期望频数 $E(\text{男生且会}) = \frac{40 \times 35}{70} = 20$ 。
- 4.7: $\chi^2 = 5.02 > 3.841 \rightarrow$ **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。

各节练习答案

- 4.1 节: 1. 圆的半径与周长——函数关系 ($C = 2\pi r$)；学习时长与数学成绩——相关关系；气温与冷饮销量——相关关系（背后可能有“季节”这一共同原因）。2. 不能。“相关”只说明两者同步变化，不能说明谁导致谁——阅读量高和成绩好可能同受“学习习惯好、家长重视”等共同原因影响；要确认因果需设计对照实验。3. 甲：正相关、较强 (r 接近 1)；乙：无明显相关 ($r \approx 0$)；丙：负相关、较弱 (r 在 -0.3 附近)。4. 点几乎贴成直线 $\rightarrow |r|$ 接近 1，方向向右上 $\rightarrow r \approx 1$ 。

- **4.2 节:** 1. $a = 7 - 2 \times 3 = 1$, 回归方程 $\hat{y} = 2x + 1$ 。2. 不合理—— $x = 100$ 超出样本范围 $[40, 60]$, 属外推, 无数据支撑。3. 预测值 3, 5, 7, 9, 残差 $-1, 0, 0, -1$; $|e|$ 最大为 1, 拟合最差的点是 $x = 1$ 与 $x = 4$ 。4. 下降—— $r < 0$, b 与 r 同号, 斜率 $b < 0$, 直线从左到右下降。

- **4.3 节:** 1. 期望: $E(\text{甲, 满意}) = \frac{60 \times 70}{100} = 42$, $E(\text{甲, 不满}) = \frac{60 \times 30}{100} = 18$, $E(\text{乙, 满意}) = \frac{40 \times 70}{100} = 28$, $E(\text{乙, 不满}) = \frac{40 \times 30}{100} = 12$ 。 $\chi^2 = \frac{(45 - 42)^2}{42} + \frac{(15 - 18)^2}{18} + \frac{(25 - 28)^2}{28} + \frac{(15 - 12)^2}{12} = \frac{9}{42} + \frac{9}{18} + \frac{9}{28} + \frac{9}{12} \approx 0.21 + 0.5 + 0.32 + 0.75 = 1.79$ 。 $1.79 < 3.841 \rightarrow$ 没有充分证据认为两店满意度有关联。 2. 拒绝—— $\chi^2 = 5.02 > 3.841$, 在犯错误概率不超过 0.05 的前提下认为有关联。 3. $\chi^2 = 2.1 < 3.841 \rightarrow$ 没有充分证据认为两变量有关联 (不能说“无关”)。

A1: 年龄与身高——相关关系 (发育期有关联但不确定); 电压与电流——函数关系 ($U = IR$); 天气与彩票号码——无关系。

A2: 甲 $r > 0$ 且 $|r|$ 较大 (强正相关); 乙 $r \approx 0$; 丙 $r < 0$ 且 $|r|$ 较小 (弱负相关)。

A3: 乙更强—— $|-0.95| = 0.95 > |0.9| = 0.9$, 强度比绝对值, 不看正负。

A4: $\hat{y} = 1.6x - 0.5$ (计算同 Try It 4.4); $x = 5$ 时 $\hat{y} = 7.5$ 。

A5: $a = 10 - 3 \times 4 = -2$ 。

A6: 会打篮球 (男 20、女 15); 不会 (男 20、女 15); 合计男 40、女 30、总 70 (三线校验通过)。

A7: $\chi^2 = 4.9 > 3.841 \rightarrow$ 拒绝“无关”, 在犯错误概率不超过 0.05 的前提下认为有关联。

A8: 没有充分证据认为两变量有关联。

B9: $\bar{x} = 20$, $\bar{y} = 9$; $S_{xy} = (-10)(-5) + (-5)(-3) + 0 + 5 \times 2 + 10 \times 6 = 50 + 15 + 0 + 10 + 60 = 135$; $S_{xx} = 250$; $b = \frac{135}{250} = 0.54$, $a = 9 - 0.54 \times 20 = -1.8$ 。 $\hat{y} = 0.54x - 1.8$; $x = 28$ 时 $\hat{y} = 0.54 \times 28 - 1.8 = 13.32 \approx 13.3$ (万度)。

B10: 预测值 \$3.6, 6.3, 9, 11.7, 14.4, 残差 \$0.4, -0.3, 0, -0.7, 0.6; $|e|$ 最大为 0.7, 拟合最差的是 $x = 25$ 的点。

B11: 列联表: 接种 200 人 (感染 20、未感染 180), 未接种 100 人 (感染 40、未感染 60), 合计感染 60、未感染 240。期望: $E(\text{接种, 感染}) = \frac{200 \times 60}{300} = 40$, $E(\text{接种, 未感染}) = \frac{200 \times 240}{300} = 160$, $E(\text{未接种, 感染}) = \frac{100 \times 60}{300} = 20$, $E(\text{未接种, 未感染}) = \frac{100 \times 240}{300} = 80$ 。 $\chi^2 = \frac{400}{40} + \frac{400}{160} + \frac{400}{20} + \frac{400}{80} = 10 + 2.5 + 20 + 5 = 37.5$ 。 $37.5 > 6.635 \rightarrow$ 在犯错误概率不超过 0.01 的前提下, 认为接种与否和是否感染有关联 (接种者感染率显著更低)。

B12: $\alpha = 0.05$: $5.2 > 3.841 \rightarrow$ 拒绝“无关”, 认为有关联; $\alpha = 0.01$: $5.2 < 6.635 \rightarrow$ 没有充分证据。同一个 χ^2 , 显著性水平更严时结论可能翻转——结论的把握度要与证据强度匹配。

C13: 回归方程相同只说明“在现有数据上拟合出的直线一样”, 但第二组是 U 形, 用直线预测必然系统性偏差 (中间高估、两端低估)。散点图决定“该不该用直线模型”——安斯库姆

四重奏告诉我们：不同形状的数据可以算出几乎相同的回归线，**必须看图选模型**。

C14：不能。观察研究只能说明相关，不能排除共同原因（如“爱喝咖啡的人收入高、体检意识强”）。要确认因果，应设计**随机对照实验**：随机分组、控制咖啡摄入、排除混杂因素，再比较发病率。

第五章 把现实装进模型：数学建模与拓展

章首情境

5.1 建模四步：从现实到数学再回现实

5.1.1 审题与建模：找变量、写定义域

学习目标

- 掌握建模流程四步法：审题 → 建模 → 求解 → 检验；
- 会从情境中找出自变量、因变量并写出**现实定义域**；
- 能根据数据规律初步判断模型类型。

前置复习：一个数每次乘 $\frac{3}{4}$ ，第 0 次是 64，第 1 次、第 2 次分别是多少？

新知：建模四步法

- **第一步·审题：**明确问题要什么 → 找自变量 x 、因变量 y → 列出已知数据 → **写清变量的现实定义域**（人数必为整数、时间非负、质量有上限）。
- **第二步·建模：**根据数据变化规律选函数模型（线性、指数、对数、分段……）。
- **第三步·求解：**用解方程、不等式、函数性质等算出数学结果。
- **第四步·检验与回写：**结果不符合现实？不符就回头改模型；最后把数学答案**翻译回现实语言**。

例题 5.1

（完整四步：咖啡降温，对抗 M1、M2）

一杯 89°C 的咖啡放在 25°C 的房间里，每 2 分钟测一次温度，得： $t = 0, 2, 4, 6$ (min) 时 $T = 89, 73, 61, 52$ ($^{\circ}\text{C}$)。

(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 $t = 8$ 、 $t = 10$ 时的温度；(4) 检验模型是否合理。

解：(1) 先看 T 的差：-16, -12, -9——**差在递减**，不是线性。把“高出室温的部分”列出来： $T - 25 = 64, 48, 36, 27$ ，相邻两项的比： $\frac{48}{64} = \frac{36}{48} = \frac{27}{36} = 0.75$ ——**比恒定**，选指数模型。(2) 设 $T - 25 = c \cdot a^t$ 。 $t = 0$ 时 $T - 25 = 64$ ，得 $c = 64$ ；每 2 分钟乘一次 0.75，即 t 增加 2 时乘 0.75，所以

$$T - 25 = 64 \times (0.75)^{t/2}, \quad \text{即} \quad T = 25 + 64 \times \left(\frac{3}{4}\right)^{t/2}$$

(3) $t = 8$: $(0.75)^4 = \frac{81}{256} = 0.316$, $64 \times 0.316 = 20.25$, $T \approx 45.3^\circ\text{C}$; $t = 10$: $(0.75)^5 = \frac{243}{1024} = 0.237$, $64 \times 0.237 = 15.19$, $T \approx 40.2^\circ\text{C}$ 。(4) 检验：模型给的温度单调下降、且永远不低于室温 25°C ($(\frac{3}{4})^{t/2} \rightarrow 0$)，符合物理直觉； $t = 8, 10$ 都在实测时间附近，内插可靠。

Try It 5.1：某手机充满电后待机耗电，电量 (%) 每小时记录：100, 88, 77, 67, 58。判断应选用什么模型 (线性/指数/二次)，并预测第 5 小时的剩余电量。(提示：先看差，再看“差的差”)

例题 5.2

(定义域的现实约束，对抗 M2)

某商店每月销售 x 台打印机 (x 为自然数)，月利润 $y = -x^2 + 13x - 30$ (万元)。问：月利润最大时卖出多少台？最大利润是多少？

解：先定定义域：利润非负要求 $-x^2 + 13x - 30 \geq 0$ ，即 $(x - 3)(x - 10) \leq 0$ ，得 $3 \leq x \leq 10$ ；又 x 为自然数，故 $x \in \{3, 4, \dots, 10\}$ 。 $y = -x^2 + 13x - 30$ 的对称轴为 $x = \frac{13}{2} = 6.5$ ，顶点对应利润 12.25 万元——但 $x = 6.5$ 不是自然数，卖不了“半台”打印机！所以看 $x = 6$ 与 $x = 7$ ：

$$y(6) = -36 + 78 - 30 = 12, \quad y(7) = -49 + 91 - 30 = 12$$

答：卖出 **6 台或 7 台** 时利润最大，最大利润 **12 万元**。

Try It 5.2：某旅行社组团旅游， x 人 (x 为正整数) 时人均费用为 $800 - 10x$ 元，总收入 $f(x) = x(800 - 10x)$ 。写出定义域并求最大收入。

5.1.2 求解与检验：算完要回头看

学习目标

- 会用“代入已知点看残差”做数值检验；
- 会用“定义域、单调性、量级常识”做现实检验；
- 理解“检验不过就回头改模型”的迭代思想。

前置复习： 100×4^2 与 100×4^5 分别等于多少？

新知：检验的两道关

- **数值检验**：把已知数据点代入模型，比较预测值与观测值，残差 $e_i = y_i - \hat{y}_i$ 过大说明模型不合适；

- **现实检验**：定义域/值域是否合理（人数不能为负、温度不能无限涨）、趋势是否符合常识（人口不会永远指数增长，资源有限时增长必放缓）。

How To：检验三步

- **回代**：把原始数据代回模型，看残差；
- **审边界**：检查极端输入下模型的输出是否合理；
- **判趋势**：长期趋势是否违背常识；不合理 → **回到建模步换模型或加约束**，迭代。

例题 5.3

（指数模型的检验，对抗 M4、M6）

细菌培养实验，每 30 分钟计数一次：

t (h)	0	0.5	1	1.5	2
N (个)	100	200	400	800	1600

（1）建立 N 关于 t 的模型；（2）用模型预测 $t = 5$ 时的数量；（3）检验这个预测是否现实。

解：（1）相邻两次的比： $\frac{200}{100} = \frac{400}{200} = \dots = 2$ ，每 30 分钟翻倍，故每小时乘 4， $N = 100 \times 4^t$ 。（2） $t = 5$ ： $N = 100 \times 4^5 = 100 \times 1024 = 102400$ （个）。（3）检验：把 $t = 5$ 回代，预测值 102400 个——**数量级上“离谱”吗？**培养皿里的营养物是有限的，细菌数量不可能永远翻倍增长。按这个模型，再往后 $t = 10$ 就有 1.05×10^8 个，早已溢出培养皿。**现实检验不过关** → 应修正模型：当数量接近环境容量时增长率放缓，改用 S 型（逻辑斯蒂）模型。

Try It 5.3：某学生用线性模型预测某地 5 年后的常住人口，算得 $P(5) = -3$ （万人）。这个结果合理吗？该怎样检验和修正？

本节练习（答案见章末）

- 数据 $x = 0, 1, 2, 3$, $y = 5, 8, 11, 14$ 。判断模型类型并写出 y 关于 x 的解析式。
- 某旅行社： $f(x) = x(800 - 10x)$ ， x 为正整数。写出定义域并求最大收入。
- 某模型预测某商品第 10 个月的销量为 -8 件。指出哪里不合理，并给出合理的处理方式。

5.2 模型选择：数据形状决定模型家族

5.2.1 差稳定还是比稳定

学习目标

- 掌握“差稳定用线性、比稳定用指数”的判据；
- 了解二阶差恒定对应二次模型；
- 能对给定数据独立判断并说明依据。

前置复习：数列 3, 6, 12, 24 的相邻差与相邻比各是多少？

新知：增长模式判据

观察方式	规律	对应模型
后项 - 前项 $\approx$ 常数	每步“加同样多”	线性 $y = kx + b$
后项 $\div$ 前项 $\approx$ 常数	每步“乘同样多”	指数 $y = c \cdot a^x$
差再差（二阶差） $\approx$ 常数	增量本身在“匀速增加”	二次 $y = ax^2 + bx + c$
不同区间规律不同	边界处换规则	分段模型

How To：选模型三步

- **描点看图**（第 4 章技能）：先看散点趋势；
- **算差与比**：实际列两行，分别看“差”和“比”哪个稳定；
- **选模型并验证**：定了类型再定参数，最后回代检验。

例题 5.4

（三组数据选模型，对抗 M3、M11）

判断下列三组数据各适合什么模型，并说明依据：

A: $x = 1, 2, 3, 4, 5, y = 3, 5, 7, 9, 11$; B: $x = 1, 2, 3, 4, 5, y = 2, 4, 8, 16, 32$; C: 某通话套餐，前 2 小时内每分钟 0.1 元，超过 2 小时后每分钟 0.08 元。

解：A: 差为 2, 2, 2, 2 恒定 $\rightarrow$ **线性**, $y = 2x + 1$ 。B: 比为 2, 2, 2, 2 恒定 $\rightarrow$ **指数**, $y = 2^x$ 。C: 前 2 小时每分钟 0.1 元（每小时 6 元），之后每分钟 0.08 元（每小时 4.8 元），规则在 2 小时处改变 $\rightarrow$ **分段模型**: $y = 6x$ ($0 \leq x \leq 2$), $y = 12 + 4.8(x - 2)$ ($x > 2$)。检验 C 的衔接: $x = 2$ 时，第一段 $y = 12$ ，第二段 $y = 12 + 0 = 12$ ——**断点处两段取值相等，衔接平滑** ✓。

Try It 5.4: 数据 $x = 1, 2, 3, 4, y = 1, 4, 9, 16$ 。算差与二阶差，判断模型类型并写出解析式。

5.2.2 简单优于复杂：别让模型“过拟合”

学习目标

- 理解过拟合：拟合已知点很准、预测新点很糟；
- 理解“奥卡姆剃刀”：能用简单模型不用复杂模型；
- 会用“外推稳定性”比较模型优劣。

前置复习：过 3 个不在同一直线上的点，可以画一条直线吗？可以画一条抛物线吗？

新知：过拟合与模型选择原则

- **拟合不等于预测：**一个模型把已知点全都精确穿过（残差为 0），只能说明它“记住了”已有数据；对没见过数据预测得准不准，才是真本事。
- **过拟合：**模型过于复杂，把噪声也“拟合”进去了，导致对新数据的预测反而更差。
- **奥卡姆剃刀：**在解释力相当的情况下，**选择更简单的模型。**

例题 5.5

（过拟合对比，对抗 M5）

三个月的销量数据：(0, 1), (1, 2), (2, 1.5) (x 为月， y 为万件)。两个候选模型：

模型甲（线性）： $\hat{y}_{\text{甲}} = 0.25x + 1.25$ ；模型乙（二次，精确穿过三点）： $y_{\text{乙}} = -0.75x^2 + 1.75x + 1$ 。

（1）分别求两模型对三个已知点的残差平方和；（2）分别预测 $x = 3$ 的销量；（3）哪个模型更值得用？

解：（1）甲在三点处的预测为 1.25, 1.5, 1.75，残差为 -0.25, 0.5, -0.25，平方和 = $0.0625 + 0.25 + 0.0625 = 0.375$ 。乙精确穿过三点，残差全为 0，平方和为 0——拟合“更好”。（2）甲预测： $x = 3$ 时 $\hat{y}_{\text{甲}} = 0.25 \times 3 + 1.25 = 2$ （万件）；乙预测： $y_{\text{乙}} = -0.75 \times 9 + 1.75 \times 3 + 1 = -0.5$ （万件）——**销量是负数，明显荒谬**（数据一直为正且在增长）。（3）甲更值得用：虽然它对已知点残差稍大，但对新点的预测合理稳定；乙“记住了”三个点，却在 $x = 3$ 处剧烈外摆，是典型的**过拟合**。

Try It 5.5：一个模型对已知 10 个点残差全为 0，但对第 11 个点预测出 1000 倍于正常量级的值。这个模型哪里出了问题？应当怎么办？

本节练习（答案见章末）

- 数据 $x = 1, 2, 3, 4$, $y = 1, 4, 9, 16$ 。判断模型类型并写出解析式。
- 两组递增数据：甲组 3, 6, 9, 12，乙组 3, 6, 12, 24。分别属于什么模型？依据是什么？
- 某同学建了个高次多项式模型，对已知点拟合完美，但外推预测时结果忽大忽小。这说明了什么？应当怎么办？

5.3 拓展 · 矩阵初步：多元关系的一张表

5.3.1 矩阵与基本运算

学习目标

- 认识矩阵及其记号（行、列、 $m \times n$ 型）；
- 掌握矩阵加法、数乘、乘法的运算法则；
- 知道矩阵乘法不满足交换律（ $AB \neq BA$ ）。

前置复习： $1 \times 5 + 2 \times 0$ 、 $3 \times 6 + 4 \times (-1)$ 分别等于多少？

新知：矩阵与运算

把数按行、列排成的矩形阵列叫**矩阵**，如 $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ 是 2×2 矩阵。

- **加法：** 同型矩阵对应元素相加；
- **数乘：** 每个元素都乘这个数；
- **乘法：** A ($m \times n$ 型) 与 B ($n \times p$ 型) 可以相乘得 C ($m \times p$ 型)， C 的第 i 行第 j 列元素 = A 第 i 行元素与 B 第 j 列元素**对应相乘后求和**（“行乘列求和”）。

例题 5.6

（矩阵运算，对抗 M7、M8）

$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$, $B = \begin{pmatrix} 5 & 6 \\ 0 & -1 \end{pmatrix}$ 。求 $A + B$ 、 $2A$ 、 AB 与 BA ，并比较 AB 与 BA 。

解： $A + B = \begin{pmatrix} 1+5 & 2+6 \\ 3+0 & 4+(-1) \end{pmatrix} = \begin{pmatrix} 6 & 8 \\ 3 & 3 \end{pmatrix}$ ； $2A = \begin{pmatrix} 2 & 4 \\ 6 & 8 \end{pmatrix}$ 。

AB ：第 1 行第 1 列： $1 \times 5 + 2 \times 0 = 5$ ；第 1 行第 2 列： $1 \times 6 + 2 \times (-1) = 4$ ；第 2 行第 1 列： $3 \times 5 + 4 \times 0 = 15$ ；第 2 行第 2 列： $3 \times 6 + 4 \times (-1) = 14$ 。

$$AB = \begin{pmatrix} 5 & 4 \\ 15 & 14 \end{pmatrix}$$

BA ：第 1 行第 1 列： $5 \times 1 + 6 \times 3 = 23$ ；第 1 行第 2 列： $5 \times 2 + 6 \times 4 = 34$ ；第 2 行第 1 列： $0 \times 1 + (-1) \times 3 = -3$ ；第 2 行第 2 列： $0 \times 2 + (-1) \times 4 = -4$ 。

$$BA = \begin{pmatrix} 23 & 34 \\ -3 & -4 \end{pmatrix}$$

$AB \neq BA$ ——矩阵乘法**不满足交换律**！这与数的乘法完全不同。另外注意乘法对“尺寸”有要求： A ($m \times n$) 能乘 B ($n \times p$) 的前提是 **A 的列数 = B 的行数**。

Try It 5.6： $A = \begin{pmatrix} 2 & 0 \\ 1 & 3 \end{pmatrix}$, $B = \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix}$ 。求 $A + B$ 与 AB 。

5.3.2 线性方程组的矩阵形式

学习目标

- 会把二元线性方程组写成矩阵形式 $A\mathbf{x} = \mathbf{b}$;
- 理解“矩阵 = 一批线性关系的打包”。

前置复习：解方程组 $\begin{cases} 2x + 3y = 8 \\ x + 4y = 9 \end{cases}$ 。

新知：方程组的矩阵形式

线性方程组

$$\begin{cases} a_{11}x + a_{12}y = b_1 \\ a_{21}x + a_{22}y = b_2 \end{cases}$$

可以写成矩阵形式

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$$

例题 5.7

(方程组矩阵形式)

把方程组 $\begin{cases} 2x + 3y = 8 \\ x + 4y = 9 \end{cases}$ 写成矩阵形式，并验证解。

解：矩阵形式为

$$\begin{pmatrix} 2 & 3 \\ 1 & 4 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 8 \\ 9 \end{pmatrix}$$

解方程组：由 $x = 9 - 4y$ 代入第一式， $2(9 - 4y) + 3y = 8$ ，得 $-5y = -10$ ， $y = 2$ ， $x = 1$ 。验证（用矩阵乘法“还原”）： $\begin{pmatrix} 2 & 3 \\ 1 & 4 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 2 \times 1 + 3 \times 2 \\ 1 \times 1 + 4 \times 2 \end{pmatrix} = \begin{pmatrix} 8 \\ 9 \end{pmatrix}$ ✓。

Try It 5.7：把方程组 $\begin{cases} x + 2y = 5 \\ 3x + y = 10 \end{cases}$ 写成矩阵形式并求其解。

本节练习（答案见章末）

• $A = \begin{pmatrix} 2 & 0 \\ 1 & 3 \end{pmatrix}$ ， $B = \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix}$ 。求 $A + B$ 与 AB 。

• 把方程组 $\begin{cases} x + 2y = 5 \\ 3x + y = 10 \end{cases}$ 写成矩阵形式并求其解。

• 判断： 2×3 矩阵与 3×4 矩阵能否相乘？乘积是几行几列？ 2×3 与 2×3 呢？

5.4 拓展 · 逻辑初步与数学史

5.4.1 命题、逆否命题与反证法

学习目标

- 理解命题、蕴含 $p \Rightarrow q$ 与逆否命题；
- 知道原命题与逆否命题同真同假；
- 会用反证法证明简单命题。

前置复习：命题“若 $x > 1$ ，则 $x^2 > 1$ ”是真的吗？它的逆命题是什么？

新知：命题与反证法

- **命题：**可以判断真假的陈述句。**蕴含** $p \Rightarrow q$ 表示“若 p 则 q ”；
- **逆否命题：**把 $p \Rightarrow q$ 的条件结论互换并都否定，得 $\neg q \Rightarrow \neg p$ 。**原命题与逆否命题同真同假**（互为等价）；
- **反证法：**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。

例题 5.8

（反证法，对抗 M9）

证明：若 n^2 是偶数，则 n 是偶数（ n 为整数）。

证明：反证。假设 n 不是偶数，即 n 是奇数，则 $n = 2k + 1$ （ k 为整数）。

$$n^2 = (2k + 1)^2 = 4k^2 + 4k + 1 = 2(2k^2 + 2k) + 1$$

这是“偶数 + 1”，仍是奇数，与已知“ n^2 是偶数”矛盾。故假设不成立， n 必为偶数。■

Try It 5.8：用反证法证明：若 $a + b$ 是奇数（ a, b 为整数），则 a, b 一奇一偶。（提示：假设 a, b 同奇或同偶会怎样？）

5.4.2 数学史与新技术：概念为什么这样长

学习目标

- 了解函数、统计相关等概念的历史由来；
- 理解“最小二乘”“相关”等名词的含义来源；
- 会解读软件/计算器的输出，并坚持“人判断原理”。

前置复习：第 4 章的残差平方和是什么？为什么用“平方”而不用“直接相减”？

新知：数学史是理解资源，不是背诵考点

- **函数概念**：从“变量说”（一个量随另一个量变）演进到“对应说”（集合间的一种对应），“对应说”更精确，避免了“什么是变量”的含糊；
- **回归与相关**：统计学家高尔顿研究父子身高时，发现极端身高会“向平均值回归”——“回归”一词由此而来；“相关”（相关系数）由皮尔逊给出，衡量“同步变化”的强度；
- **最小二乘**：为什么不直接用“残差之和”而用“残差平方和”？因为残差有正有负、会互相抵消，平方让正负都变成正、且放大离得远的点——这也是第 4 章方差思想的来源。

例题 5.9

（新技术输出解读，对抗 M10）

某同学用软件对一组“投入—产出”数据做回归，软件输出：线性回归 $\hat{y} = 2.1x + 3.5$, $R^2 = 0.96$ （拟合度很高）。但他在生成散点图时发现数据呈明显的 U 形弯曲。

（1）这个 $R^2 = 0.96$ 能说明线性模型合适吗？（2）应该怎么办？

解：（1）不能。 R^2 高只说明“在直线假设下拟合得好”，而散点图已经明明白白告诉你是 U 形——用直线去拟合弯的数据， R^2 再高也是自欺欺人。（2）回到原理判断：放弃线性模型，改选二次或分段模型，再重新拟合验证。工具负责“算”，“选什么模型、结论怎么读”始终是人做的判断。

Try It 5.9：软件对某组数据输出指数拟合 $N = 500 \times 1.5^t$, $R^2 = 0.99$ 。可你知道该场景是“培养皿细菌”且营养早已耗尽。这个输出能直接用吗？为什么？

本节练习（答案见章末）

- 判断命题“若 $x > 1$ ，则 $x^2 > 1$ ”的真假，并写出它的逆否命题。
- 用反证法证明：若 $a + b$ 是奇数（ a, b 为整数），则 a, b 一奇一偶。
- 为什么“相关关系”不能改叫“因果关系”？（从数学史或统计含义角度说明）

本章回顾

概念地图

· 建模四步：审题 → 建模 → 求解 → 检验（不过关回头改，闭环）· 审题：找变量 + 写清**现实定义域**（整数/非负/有限制）· 建模：选模型（差稳定 → 线性；比稳定 → 指数；二阶差恒定 → 二次；分段看断点）· 求解：方程、不等式、函数性质、统计工具· 检验：数值（残差 $e_i = y_i - \hat{y}_i$ ）+ 现实（常识、量级、趋势）· 原则：奥卡姆剃刀——简单优于复杂，防过拟合· 矩阵：数表打包；加法对应元素加、数乘逐个乘、乘法行乘列求和； $AB \neq BA$ · 方程组矩阵形式： $Ax = b$ （一堆线性关系的打包）· 逻辑：原命题 $\Leftrightarrow$ 逆否命题；反证法 = 假设反面推出矛盾· 数学史与新技术：名字背后是故事；工具算、人判断

关键公式速查

内容	要点
差判据	后项 - 前项 $\approx$ 常数 $\rightarrow$ 线性
比判据	后项 $\div$ 前项 $\approx$ 常数 $\rightarrow$ 指数
二阶差	差再差 $\approx$ 常数 $\rightarrow$ 二次
指数冷却/增长	$y = c \cdot a^x$, $0 < a < 1$ 衰减、 $a > 1$ 增长
检验	残差小 + 现实合理 + 内插不外推
矩阵乘法	行乘列求和; $(m \times n)(n \times p) \rightarrow m \times p$; $AB \neq BA$ 一般成立
逻辑	原命题 $\Leftrightarrow$ 逆否命题
反证法	假设反面 $\rightarrow$ 推出矛盾 $\rightarrow$ 原命题成立

分层练习

A 基础层

- 数据 3, 6, 12, 24, 48。判断适合什么模型, 说明依据。
- 用 24 m 长的栅栏围一块矩形菜地, 设长为 x m。写出面积关于 x 的函数及定义域, 并求最大面积。
- 某模型预测某商品第 12 个月的销量为 -8 件。指出不合理之处。
- 计算: $\begin{pmatrix} 1 & 3 \\ 2 & 0 \end{pmatrix} + \begin{pmatrix} 4 & 1 \\ 1 & 5 \end{pmatrix}$ 。
- 计算: $\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 1 & 1 \end{pmatrix}$ 。
- 判断命题“若两个角是对顶角, 则它们相等”的真假, 并判断其逆命题“若两角相等, 则它们是对顶角”的真假。
- 把方程组 $\begin{cases} 2x + y = 7 \\ x + 3y = 11 \end{cases}$ 写成矩阵形式并求其解。
- 两组递增数据: 甲 5, 8, 11, 14, 乙 5, 10, 20, 40。分别判断模型类型。

B 应用层

- 某产品年销量(万台)随年份数据: 第 1~5 年依次为 5, 8, 13, 20, 29。判断模型类型, 建立模型并预测第 6 年销量。
- 某市居民用电阶梯电价: 月用电 200 度及以下每度 0.5 元, 超过 200 度的部分每度 0.8 元。写出月电费 y 与用电量 x 的分段函数, 并计算小明家月用电 180 度、350 度各付多少。检验分段函数在 $x = 200$ 处是否衔接。
- 甲、乙两家餐厅的菜品单价(元)分别为: 甲 $\begin{pmatrix} 12 & 15 \end{pmatrix}$, 乙 $\begin{pmatrix} 10 & 18 \end{pmatrix}$; 小明点菜量为 $\begin{pmatrix} 3 \\ 4 \end{pmatrix}$ (第一样菜 3 份、第二样菜 4 份)。分别求两家餐厅的花费。

- 某软件对一组数据输出线性回归， $R^2 = 0.99$ ，但散点图呈明显的弯曲形状。这个结果能否直接用？为什么？应当怎么做？

C 拓展层

- 某地区人口（万人）数据：第 0~6 年依次为 100, 120, 144, 170, 190, 205, 215。前三年“比”约为 1.2（指数增长），后三年“比”明显下降。这说明指数模型在该数据上出了什么问题？应当如何修正？
- 用反证法证明： $\sqrt{2}$ 不是有理数。（提示：可先用 5.4.1 例题 5.8 的结论：“若 n^2 是偶数，则 n 是偶数”）

参考答案（章末集中）

Try It 即时练习答案

- 5.1:** 差为 $-12, -11, -10, -9$ 不恒定，但“差的差”为 $1, 1, 1$ 恒定 $\rightarrow$ **二次模型**。设 $y = at^2 + bt + c$ ：由 $y(0) = 100$ 得 $c = 100$ ； $y(1) = a + b + 100 = 88$ ， $y(2) = 4a + 2b + 100 = 77$ ，解得 $a = 0.5$ ， $b = -12.5$ ，模型为 $y = 0.5t^2 - 12.5t + 100$ 。 $t = 5$ 时 $y = 12.5 - 62.5 + 100 = 50$ （%）。预测第 5 小时剩余电量约为 **50%**。
- 5.2:** 定义域： x 为正整数且 $800 - 10x \geq 0$ ，故 $1 \leq x \leq 80$ 。 $f(x) = -10x^2 + 800x$ ，对称轴 $x = 40$ （为整数，可取）， $f_{\max} = f(40) = 40 \times 400 = 16000$ （元）。
- 5.3:** 不合理——人口不可能为负。检验：模型斜率应为正、值域应为正；应检查数据是否线性（若数据增长放缓，应改指数或 S 型模型，并重新定参数）。
- 5.4:** 差为 $3, 5, 7$ 递增，二阶差为 $2, 2$ 恒定 $\rightarrow$ **二次模型**， $y = x^2$ 。
- 5.5:** 过拟合——模型把噪声也拟合进去了，对新数据剧烈失真。应改用更简单的模型（如线性），用残差、外推稳定性重新评估。
- 5.6:** $A + B = \begin{pmatrix} 3 & 1 \\ 1 & 5 \end{pmatrix}$ ； $AB = \begin{pmatrix} 2 & 2 \\ 1 & 7 \end{pmatrix}$ （第 1 行第 1 列 $2 \times 1 + 0 \times 0 = 2$ ；第 2 行第 2 列 $1 \times 1 + 3 \times 2 = 7$ ）。
- 5.7:** 矩阵形式 $\begin{pmatrix} 1 & 2 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 5 \\ 10 \end{pmatrix}$ 。由 $x = 5 - 2y$ 代入 $3(5 - 2y) + y = 10$ ， $-5y = -5$ ， $y = 1$ ， $x = 3$ 。
- 5.8:** 反证：假设 a, b 同奇或同偶。若同偶， $a + b$ 为偶数；若同奇， $a + b = 2k + 1 + 2m + 1 = 2(k + m + 1)$ 也为偶数——都与“ $a + b$ 为奇数”矛盾。故 a, b 一奇一偶。■
- 5.9:** 不能直接用——营养早已耗尽，指数增长在现实上不可持续（检验环节不过关）。应结合背景改用 S 型模型；工具输出再漂亮，也要先过“现实检验”这一关。

各节练习答案

- **5.1 节**: 1. 差为 3, 3, 3 恒定 $\rightarrow$ 线性, $y = 3x + 5$ 。2. 定义域: x 正整数且 $x \leq 80$; $f(x) = -10x^2 + 800x$, 顶点 $x = 40$ (整数), 最大收入 $f(40) = 40 \times 400 = 16000$ 元。3. 销量不可能为负——数值检验不过关。应检查模型与数据 (可能数据是递增、模型斜率应为正), 修正模型或扩大样本重新拟合。
- **5.2 节**: 1. 二阶差恒定 (3, 5, 7, 再差 2, 2) $\rightarrow$ 二次模型, $y = x^2$ 。2. 甲组差 3, 3, 3 $\rightarrow$ 线性; 乙组比 2, 2, 2 $\rightarrow$ 指数。依据: 差稳定 vs 比稳定。3. 过拟合——模型记住了噪声, 外推不可靠。应换简单模型 (如线性/二次) 重新拟合, 并用外推预测的稳定性判断。
- **5.3 节**: 1. $A + B = \begin{pmatrix} 3 & 1 \\ 1 & 5 \end{pmatrix}$; $AB = \begin{pmatrix} 2 & 2 \\ 1 & 7 \end{pmatrix}$ 。2. 矩阵形式 $\begin{pmatrix} 1 & 2 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 5 \\ 10 \end{pmatrix}$, 解 $x = 3, y = 1$ 。3. 2×3 与 3×4 可乘 ($3 = 3$), 乘积为 2×4 ; 2×3 与 2×3 不可乘 ($3 \neq 2$)。
- **5.4 节**: 1. 真命题; 逆否命题: “若 $x^2 \leq 1$, 则 $x \leq 1$ ” (也为真)。2. 反证: 假设 a, b 同奇或同偶, 则 $a + b$ 为偶数, 与 “奇数” 矛盾 $\rightarrow$ 一奇一偶。3. 相关只刻画 “同步变化的趋势”, 不说明谁导致谁; 因果需对照实验/控制变量等更强证据 (第 4 章 M1 反例: 冰淇淋—溺水同涨因夏天)。从数学史上讲, “相关” 由统计学家对 “同步偏离” 的度量而得名, 从未包含因果含义。

A1: 比为 2, 2, 2, 2 $\rightarrow$ 指数模型。

A2: 设长 x 、宽 $12 - x$, $S = x(12 - x) = -(x - 6)^2 + 36$, 定义域 $0 < x < 12$; $x = 6$ 时 $S_{\max} = 36$ (m^2), 即围成边长 6 m 的正方形。

A3: 销量为负不合理——数值/现实检验不过关, 应修正模型 (检查数据趋势、换模型重新拟合)。

A4: $\begin{pmatrix} 5 & 4 \\ 3 & 5 \end{pmatrix}$ 。

A5: $\begin{pmatrix} 1 \times 2 + 2 \times 1 & 1 \times 0 + 2 \times 1 \\ 3 \times 2 + 4 \times 1 & 3 \times 0 + 4 \times 1 \end{pmatrix} = \begin{pmatrix} 4 & 2 \\ 10 & 4 \end{pmatrix}$ 。

A6: 原命题真 (对顶角必相等); 逆命题假 (两个相等的角不一定是对顶角)。

A7: $\begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 7 \\ 11 \end{pmatrix}$ 。由 $y = 7 - 2x$ 代入 $x + 3(7 - 2x) = 11$, $-5x = -10$, $x = 2$, $y = 3$ 。

A8: 甲差 3, 3, 3 $\rightarrow$ 线性; 乙比 2, 2, 2 $\rightarrow$ 指数。

B9: 差 3, 5, 7, 9 递增, 二阶差 2, 2, 2 恒定 $\rightarrow$ 二次模型 $y = x^2 + 4$ ($x = 1$ 时 5, $x = 5$ 时 29 $\checkmark$)。第 6 年预测: $6^2 + 4 = 40$ (万台)。检验: 回代五年的残差全为 0, 二阶差恒定, 模型合理。

B10: $y = \begin{cases} 0.5x, & 0 \leq x \leq 200 \\ 100 + 0.8(x - 200), & x > 200 \end{cases}$ 。衔接: $x = 200$ 时两段均为 100 ✓。180 度: $0.5 \times 180 = 90$ 元; 350 度: $100 + 0.8 \times 150 = 220$ 元。

B11: 甲: $12 \times 3 + 15 \times 4 = 96$ (元); 乙: $10 \times 3 + 18 \times 4 = 102$ (元)。

B12: 不能直接用——散点图弯曲说明“直线假设”本身错了, R^2 再高也是拟合错了形状。应改选二次/分段模型重新拟合; 先看图再算数是铁律。

C13: 前三年“比”稳定为 1.2 可用指数; 后三年“比”降至 1.08 左右, 说明增长放缓——指数模型会无限增长, 违背现实(资源、环境容量有限)。应修正为 S 型(逻辑斯蒂)模型: 初期指数、接近容量时放缓。

C14: 假设 $\sqrt{2} = \frac{p}{q}$ (p, q 为互质的正整数)。两边平方得 $p^2 = 2q^2$, 故 p^2 是偶数。由例题 5.8 的结论, p 是偶数, 设 $p = 2k$, 则 $4k^2 = 2q^2$, $q^2 = 2k^2$, q^2 是偶数, q 也是偶数。于是 p, q 都是偶数, 与“ p, q 互质”矛盾。故假设不成立, $\sqrt{2}$ 不是有理数。■

第六章 卷末综合检测

一、选择题（本大题共 6 小题，每小题 5 分，共 30 分）

Z1

某校食堂午餐设三类窗口：盖饭窗口有 4 种，面条窗口有 3 种，轻食窗口有 2 种。小明只买一份午餐（只在一个窗口购买），则他不同的购买选择共有（ ）种。

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

答案

A

解析

这件事“买一份午餐”，选到任何一个窗口中的任意一种都立即完成，属于分类问题，各类互不重叠且覆盖全部，由分类加法计数原理得 $N = 4 + 3 + 2 = 9$ 。选 C 是误用乘法（把三类当成“先…再…”的步骤）；选 B 误算为 4×3 ；选 D 误算为 $4 + 3 + 2 + 5$ 之类。判据一句话：做完这一步“整件事完成了吗”——完成了就加，没完成才乘。

Z2

袋中有 3 个红球、2 个白球，除颜色外完全相同。从中不放回地任取 2 个球，则“恰好取到 1 个红球、1 个白球”的概率为（ ）。

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

答案

B

解析

古典概型。取 2 个球的所有等可能结果数为 $C_5^2 = 10$ ；“1 红 1 白”需分两步：先取红球 C_3^1 种、再取白球 C_2^1 种，由分步乘法计数原理共 $C_3^1 \times C_2^1 = 6$ 种。故 $P = \frac{6}{10} = \frac{3}{5}$ 。若只取红或只取白，都会漏算“红白搭配”的两步选择。

Z3

$(1 + 2x)^5$ 的展开式中 x^2 项的系数为 ()。

- A. 40
- B. 80
- C. 10
- D. 20

答案

A

解析

二项式定理。通项 $T_{r+1} = C_5^r \cdot 1^{5-r} \cdot (2x)^r = C_5^r \cdot 2^r \cdot x^r$ 。令 $r = 2$ ，得 x^2 项系数 $C_5^2 \times 2^2 = 10 \times 4 = 40$ 。选 B 是把 C_5^2 误作 $C_5^3 = 10$ 再乘 $2^3 = 8$ ；选 C 忘了乘 2^r ；选 D 只取了 C_5^2 没取 2^2 。注意：题目问的是**系数**（含数 2 的幂），不是只问组合数。

Z4

设随机变量 X 的分布列为 $P(X = 1) = 0.4$, $P(X = 2) = 0.3$, $P(X = 3) = 0.3$ ，则 $D(X)$ 的值为 ()。

- A. 0.61
- B. 0.69
- C. 0.81
- D. 0.49

答案

B

解析

先求期望： $E(X) = 1 \times 0.4 + 2 \times 0.3 + 3 \times 0.3 = 1.9$ 。再用公式 $D(X) = E(X^2) - [E(X)]^2$ ： $E(X^2) = 1^2 \times 0.4 + 2^2 \times 0.3 + 3^2 \times 0.3 = 0.4 + 1.2 + 2.7 = 4.3$ ， $D(X) = 4.3 - 1.9^2 = 4.3 - 3.61 = 0.69$ 。选 A 是误用 $D(X) = E(X^2) - E(X) = 4.3 - 1.9 = 2.4$ 一类的错误变形（忘记平方）；选 C、D 是计算错位。务必按“先期望、再平方期望、后作差”的顺序。

Z5

某校高二年级学生的数学期末成绩近似服从正态分布 $N(75, \sigma^2)$, 已知 $P(X \geq 85) = 0.15$, 则 $P(65 \leq X \leq 75)$ 等于 ()。

- A. 0.30
- B. 0.35
- C. 0.70
- D. 0.15

答案

B

解析

正态分布密度曲线关于均值 75 对称。85 与 65 关于 75 对称, 故 $P(X \leq 65) = P(X \geq 85) = 0.15$ 。于是 $P(65 \leq X \leq 85) = 1 - 0.15 - 0.15 = 0.70$, 再由对称性, $P(65 \leq X \leq 75)$ 是它的一半, 即 0.35。选 C 是忘了除以 2 (把整段 $[65, 85]$ 当成了答案); 选 A、D 是错误套用数据。做正态题先画对称轴, 把“已知一侧”通过对称翻译到另一侧。

Z6

已知变量 x 与 y 的一组样本数据经计算得回归直线方程为 $\hat{y} = 2.3x + 5$, 且 $\bar{x} = 10$, 则 $\bar{y}$ 的值为 ()。

- A. 23
- B. 28
- C. 25
- D. 30

答案

B

解析

回归直线必过样本中心点 $(\bar{x}, \bar{y})$ 。将 $(\bar{x}, \bar{y})$ 代入 $\hat{y} = 2.3x + 5$: $\bar{y} = 2.3 \times 10 + 5 = 28$ 。选 A 漏加了截距 5; 选 C、D 是把 $\bar{x}$ 代错 (比如把 $\bar{x}$ 当 10 以外数值)。“回归直线过中心点”是回归分析的标志性结论, 所有求 $\bar{y}$ 、求截距的题目都从它出发。

二、填空题 (本大题共 4 小题, 每小题 5 分, 共 20 分)

Z7

从 6 名同学中任选 3 名参加数学建模比赛, 共有 _ _ 种不同的选法。

答案

20

解析

从 6 人中选 3 人，只选人不排序，是组合问题： $C_6^3 = \frac{6 \times 5 \times 4}{3 \times 2 \times 1} = 20$ 。若误用排列 $A_6^3 = 120$ ，则把“三人团队内部顺序”也数了进去——选人组队，顺序无关，一律用组合。

Z8

从 1, 2, 3, 4, 5, 6 这 6 个数中随机取一个数，已知取到的是偶数，则该数能被 3 整除的概率为__。

答案 $\frac{1}{3}$ **解析**

条件概率。记 $A =$ “取到偶数”， $A = \{2, 4, 6\}$ ， $n(A) = 3$ ；记 $B =$ “能被 3 整除”， $B = \{3, 6\}$ 。 $A \cap B = \{6\}$ ， $n(A \cap B) = 1$ 。由条件概率公式 $P(B | A) = \frac{n(A \cap B)}{n(A)} = \frac{1}{3}$ 。注意：条件概率的分母是“已知发生的事件”的样本点数 3，而不是整个样本空间的点数 6——这是最易错的一步。

Z9

若一组数据 x_1, x_2, x_3, x_4, x_5 的平均数为 3，方差为 2，则数据 $2x_1 + 1, 2x_2 + 1, \dots, 2x_5 + 1$ 的平均数为__，方差为__。

答案

平均数为 7；方差为 8。

解析

数据整体线性变换 $y_i = ax_i + b$ 的数字特征规律：平均数 $\bar{y} = a\bar{x} + b$ ，方差 $s_y^2 = a^2 s_x^2$ （加减常数 b 只平移位置、不改变波动）。这里 $a = 2$ ， $b = 1$ ： $\bar{y} = 2 \times 3 + 1 = 7$ ， $s_y^2 = 2^2 \times 2 = 8$ 。若方差误写成 $2 \times 2 + 1 = 5$ 或 $2 + 1$ ，就是忘掉“方差对缩放取平方、对平移不敏感”。

Z10

某校为调查“是否喜欢数学与性别是否有关”，随机抽取 100 名学生，得到如下 2×2 列联表：

	喜欢数学	不喜欢数学	合计
男生	30	20	50
女生	20	30	50
合计	50	50	100

根据表中数据计算得 $\chi^2 = \underline{\quad} \underline{\quad}$ (保留一位小数)。在犯错误的概率不超过 0.05 的前提下, 可以认为喜欢数学与性别 $\underline{\quad} \underline{\quad}$ (填“有关”或“无关”)。

答案

$\chi^2 = 4.0$; 有关。

解析

独立性检验公式 $\chi^2 = \frac{n(ad - bc)^2}{(a + b)(c + d)(a + c)(b + d)}$ 。这里 $n = 100$, $a = 30$, $b = 20$, $c = 20$, $d = 30$: $\chi^2 = \frac{100 \times (30 \times 30 - 20 \times 20)^2}{50 \times 50 \times 50 \times 50} = \frac{100 \times 500^2}{6250000} = \frac{100 \times 250000}{6250000} = 4.0$ 。4.0 > 3.841 (显著性水平 0.05 下的临界值), 所以在犯错误的概率不超过 0.05 的前提下, 认为喜欢数学与性别有关。注意临界值比较是“ χ^2 越大越拒绝原假设”, 若算出的值小于临界值则不能拒绝, 只能写“没有足够证据认为有关”。

三、解答题 (本大题共 4 小题, 共 50 分)

Z11 (本题满分 12 分)

某校采用分层抽样从高一、高二、高三三个年级共抽取 100 名学生, 调查其每天晚上的睡眠时间, 整理得如下频率分布表:

睡眠时间 (小时)	[5, 6)	[6, 7)	[7, 8)	[8, 9)	[9, 10]
人数	5	25	35	25	10

(1) 估计该校学生每天晚上的平均睡眠时间; (2) 估计该校学生睡眠时间的第 80 百分位数; (3) 若该校高一、高二、高三三个年级的人数之比为 \$2 : 3 : 5\$, 则三个年级各应抽取多少人?

答案

(1) 平均睡眠时间约为 7.6 小时。(2) 第 80 百分位数约为 8.6 小时。(3) 高一抽取 20 人, 高二抽取 30 人, 高三抽取 50 人。

解析

(1) 用各组组中值估计组内取值: 5.5, 6.5, 7.5, 8.5, 9.5。总人数 $n = 100$, 各组的频率依次为 \$0.05, 0.25, 0.35, 0.25, 0.10\$。平均数的估计值为 $\bar{x} = 5.5 \times 0.05 + 6.5 \times 0.25 + 7.5 \times$

$$0.35 + 8.5 \times 0.25 + 9.5 \times 0.10 = 7.6。$$

(2) 先求累计频率：0.05, 0.30, 0.65, 0.90, 1.00。第 80 百分位数是使累计频率达到 0.80 的值：0.80 落在区间 $[8, 9)$ 内（该组累计频率从 0.65 升到 0.90，频率为 0.25）。在组内按比例取位置： $8 + \frac{0.80 - 0.65}{0.25} \times (9 - 8) = 8 + 0.6 = 8.6$ 。即约有 80% 的学生睡眠时间不超过 8.6 小时。

(3) 三个年级人数之比为 $2 : 3 : 5$ ，总份数为 10，样本量 100 按比例分配：高一 $100 \times \frac{2}{10} = 20$ 人，高二 $100 \times \frac{3}{10} = 30$ 人，高三 $100 \times \frac{5}{10} = 50$ 人。

Z12 (本题满分 12 分)

某工厂有甲、乙两条生产线。甲生产线产量占总产量的 60%，次品率为 2%；乙生产线产量占 40%，次品率为 3%。两条生产线的产品混合后统一出厂。

(1) 任取一件出厂产品，求它是次品的概率；(2) 已知取到的一件出厂产品是次品，求它来自甲生产线的概率；(3) 从出厂产品中独立抽取 10 件，求其中恰好有 1 件次品的概率（结果保留 3 位小数），并求这 10 件产品中次品件数的数学期望。

答案

(1) 0.024。(2) $\frac{1}{2}$ 。(3) $P(X = 1) \approx 0.193$ ； $E(X) = 0.24$ 。

解析

设 $A_1 =$ “产品来自甲生产线”， $A_2 =$ “产品来自乙生产线”， $B =$ “产品是次品”。则 $P(A_1) = 0.6$ ， $P(A_2) = 0.4$ ， $P(B | A_1) = 0.02$ ， $P(B | A_2) = 0.03$ 。

(1) 由全概率公式： $P(B) = P(A_1)P(B | A_1) + P(A_2)P(B | A_2) = 0.6 \times 0.02 + 0.4 \times 0.03 = 0.012 + 0.012 = 0.024$ 。

(2) 由条件概率公式（贝叶斯思想）： $P(A_1 | B) = \frac{P(A_1)P(B | A_1)}{P(B)} = \frac{0.012}{0.024} = 0.5$ 。

(3) 10 件产品独立抽取，每件次品率恒为 0.024，故次品件数 $X \sim B(10, 0.024)$ ： $P(X = 1) = C_{10}^1 \times 0.024 \times (1 - 0.024)^9 = 10 \times 0.024 \times 0.976^9 \approx 0.193$ 。数学期望 $E(X) = np = 10 \times 0.024 = 0.24$ 。

Z13 (本题满分 13 分)

某公司统计了近 6 个月的广告投入 x （万元）与当月销售额 y （万元）的成对数据：

月份	1	2	3	4	5	6
广告投入 x (万元)	1	2	3	4	5	6
销售额 y (万元)	10	12	15	17	20	22

(1) 求销售额 y 关于广告投入 x 的回归直线方程（系数保留两位小数）；(2) 求样本相关系数 r （保留三位小数），并据此说明 x 与 y 相关的强弱与方向；(3) 用回归方程预测广告投

入为 7 万元时的销售额；(4) 有同学说：“既然 r 接近 1，就说明广告投入一定导致销售额增长。”你同意吗？请说明理由。

答案

(1) $\hat{y} = 2.46x + 7.40$ 。(2) $r \approx 0.998$, x 与 y 高度正相关。(3) 预测销售额约为 24.62 万元。(4) 不同意。强相关只说明两个变量“同步变化”的趋势很强，不等于“广告投入导致销售额增长”；销售额还可能受季节、市场行情、促销活动等共同因素影响，因果结论需要控制变量或随机对照实验等更严格证据。

解析

(1) 先算样本中心点： $\bar{x} = \frac{1+2+3+4+5+6}{6} = 3.5$, $\bar{y} = \frac{10+12+15+17+20+22}{6} = 16$ 。计算离差乘积与离差平方和： $\sum_{i=1}^6 (x_i - \bar{x})^2 = 2.5^2 + 1.5^2 + 0.5^2 + 0.5^2 + 1.5^2 + 2.5^2 = 17.5$ ； $\sum_{i=1}^6 (x_i - \bar{x})(y_i - \bar{y}) = 2.5 \times 6 + 1.5 \times 4 + 0.5 \times 1 + 0.5 \times 1 + 1.5 \times 4 + 2.5 \times 6 = 43$ 。于是 $\hat{b} = \frac{43}{17.5} \approx 2.46$, $\hat{a} = \bar{y} - \hat{b}\bar{x} = 16 - 2.46 \times 3.5 = 16 - 8.61 = 7.39 \approx 7.40$ 。回归直线方程 $\hat{y} = 2.46x + 7.40$ 。

(2) $\sum_{i=1}^6 (y_i - \bar{y})^2 = 36 + 16 + 1 + 1 + 16 + 36 = 106$, $r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \cdot \sum (y_i - \bar{y})^2}} = \frac{43}{\sqrt{17.5 \times 106}} \approx \frac{43}{43.07} \approx 0.998$ 。 $r > 0$ 说明正相关， $|r|$ 接近 1 说明相关程度很强。

(3) 令 $x = 7$: $\hat{y} = 2.46 \times 7 + 7.40 = 17.22 + 7.40 = 24.62$ (万元)。

(4) 理由见【答案】。相关性度量的是“同向变化的一致性”，不提供因果方向的证据；要下因果结论须排除共同原因（如夏天同时拉动广告投放与销售额的市场景气）。

为什么错

本题第 (4) 问是最典型的失分点，常见错误写法有三类：

- **把“相关”写成“因果”**：“因为 $r = 0.998$ ，所以广告投入导致销售额增长。”——错在 r 只度量相关性，任何相关结论都不能直接推出因果关系；
- **忽略共同原因**：“广告和销售一起涨，一定是广告的功劳。”——可能两者都由同一原因（市场景气、节日大促）驱动；
- **过度外推**：用样本外数据（如 $x = 50$ ）随意预测而不说明“样本范围内预测才可靠”。

正确姿势：承认强正相关，同时强调“相关不等于因果”，需要设计对照实验或控制变量才能谈因果。这一问考查的正是第 04 章全章的素养落点——会读相关系数，更会守住结论的边界。

Z14 (本题满分 13 分)

某水果店计划每天进货某种时令水果。该水果进价每斤 2 元，售价每斤 4 元；若当天卖不完，剩余部分按每斤 1 元处理（即每剩余 1 斤净亏 1 元）。根据历史销售记录，该水果的

日需求量 X (斤) 的分布列为:

日需求量 X (斤)	80	90	100	110
概率 P	0.2	0.3	0.3	0.2

店主考虑两种进货方案: 方案一, 每天进货 100 斤; 方案二, 每天进货 90 斤。

(1) 若采用方案一, 写出当天利润 Y (元) 的分布列, 并求 $E(Y)$; (2) 若采用方案二, 求 $E(Y)$, 并据此判断哪种进货方案更优; (3) 从数学建模的角度反思: 上述决策模型忽略了哪些现实因素? 若要进一步提高决策质量, 应如何改进模型? (至少各写出两条)

答案

(1) 方案一的利润分布列: $Y = 140$ ($P = 0.2$), $Y = 170$ ($P = 0.3$), $Y = 200$ ($P = 0.5$); $E(Y) = 179$ 元。(2) 方案二的利润分布列: $Y = 150$ ($P = 0.2$), $Y = 180$ ($P = 0.8$); $E(Y) = 174$ 元。 $179 > 174$, 方案一 (进货 100 斤) 更优。(3) 开放题, 示例: 模型忽略了水果保鲜期与变质损耗、进货运输与仓储成本、节假日与天气导致的需求分布变化、资金占用成本、因缺货而流失的顾客价值等。改进方向: 把需求分布按季节/节假日分段估计并动态更新, 加入损耗率与资金成本参数, 引入“缺货损失”与“库存持有成本”重新建模, 必要时用模拟或历史数据回测验证。

解析

(1) 方案一进货 100 斤, 成本固定为 $100 \times 2 = 200$ 元。

- 日需求 80 斤 ($P = 0.2$): 卖出 80 斤收入 $80 \times 4 = 320$ 元, 剩余 20 斤处理收入 $20 \times 1 = 20$ 元, 利润 $Y = 320 + 20 - 200 = 140$ 元;
- 日需求 90 斤 ($P = 0.3$): 卖出 90 斤收入 360 元, 剩余 10 斤处理收入 10 元, 利润 $Y = 360 + 10 - 200 = 170$ 元;
- 日需求 100 或 110 斤 ($P = 0.3 + 0.2 = 0.5$): 全部售出, 收入 $100 \times 4 = 400$ 元, 利润 $Y = 400 - 200 = 200$ 元。

故分布列为 $P(Y = 140) = 0.2$, $P(Y = 170) = 0.3$, $P(Y = 200) = 0.5$, $E(Y) = 140 \times 0.2 + 170 \times 0.3 + 200 \times 0.5 = 28 + 51 + 100 = 179$ 元。

(2) 方案二进货 90 斤, 成本 $90 \times 2 = 180$ 元。

- 日需求 80 斤 ($P = 0.2$): 卖出 80 斤收入 320 元, 剩余 10 斤处理收入 10 元, 利润 $Y = 320 + 10 - 180 = 150$ 元;
- 日需求 90、100、110 斤 ($P = 0.3 + 0.3 + 0.2 = 0.8$): 全部售出, 收入 $90 \times 4 = 360$ 元, 利润 $Y = 360 - 180 = 180$ 元。

$E(Y) = 150 \times 0.2 + 180 \times 0.8 = 30 + 144 = 174$ 元。比较期望利润 $179 > 174$ ，方案一更优。注意“期望更大”指长期平均每天多赚 5 元，不等于方案一每天都赚得更多（需求为 80、90 时方案二反而更稳），这正是“用期望做决策”的含义。

(3) 见【答案】。此问对应第 05 章建模四步法中的“检验与修正”：模型只是现实的近似，必须回头审视被忽略的因素，必要时回改模型。

为什么错

本题常见失分点：

- **忘记剩余水果的亏损**：卖不完的 20 斤只算“没赚”而不算“亏 1 元/斤”，会把方案一的低需求利润错算成 160 而非 140。处理价 1 元低于进价 2 元，每剩 1 斤净亏 1 元，必须扣掉；
- **把“最可能发生的需求”当答案**：误以为“日需求 100 斤概率最大，所以方案一最优”——决策要看长期平均（期望），而不是只看概率最高的一天；
- **分布列概率漏项或加起来不等于 1**：方案一中需求 100、110 两行合并为 $P = 0.5$ 时若漏加，分布列会不完整，期望随之算错；
- **第 (3) 问空泛**：只写“考虑实际情况”不举例，会被扣分。建模反思要具体到“哪个因素、影响哪一项、怎么改进”。

全卷答案速查

题号	Z1	Z2	Z3	Z4	Z5	Z6	Z7	Z8	Z9	Z10
答案	A	B	A	B	B	B	20	$\frac{1}{3}$	7; 8	4.0; 有关