# 第六章　计数原理

**重点与难点**

1．分类加法计数原理

2．分步乘法计数原理

3．排列与排列数

4．组合与组合数

5．捆绑法与插空法

6．二项式定理

7．二项式系数的性质

## 6．1　分类加法计数原理与分步乘法计数原理

### 第 1 课时　分类加法计数原理与分步乘法计数原理

预习评价▶新知初探

**学习目标**

1．了解计数问题，理解分类加法计数原理与分步乘法计数原理．

2．会用分类加法计数原理与分步乘法计数原理进行简单的计数．

**问题导引**

1．分类加法计数原理与分步乘法计数原理的主要作用是什么？

2．分类加法计数原理与分步乘法计数原理相同点和不同点分别是什么？

**现学现用**

1．概念辨析(正确的打“√”，错误的打“×”)

(1) 在分类加法计数原理中，两类不同方案中的方法可以相同．(　×　)

(2) 在分类加法计数原理中，每类方案中的方法都能完成这件事．(　√　)

(3) 在分步乘法计数原理中，事情若是分两步完成的，那么其中任何一个单独的步骤都不能完成这件事，只有两个步骤都完成后，这件事情才算完成．(　√　)

2．某中学需从 2021 年师范大学毕业的 3 名女大学生和 2 名男大学生中选聘 1 人，则不同的选法种数为(　B　)

A．6　　　　B．5

C．3　　　　　　　　D．2

3．从 $A$ 地到 $B$ 地，可乘汽车、火车、轮船三种交通工具，如果一天内汽车有 3 班，火车有 4 班，轮船有 2 班，那么一天内乘坐这三种交通工具的不同走法数为(　B　)

A．1＋1＋1＝3　　　　　　　　B．3＋4＋2＝9

C．3×4×2＝24　　　　　　　　D．以上都不对

4．教学大楼共有四层，每层都有东西两个楼梯，由一层到四层的走法共有(　A　)

A．8 种　　　　　　　　B．23 种

C．42 种　　　　　　　　D．24 种

【解析】 由一层到二层有 2 种选择，二层到三层有 2 种选择，三层到四层有 2 种选择．

## 对点练 ▶ 先掌握基础

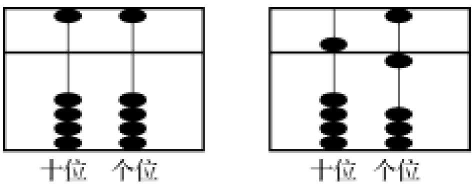
类型 1　分类加法计数原理的应用

例1　(1) 三个人踢毽子，互相传递，每人每次只能踢一下，由甲开始踢，经过 4 次传递后，毽子又被踢回给甲，则不同的传递方式共有(　B　)

A．4 种　　　　　　　　B．6 种

C．10 种　　　　　　　　D．16 种

(2) 现有一种算盘(如图一)，共两档，自右向左分别表示个位和十位，档中横以梁，梁上一珠拨下，记作数字 5，梁下四珠，上拨每珠记作数字 1(例如图二中算盘表示整数 51)．如果拨动图一算盘中的三枚算珠，可以表示不同整数的个数为(　C　)



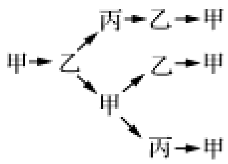

图一　　　　图二

(例 1(2))

A．16　　　　　　　　B．15

C．12　　　　　　　　D．10

【解析】(1) 分两类：甲第一次踢给乙时，满足条件的有 3 种传递方式(如图)，同理，甲先踢给丙时，满足条件也有 3 种传递方式．由分类加法计数原理，共有 $3+3=6$ 种传递方式．




(例 1(1))

(2) 由题意，拨动三枚算珠，有四种拨法：

①十位拨动 0 枚，个位拨动 3 枚，有 2 种结果：7 和 3；②十位拨动 1 枚，个位拨动 2 枚，有 4 种结果：12,16,52,56；③十位拨动 2 枚，个位拨动 1 枚，有 4 种结果：21,25,61,65；④十位拨动 3 枚，个位拨动 0 枚，有 2 种结果：30,70.综上，拨动图一算盘中的三枚算珠，可以表示不同整数的个数为 12.

【规律总结】 用分类加法计数原理解题应注意以下问题：

(1) 明确题目中所指的“完成一件事”是什么事，完成这件事可以有哪些办法，怎样才算完成这件事．

(2) 分类加法计数原理中的“分类”要全面、不能遗漏，但也不能重复、交叉．

(3) 若完成某件事情有 $n$ 类办法，则它们两两的交集为空集，并集为全集．

类型 2　分步乘法计数原理的应用

**例2**　已知集合 $M=\{-3，-2，-1,0,1,2\}$，$P(a，b)(a，b\in M)$表示平面上的点，问：

(1) 点 $P$ 可表示平面上多少个不同的点？

(2) 点 $P$ 可表示平面上第二象限内多少个不同的点？

【解答】 (1) 确定平面上的点 $P(a，b)$，可分两步完成：第一步确定 $a$ 的值，

有 6 种不同方法；第二步确定 $b$ 的值，也有 6 种不同方法．根据分步乘法计数原理，得点 $P$ 可表示平面上不同的点共有 $6\times6=36$ 个．

(2) 确定平面上第二象限内的点 $P(a，b)$，可分两步完成：第一步确定 $a$ 的值，因为 $a<0$，所以有 3 种不同方法；第二步确定 $b$ 的值，因为 $b>0$，所以有 2 种不同方法．由分步乘法计数原理，得点 $P$ 可表示平面上第二象限内不同的点共有 $3\times2=6$ 个．

【规律总结】 利用分步乘法计数原理应注意：

(1) 要按事件发生的过程合理分步，即分步是有先后顺序的．

(2) “步”与“步”之间是连续的、不间断的、缺一不可的，但也不能重复、交叉．

(3) 若完成某件事情需 $n$ 步，则必须且只需依次完成这 $n$ 个步骤后，这件事情才算完成．

类型 3　两个计数原理的结合应用

例3　某座山，若从东侧通往山顶的道路有 3 条，从西侧通往山顶的道路有 2 条，那么游人从上山到下山共有多少种不同的走法？

【解答】 完成从上山到下山这件事可分为四类：①从东侧上山，且从东侧下山，走法有 $3\times3$ 种；②从东侧上山，从西侧下山，走法有 $3\times2$ 种；③从西侧上山，从东侧下山，走法有 $2\times3$ 种；④从西侧上山，且从西侧下山，走法有 $2\times2$ 种．据分类计数原理知，符合条件的走法共有 $3\times3+3\times2+2\times3+2\times2=25$ 种．

变式　现有 5 幅不同的国画，2 幅不同的油画，7 幅不同的水彩画．

(1) 从这些国画、油画、水彩画中各选一幅布置房间，有几种不同的选法？

(2) 从这些画中选出两幅不同种类的画布置房间，有几种不同的选法？

【解答】 (1) 分为三步：国画、油画、水彩画分别有 5 种、2 种、7 种不同的选法，根据分步乘法计数原理，共有 $5\times2\times7=70$ 种不同的选法．

(2) 分为三类：第一类是一幅选自国画，一幅选自油画．由分步乘法计数原理知，有 $5\times2=10$ 种不同的选法．

第二类是一幅选自国画，一幅选自水彩画，有 $5\times7=35$ 种不同的选法．第三类是一幅选自油画，一幅选自水彩画，有 $2\times7=14$ 种不同的选法．所以共有 $10+35+14=59$ 种不同的选法．

【规律总结】 两个计数原理结合解决问题的一般思路：

(1) 弄清完成一件事是做什么．

(2) 确定是先分类后分步，还是先分步后分类．

(3) 弄清分步、分类的标准是什么．

(4) 利用两个计数原理求解．

课堂评价▶基础达标

1．把 10 个苹果分成三堆，要求每堆至少 1 个，至多 5 个，则不同的分法共有(　A　)

A．4 种　　B．5 种

C．6 种　　D．7 种

【解析】 分类：三堆中“最多”的一堆为 5 个，其他两堆总和为 5，每堆至少 1 个，只有 2 种分法，即 1 和 4,2 和 3 两种方法．三堆中“最多”的一堆为 4 个，其他两堆总和为 6，每堆至少 1 个，只有 2 种分法，即 2 和 4,3 和 3 两种方法．三堆中“最多”的一堆为 3 个，那是不可能的．所以不同的分法共有 $2+2=4$ 种．

2．晓芳有 4 件不同颜色的衬衣，3 件不同花样的裙子，另有 2 套不同样式的连衣裙．“五一”节需选择一套服装参加歌舞演出，则晓芳选择穿衣服的不同方式有(　B　)

A．24 种　　B．14 种

C．10 种　　D．9 种

【解析】 首先分两类．第一类是穿衬衣和裙子，由分步乘法计数原理知共有

$4\times3=12$ 种．第二类是穿连衣裙有 2 种，所以由分类加法计数原理知共有 $12+2=14$ 种不同的穿衣服方式．

3．古代“五行”学认为：“物质分金、木、土、水、火五种属性，金克木，木克土，土克水，水克火，火克金．”将五种不同属性的物质任意排成一列，但排列中属性相克的两种物质不相邻，则这样的排列方法共有( B )

A．5 种　　B．10 种

C．20 种　　D．120 种

【解析】 由题意，可看作五个位置排列五种事物，第一个位置有五种排列方法，不妨假设排的是金，则第二步只能从土与水两者中选一种排列，故有两种选择，不妨假设排的是水，第三步只能排木，第四步只能排火，第五步只能排土，故总的排列方法种数有 $5\times2\times1\times1\times1=10$ 种．

4．若 3 张不同的电影票全部分给 10 个人，每人至多一张，则不同分法的种数是__720__．

【解析】 由题意知，因为 3 张不同的电影票全部分给 10 个人，每人至多一张，所以第一张有 10 种不同的分法，第二张有 9 种不同的分法，第三张有 8 种不同的分法，由分步乘法计数原理知有 $10\times9\times8=720$ 种不同的分法．

5．有一项活动，需要在 3 名老师、8 名男同学和 5 名女同学中选人参加．

(1) 若只需 1 人参加，则有__16__种不同的选法．

(2) 若需要老师、男同学、女同学各 1 人参加，则有__120__种不同的选法．

(3) 若需要 1 名老师、1 名学生参加，则有__39__种不同的选法．

【解析】 (1) 只需一人参加，有三类：第一类选老师，有 3 种不同的选法；第二类选男生，有 8 种不同的选法；第三类选女生，有 5 种不同的选法．共有 $3+8+5=16$ 种不同的选法．

(2) 需老师、男同学、女同学各一人，则分 3 步，第一步选老师，有 3 种不同的选法；第二步选男生，有 8 种不同的选法；第三步选女生，有 5 种不同的选

法．共有 $3\times8\times5=120$ 种不同的选法．

(3) 第一步选老师，有 3 种不同的选法；第二步选学生，有 $8+5=13$ 种不同的选法．共有 $3\times13=39$ 种不同的选法．

## 第 2 课时　两个基本原理的应用

### 预习评价▶新知初探

**学习目标**

1．进一步理解分类计数原理与分步计数原理．

2．会用分类计数原理与分步计数原理综合解决计数问题．

**问题导引**

1．用分类加法计数原理与分步乘法计数原理解题的步骤是什么？

2．用分类加法计数原理与分步乘法计数原理计数，你有哪些方法？

**现学现用**

1．判断正误(正确的打“√”，错误的打“×”)

(1) 从甲地到乙地有两类交通方式：坐飞机和乘轮船，其中飞机每天有 3 班，轮船有 4 班．若李先生从甲地去乙地，则不同的交通方式共有 7 种．(　√　)

(2) 现有 4 件不同款式的上衣和 3 条不同颜色的长裤，若一条长裤与一件上衣配成一套，则不同的配法种数为 12.(　√　)

(3) 从 $A$ 地到 $B$ 地要经过 $C$ 地和 $D$ 地，从 $A$ 地到 $C$ 地有 3 条路，从 $C$ 地到 $D$ 地有 2 条路，从 $D$ 地到 $B$ 地有 4 条路，则从 $A$ 地到 $B$ 地不同走法的种数是 24.(　√　)

2．完成一项工作，有两种方法，有 5 个人只会第一种方法，另外有 4 个人只会第二种方法，从这 9 个人中选 1 个人完成这项工作，则不同的选法共有(　C　)

A．5 种　　B．4 种

C．9 种　　D．45 种

3．将 3 名防控新冠疫情志愿者全部分配给 2 个不同的社区服务，不同的分配方案有(　C　)

A．12 种　　B．9 种

C．8 种　　D．6 种

【解析】 每名志愿者都有两种不同的分配方法，根据分步计数原理可知，不同的分配方案有 $2^3=8$ 种．

### 对点练▶先掌握基础

类型 1　两个计数原理在组数中的应用

例1　用 0,1，…，9 这十个数字，可以组成多少个：

(1) 三位整数？

(2) 无重复数字的三位整数？

【解答】 由于 0 不可在最高位，因此应对它进行单独考虑.

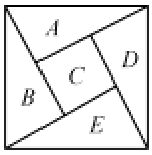
(1) 百位数字有 9 种选择，十位数字和个位数字都各有 10 种选择．由分步乘法计数原理知，满足题意的三位数共有 $9\times10\times10=900$ 个.

(2) 由于数字不可重复，可知百位数字有 9 种选择，十位数字也有 9 种选择，但个位数字仅有 8 种选择．由分步乘法计数原理知，满足题意的三位数共有 $9\times9\times8=648$ 个.

【规律总结】排数问题实际就是分步问题，需要用分步乘法计数原理解决．此题中，由于数字“0”的出现，又进行了分类讨论，即在解决相关的排数问题时，要注意两个计数原理的综合应用.

类型 2　两个计数原理在涂色中的应用

例2　如图为我国数学家赵爽(约 3 世纪初)在为《周髀算经》作注时验证勾股定理的示意图，现在提供 6 种颜色给其中 5 个小区域涂色，规定每个区域只涂一种颜色且相邻颜色不同，求不同涂色的方法种数.



(例 2)

【解答】 分 4 步进行分析：

①，对于区域 $A$，有 6 种颜色可选．②，对于区域 $B$，与 $A$ 区域相邻，有 5 种颜色可选．③，对于区域 $C$，与 $A$、$B$ 区域相邻，有 4 种颜色可选．④，对于区域 $D$、$E$，若 $D$ 与 $B$ 颜色相同，$E$ 区域有 4 种颜色可选；若 $D$ 与 $B$ 颜色不相同，$D$ 区域有 3 种颜色可选，$E$ 区域有 3 种颜色可选，则区域 $D$、$E$ 有 $4+3\times3=13$

种选择．综上，不同的涂色方案有 $6\times5\times4\times13=1\ 560$ 种．

【规律总结】 解决涂色问题的一般思路：

(1) 按涂色的顺序分步进行，用分步乘法计数原理计数．

(2) 按颜色恰当选取情况分类，用分类加法计数原理计数．

(3) 注意相邻区域同色与不同色的情况．

类型 3　放球(填数字、排序)问题

例3　(1) 将编号 1,2,3,4 的小球放入编号为 1,2,3 的盒子中，要求不允许有空盒子，且球与盒子的号不能相同，则放球方法有(　B　)

A．16 种　　　　B．12 种

C．9 种　　　　D．6 种

(2) 将 1,2,3，…，9 这 9 个数字填在如图的 9 个空格中，要求每一行从左到右，每一列从上到下分别依次增大．当 3,4 固定在图中的位置时，填写空格的方法有(　A　)

| | | |
|---|---|---|
| | | |
| 3 | 4 | |
| | | |

(例 3(2))

A．6 种　　　　B．12 种

C．18 种　　　　D．24 种

【解析】 (1) 由题意可知，这四个小球有两个小球放在一个盒子中，当 1 与 2 号球放在同一盒子中时，有 2 种不同的放法；当 1 与 3 号球放在同一盒子中时，有 2 种不同的放法；当 1 与 4 号球放在同一盒子中时，有 2 种不同的放法；当 2 与 3 号球放在同一盒子中时，有 2 种不同的放法；当 2 与 4 号球放在同一盒子中时，有 2 种不同的放法；当 3 与 4 号球放在同一盒子中时，有 2 种不同的放法．因此，不同的放球方法有 12 种．

(2) 因为每一行从左到右，每一列从上到下分别依次增大，1,2,9 只有一种填

法，5 只能填在右上角或左下角，5 填好后与之相邻的空格可填 6,7,8 任一个；余下两个数字按从小到大只有一种填法．共有 $2\times3=6$ 种方法．

【规律总结】 解答此类问题，首先必须弄清是“分类”还是“分步”，其次要搞清“分类”或“分步”的具体标准是什么，选择合理的标准处理事件，关键是看能否独立完成这件事，避免计数的重复或遗漏．

课堂评价▶基础达标

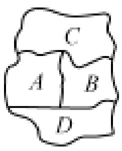
1．汽车维修师傅在安装好汽车轮胎后，需要紧固轮胎上的五个螺栓，记为 $A$，$B$，$C$，$D$，$E$(在正五边形的顶点上)，紧固时需要按一定的顺序固定每一个螺栓，但不能连续固定相邻的两个，则不同固定螺栓顺序的种数为(　C　)

A．20　　B．15

C．10　　D．5

【解析】 此题相当于在正五边形中，对五个字母排序，要求五边形的任意相邻两个字母不能排在相邻位置．考虑 $A$ 放第一个位置，第二步只能放 $C$ 或 $D$，只有 $ACEBD$ 和 $ADBEC$ 两种情况；同理，分别让 $B$，$C$，$D$，$E$ 放第一个位置，则各有两种情况，共 $2\times5=10$ 种情况．

2．用 5 种不同颜色对如图所示的四个部分进行着色，要求有公共边界的两块不能用同一种颜色，则不同的着色方法共有(　B　)




(第 2 题)

A．150 种　　B．180 种

C．240 种　　D．120 种

【解析】 分步涂色，第一步对 $A$ 涂色有 5 种方法，第二步对 $B$ 涂色有 4 种方法，第三步对 $C$ 涂色有 3 种方法，第四步对 $D$ 涂色有 3 种方法，所以总的方法数为 $5\times4\times3\times3=180$．

3．某年级要从 3 名男生，2 名女生中选派 3 人参加某次社区服务，如果要求至少有 1 名女生，那么不同的选派方案有(　D　)

A．6 种　　　　B．7 种

C．8 种　　　　D．9 种

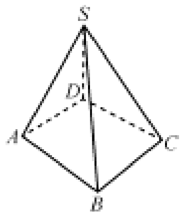
**【解析】** 可按女生人数分类：若选派 1 名女生，有 $2\times3=6$ 种方案；若选派 2 名女生，有 3 种方案．由分类加法计数原理知，共有 9 种不同的选派方案．

4．用 0,1,2,3,4,5 可以组成无重复数字的比 2 000 大的四位奇数有 <u>120</u> 个．

**【解析】** 按末位是 1,3,5 分三类计数，第一类：末位是 1，共有 $4\times4\times3=48$ 个；第二类：末位是 3，共有 $3\times4\times3=36$ 个；第三类：末位是 5，共有 $3\times4\times3=36$ 个．由分类加法计数原理知，共有 $48+36+36=120$ 个．

5．给一个四棱锥的每个顶点染上一种颜色，并使同一条棱上的两端异色，若只有 5 种颜色可供使用，则不同的染色方法共有 <u>420</u> 种．

**【解析】** 如图，按 $S\to A\to B\to C\to D$ 依次染色，当 $A$，$C$ 同色时有 $5\times4\times3\times1\times3=180$ 种．当 $A$，$C$ 不同色时，有 $5\times4\times3\times2\times2=240$ 种．因此共有 $180+240=420$ 种．




(第 5 题)

# 6．2　排列与组合

## 第1课时　排　　列

预习评价▶新知初探

**学习目标**

1．了解排列及排列数的意义．

2．会用枚举法或计数原理计算排列数．

**问题导引**

1．排列的概念是什么？两个排列是相同排列的条件是什么？

2．排列数的定义是什么？

**现学现用**

1．概念辨析(正确的打“√”，错误的打“×”)

(1) $a$，$b$，$c$与$b$，$a$，$c$是同一个排列．(　×　)

(2) 在一个排列中，同一个元素不能重复出现．(　√　)

(3) 在一个排列中，若交换两个元素的位置，则该排列不发生变化．(　×　)

2．从甲、乙、丙三人中选两人站成一排的所有站法为<u>甲乙，甲丙，乙甲，乙丙，丙甲，丙乙</u>.

3．在$A$，$B$，$C$，$D$四位学生中，选出两人担任正、副班长，共有选法(　B　)

A．4种　　　　B．12种

C．$4^2$种　　　　D．$2^4$种

4．(多选)下列问题中是排列问题的有(　AD　)

A．从甲、乙、丙三名同学中选出两名分别参加数学和物理学习小组

B．从甲、乙、丙三名同学中选出两名同学参加一项活动

C．从$a$，$b$，$c$，$d$四个字母中取出两个字母

D．从1,2,3,4,5五个数字中取出3个数字组成一个三位数

## 对点练▶先掌握基础

**类型1　判断是否为排列问题**

**例1**　(1) 北京、上海、天津三个民航站之间的直达航线的飞机票的价格(假设来回的票价相同)；

(2) 选2个小组分别去植树和种菜；

(3) 选 2 个小组去种菜；

(4) 选 10 人组成一个学习小组；

(5) 选 3 个人分别担任班长、学习委员、生活委员；

(6) 某班 40 名学生在假期相互通信.

【解答】 (1) 中票价只有三种，虽然机票是不同的，但票价是一样的，不存在顺序问题，所以不是排列问题.

(2) 植树和种菜是不同的，存在顺序问题，属于排列问题.

(3)(4) 不存在顺序问题，不属于排列问题.

(5) 中每个人的职务不同，例如甲当班长或当学习委员是不同的，存在顺序问题，属于排列问题.

(6) $A$ 给 $B$ 写信与 $B$ 给 $A$ 写信是不同的，所以存在顺序问题，属于排列问题.

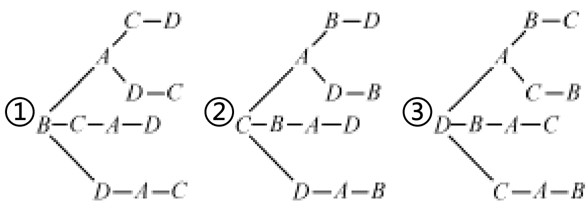
所以在上述各题中(2)(5)(6)属于排列问题.

【规律总结】 判断一个具体问题是否为排列问题，就看取出元素后排列是有序的还是无序的，而检验它是否有序的依据就是变换元素的“位置”(这里的“位置”应视具体问题的性质和条件来决定)，看其结果是否有变化，有变化就是排列问题，无变化就不是排列问题.

类型 2 “树状图”(列表法)枚举排列问题

例2 用 $A$，$B$，$C$，$D$ 四个不同的元素排成一列，$A$ 不在两端的排法有多少种？请列举出来.

【解答】 画树状图如下图：




(例 2)

故有 $BACD$，$BADC$，$BCAD$，$BDAC$，$CABD$，$CADB$，$CBAD$，$CDAB$，$DABC$， *DACB*，*DBAC*，*DCAB*，共 12 种．

【规律总结】在画树状图时，先以安排哪个元素在首位为分类标准进行分类，在每类中，再按余下元素在前面元素不变的情况下定第二位并按序分类，依次一直进行到完成一个排列，最后应把所有排列列举出来．

**变式** 将玫瑰花、月季花、莲花各一束分别送给甲、乙、丙三人，每人一束，共有多少种不同的分法？请将它们列出来．

【解答】 按分步乘法计数原理的步骤：

第一步，分给甲，有 3 种分法；第二步，分给乙，有 2 种分法；第三步，分给丙，有 1 种分法．故共有 $3\times2\times1=6$ 种不同的分法．列出这 6 种分法，如下表所示：

| 甲 | 乙 | 丙 |
|---|---|---|
| 玫瑰花 | 月季花 | 莲花 |
| 玫瑰花 | 莲花 | 月季花 |
| 月季花 | 玫瑰花 | 莲花 |
| 月季花 | 莲花 | 玫瑰花 |
| 莲花 | 玫瑰花 | 月季花 |
| 莲花 | 月季花 | 玫瑰花 |

课堂评价▶基础达标

1．已知下列问题：①从 1 到 10 十个自然数中任取两个数组成直角坐标平面内的点的坐标，可得多少个不同的点的坐标？②从 10 名同学中任抽两名同学去学校开座谈会，有多少种不同的抽取方法？③某商场有四个大门，若从一个门进去，购买物品后再从另一个门出来，不同的出入方式共有多少种？其中是排列问题的个数为(　C　)

A．0　　B．1

C．2　　D．3

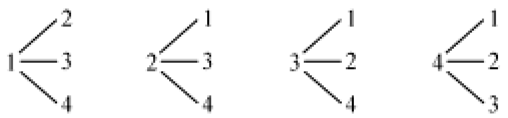
【解析】 ①因为取出的两数组成的点的坐标与哪一个数作横坐标，哪一个数作纵坐标的顺序有关，所以这是一个排列问题．②因为从 10 名同学中抽取两人去学校开座谈会的方式不用考虑两人的顺序，所以这不是排列问题．③因为从一门进，从另一门出是有顺序的，所以是排列问题．综上，①③是排列问题，②不是排列问题．

2．从 1,2,3,4 这四个数字中任取两个不同的数字，则可组成的不同两位数有(　B　)

A．9 个　　　　B．12 个

C．15 个　　　　D．18 个

【解析】 用树状图表示如图，由此可知共有 12 个．



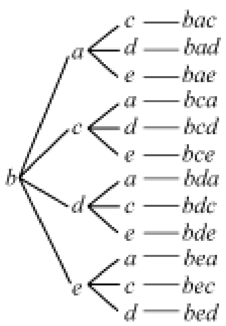

(第 2 题)

3．从 $a$，$b$，$c$，$d$，$e$ 五个元素中每次取出三个元素，可组成以 $b$ 为首的不同排列的个数为(　C　)

A．6　　　　B．10

C．12　　　　D．16

【解析】 画出树状图如图，可知共 12 个，它们分别是 $bac$，$bad$，$bae$，$bca$，$bcd$，$bce$，$bda$，$bdc$，$bde$，$bea$，$bec$，$bed$.




(第 3 题)

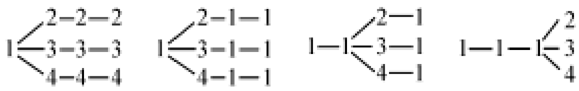
4．由 1,2,3,4 这四个数字组成的首位数字是 1，且恰有三个相同数字的四位数

有(　B　)

A．9 个　　　　　　　　　　　　B．12 个

C．15 个　　　　　　　　　　　　D．18 个

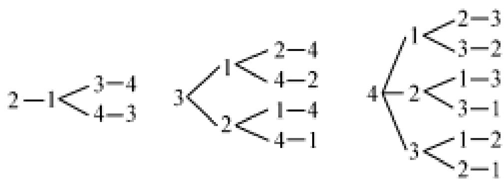
【解析】 用树状图表示如图，由此可知共有 12 个．



(第 4 题)

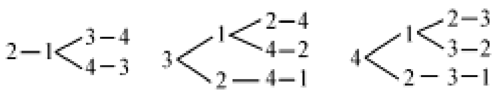
5．在 1,2,3,4 的排列 $a_1a_2a_3a_4$ 中，满足 $a_1>a_2$，$a_3>a_2$，$a_3>a_4$ 的排列的个数是＿5＿．

【解析】 首先注意 $a_1$ 位置的数比 $a_2$ 位置的数大，可以借助树形图进行筛选．满足 $a_1>a_2$ 的树状图如图(1)：



(第 5 题(1))

其次满足 $a_3>a_2$ 的树状图如图(2)：



(第 5 题(2))

最后满足 $a_3>a_4$ 的排列有 2143,3142,3241,4132,4231，共 5 个．

第 2 课时　排列数

预习评价▶新知初探

**学习目标**

1．了解排列的应用．

2．理解排列数公式的推导并应用．

**问题导引**

1．什么是排列数公式？如何推导？

2．有关排列计数应用题的解题步骤是什么？要注意哪些问题？

**现学现用**

1．概念辨析(正确的打“√”，错误的打“×”)

(1) 因为排列数的阶乘式是一个分式，所以其化简的结果不一定是整数．(　×　)

(2) 若 $A_n^m=10\times9\times8\times7\times6$，则 $n=10$，$m=6$.(　×　)

(3) $n!=1\times2\times3\times\cdots\times(n-1)\times n$．(　√　)

2．$A_4^2=$__12__，$A_3^3=$__6__．

3．已知 $A_n^2=132$，则 $n$ 等于(　A　)

A．12　　　　B．24

C．36　　　　D．48

【解析】 由已知得 $n(n-1)=132$，即 $n^2-n-132=0$，所以 $n=12$ 或 $n=-11$(舍去)．

4．在新冠肺炎疫情防控期间，某记者要去武汉 4 个方舱医院采访，则不同的采访顺序有(　D　)

A．4 种　　　　B．12 种

C．18 种　　　　D．24 种

**对点练▶先掌握基础**

类型 1　排列数公式及其应用

例1　(1) 计算：$\dfrac{2A_8^5+7A_8^4}{A_8^8-A_9^5}$；

(2) 解方程 $3A_x^3=2A_{x+1}^2+6A_x^2$；

(3) 解不等式 $A_8^{x+2}<6A_8^x$.

【解答】 (1) $\dfrac{2A_8^5+7A_8^4}{A_8^8-A_9^5}$

$=\dfrac{2\times8\times7\times6\times5\times4+7\times8\times7\times6\times5}{8\times7\times6\times5\times4\times3\times2\times1-9\times8\times7\times6\times5}$

$=\dfrac{8\times7\times6\times5\times(8+7)}{8\times7\times6\times5\times(24-9)}=1.$

(2) 由 $3A_x^3=2A_{x+1}^2+6A_x^2$,

得 $3x(x-1)(x-2)=2(x+1)x+6x(x-1)$.

因为 $x\geqslant3$，且 $x\in\mathbf{N}^*$,

所以 $3(x-1)(x-2)=2(x+1)+6(x-1)$,

即 $3x^2-17x+10=0$,

解得 $x=5$ 或 $x=\dfrac{2}{3}$(舍去)，所以 $x=5$.

(3) 原不等式可化为$\dfrac{8!}{(8-x-2)!}<6\times\dfrac{8!}{(8-x)!}$,

即 $x^2-15x+50<0$,

即$(x-5)(x-10)<0$,

得 $5<x<10$.又$\begin{cases}x+2\leqslant8,\\x\leqslant8,\end{cases}$

所以 $5<x\leqslant6$，$x\in\mathbf{N}^*$，所以 $x=6$.

【规律总结】 (1) 排列数的第一个公式 $A_n^m=n(n-1)\cdots\cdots(n-m+1)$适用于具体计算以及解当 $m$ 较小时的含有排列数的方程和不等式；在运用该公式时要注意它的特点.

(2) 排列数的第二个公式 $A_n^m=\frac{n!}{(n-m)!}$适用于与排列数有关的证明、解方程、解不等式等，在具体运用时，应注意先提取公因式，再计算，同时还要注意隐含条件“$m\leqslant n$ 且 $n\in\mathbf{N}^*$，$m\in\mathbf{N}^*$”的运用.

类型 2　特殊元素、特殊位置问题

例2　用数字 0,1,2,3,4,5 组成没有重复数字的四位数.

(1) 可组成多少个不同的四位数？

(2) 可组成多少个不同的四位偶数？

【解答】 (1) 根据题意分步完成任务：

第一步：排千位数字，从 1,2,3,4,5 这 5 个数字中选 1 个来排，有 $A_5^1=5$ 种不同排法；

第二步：排百位、十位、个位数字，从排了千位数字后剩下的 5 个数字中选 3 个来排列，有 $A_5^3=5\times4\times3=60$ 种不同排法，所以组成不同的四位数有 $5\times60=300$ 种.

(2) 根据题意分类完成任务：

第一类：个位数字为 0，则从 1,2,3,4,5 这 5 个数字中选 3 个来排在千位、百位、十位，有 $A_5^3=5\times4\times3=60$ 种不同排法；

第二类：个位数字为 2 或 4，则 0 不能排在千位，有 $A_2^1A_4^1A_4^2=2\times4\times4\times3=96$ 种不同排法.

所以组成的不同四位偶数有 $60+96=156$ 种.

【规律总结】对于有限制条件的排列问题，先考虑安排好特殊元素(或位置)，再安排一般的元素(或位置)，即先特殊后一般，此方法是直接分步法；或按特殊元素当选情况(或特殊位置放哪个元素)分类，再安排一般的元素(或位置)，即先分类后分步，此方法是直接分类法；也可以先不考虑特殊元素(或位置)，而列出所

有元素的全排列数，再从中减去不满足特殊元素(或位置)要求的排列数，即先全体后排除，此方法是间接法(排除法)．

类型 3　“相邻”与“不相邻”问题

**例3**　3 名男生、4 名女生按照不同的要求排队，求不同的排队方法的种数．

(1) 全体站成一排，男、女各站在一起；

(2) 全体站成一排，男生必须站在一起；

(3) 全体站成一排，男生不能站在一起；

(4) 全体站成一排，男、女各不相邻．

**【解答】** (1) 男生必须站在一起是男生的全排列，有 $A_3^3$种排法；女生必须站在一起是女生的全排列，有 $A_4^4$种排法；全体男生、女生各视为一个元素，有 $A_2^2$种排法．由分步计数原理知，共有 $A_3^3\cdot A_4^4\cdot A_2^2=288$ 种排队方法．

(2) 三个男生全排列有 $A_3^3$种排法，把所有男生视为一个元素，与 4 名女生组成 5 个元素全排列，有 $A_5^5$种排法．故有 $A_3^3\cdot A_5^5=720$ 种排队方法．

(3) 先安排女生，共有 $A_4^4$种排法；男生在 4 个女生隔成的五个空中安排，共有 $A_5^3$种排法，故共有 $A_4^4\cdot A_5^3=1\ 440$ 种排法．

(4) 排好男生后让女生插空，共有 $A_3^3\cdot A_4^4=144$ 种排法．

**变式**　某节目组决定把《将进酒》《山居秋暝》《望岳》《送杜少府之任蜀州》和另外确定的两首诗词排在后六场做节目开场诗词，并要求《将进酒》与《望岳》相邻，且《将进酒》排在《望岳》的前面，《山居秋暝》与《送杜少府之任蜀州》不相邻，且均不排在最后，则后六场开场诗词的排法有(　C　)

A．72 种　　　　B．48 种

C．36 种　　　　D．24 种

**【解析】** 首先可将《将进酒》与《望岳》捆绑在一起和另外确定的两首诗词进行全排列，共有 $A_3^3=6$ 种排法，再将《山居秋暝》与《送杜少府之任蜀州》插排在 3 个空里(最后一个空不排)，共有 $A_3^2=6$ 种排法，则后六场开场诗词的排法

有 $6\times6=36$ 种.

【规律总结】元素相邻问题,一般用“捆绑法”,先把相邻的若干个元素“捆绑”为一个大元素与其余元素全排列，然后再松绑，将这若干个元素内部全排列．元素不相邻问题，一般用“插空法”，先将不相邻元素以外的“普通”元素全排列，然后在普通元素之间及两端插入不相邻元素．

课堂评价▸基础达标

1．$89\times90\times91\times\cdots\times100$ 可表示为(　C　)

A．$A_{100}^{10}$　　B．$A_{100}^{11}$

C．$A_{100}^{12}$　　D．$A_{100}^{13}$

【解析】 $A_{100}^{12}=100\times99\times\cdots\times(100-12+1)=100\times99\times\cdots\times89$.

2．已知 $A_{2x}^{3}=100A_{x}^{2}$，则 $x$ 等于(　C　)

A．11　　B．12

C．13　　D．14

【解析】由 $A_{2x}^{3}=100A_{x}^{2}$,得 $2x\cdot(2x-1)\cdot(2x-2)=100x\cdot(x-1)$,则 $2x\cdot(2x-1)\cdot2(x-1)=100x\cdot(x-1)$，即 $2x-1=25$，解得 $x=13$，经检验满足题意．

3．(多选)满足不等式 $\frac{A_n^7}{A_n^5}>12$ 的 $n$ 的值可能为(　ABC　)

A．12　　B．11

C．10　　D．8

【解析】 由排列数公式得 $\frac{n!\ \cdot(n-5)!}{(n-7)!\ \cdot n!}>12$，则 $(n-5)\cdot(n-6)>12$，解得 $n>9$ 或 $n<2$(舍去)．因为 $n\in\mathbf{N}^*$，所以 $n$ 可以取 10,11,12.

4．某天上午要排语文、数学、体育、计算机四节课，其中体育不排在第一节，那么这天上午课程表的不同排法共有(　C　)

A．6 种　　B．9 种

C．18 种　　D．24 种

【解析】 先排体育有 $A_3^1$ 种，再排其他的三科有 $A_3^3$ 种，共有 $3\times6=18$ 种．

5．5 人随机排成一排且甲、乙不相邻的站法有( B )

A．36 种　　B．72 种

C．120 种　　D．144 种

【解析】 先将 5 人中除甲、乙之外的 3 人排成一排，然后将甲、乙插入，故共有 $A_3^3A_4^2=6\times12=72$ 种排列方法．

# 第3课时　组　合

## 预习评价▶新知初探

### 学习目标

1．了解组合的概念．

1．了解组合数的意义．

2．会用枚举法或计数原理计算组合数．

### 问题导引

1．组合的概念是什么？两个组合是相同组合的条件是什么？

2．排列数和组合数有什么区别？

### 现学现用

1．概念辨析(正确的打“√”，错误的打“×”)

(1) 从$a,b,c,d$中选取2个合成一组,其中$a,b$与$b,a$是同一个组合.(　√　)

(2) 组合要求$n$个元素是不同的，取出的$m$个元素也是不同的，即从$n$个不同的元素中进行$m$次不放回地取出．(　√　)

(3) 组合和排列一样，都与顺序有关．(　×　)

2．下列各事件中，属于组合问题的是(　C　)

A．从3名教师中，选出2名分别去北京、上海学习

B．从10名司机中选出4名，分配到4辆汽车上

C．某同学从4门课程中选修2门

D．从13位同学中任选出两位担任学习委员、体育委员

3．甲、乙、丙三地之间有直达的火车，相互之间的距离均不相等，则车票票价的种数是(　C　)

A．1　　　　B．2

C．3　　　　D．6

4．从2,3,4,5四个数中任取2个数作为对数式$\log_a b$的底数与真数，得到的对数的个数有多少，是<u>　排列　</u>问题；若求两个数相乘得到的积有几种，则是<u>　组合　</u>问题．(填“排列”或“组合”)

## 对点练▶先掌握基础

### 类型1　组合概念的理解

例1　在下列问题中，判断哪些是组合问题，哪些是排列问题.

(1) 从 $a$，$b$，$c$，$d$ 四名学生中选出 2 名学生，有多少种不同的选法？

(2) 从 $a$，$b$，$c$，$d$ 四名学生中选出 2 名学生完成两件不同的工作，有多少种不同的选法？

(3) $a$，$b$，$c$，$d$ 四支足球队之间进行单循环比赛，共需赛多少场？

(4) $a$，$b$，$c$，$d$ 四支足球队争夺冠亚军，有多少种不同的结果？

【解答】 (1) 没有顺序，是组合问题.

(2) 2 名学生完成两件不同的工作，有顺序，是排列问题.

(3) 单循环比赛要求每两支球队之间只打一场比赛，没有顺序，是组合问题.

(4) 争夺冠亚军是有顺序的，是排列问题.

【规律总结】 区分某一问题是组合问题还是排列问题，关键是看取出的元素是否有顺序，有顺序就是排列问题，没有顺序就是组合问题.

类型 2　枚举组合数

例2　平面内有 $A$，$B$，$C$，$D$ 四个不同的点，其中任意 3 个点不共线.

(1) 试写出以其中任意两个点为端点的有向线段.

(2) 试写出以其中任意两个点为端点的线段.

(3) 试写出以其中任意三点为顶点的三角形.

【解答】 (1) 以其中任意两个点为端点的有向线段为一个排列问题，共有有向线段：$AB$，$AC$，$AD$，$BA$，$BC$，$BD$，$CA$，$CB$，$CD$，$DA$，$DB$，$DC$.

(2) 以其中任意两个点为端点的线段为一个组合问题，共有线段：$AB$，$AC$，$AD$，$BC$，$BD$，$CD$.

(3) 以其中任意三点为顶点的三角形是一个组合问题，共有$\triangle ABC$，$\triangle ABD$，$\triangle BCD$，$\triangle ACD$.

【规律总结】 要想列出所有组合，做到不重不漏，应先将元素按照一定的顺序排好，然后按顺序用枚举的方法将各个组合逐个地标出来.

课堂评价▶基础达标

1．下列问题不是组合问题的是(　D　)

A．10 个朋友聚会，每两人握手一次，一共握手多少次？

B．平面上有 2 015 个不同的点，它们中任意三点不共线，连接任意两点可以构成多少条线段？

C．集合$\{a_1,a_2,a_3,\cdots,a_n\}$的含有三个元素的子集有多少个？

D．从高三(19)班的 54 名学生中选出 2 名学生分别参加校庆晚会的独唱、独舞节目，有多少种选法？

**【解析】** 组合问题与次序无关，排列问题与次序有关．D 项中，选出的 2 名学生，如甲、乙，其中“甲参加独唱、乙参加独舞”与“乙参加独唱、甲参加独舞”是两个不同的选法，因此是排列问题，不是组合问题．

2．7 个朋友聚会，每两人握手 1 次，共握手次数为(　A　)

A．$C_7^2$　　　　B．$A_7^2$

C．$C_7^1C_6^1$　　　　D．$A_7^1A_6^1$

3．已知平面内 $A$，$B$，$C$，$D$ 这 4 个点中任何 3 点均不共线，则以其中任意 3 个点为顶点的所有三角形的个数为(　B　)

A．3　　　　B．4

C．12　　　　D．24

4．现有 10 名教师，其中男教师 6 名，女教师 4 名．选出 2 名男教师或 2 名女教师参加会议，则不同的选派方法种数是(　D　)

A．$C_{10}^4$　　　　B．$A_{10}^4$

C．$A_6^2+A_4^2$　　　　D．$C_6^2+C_4^2$

**【解析】** 可把问题分两类情况：第 1 类，选出的 2 名是男教师，有 $C_6^2$种方法；第 2 类，选出的 2 名是女教师，有 $C_4^2$种方法．根据分类加法计数原理，共有 $C_6^2+C_4^2$种．

5．从 5 个不同元素 $a$，$b$，$c$，$d$，$e$ 中取出 2 个，所有不同的组合有 $\underline{ab,ac,ad,ae,bc,bd,be,cd,ce,de}$.

## 第 4 课时　组合数

### 预习评价▶新知初探

**学习目标**

1．理解组合数公式，并会解决实际问题．

2．了解组合数的性质．

3．掌握几种有限制条件的组合问题的解法．

**问题导引**

1．组合数公式是如何推导的？．组合数有怎样的性质？

2．有关组合计数应用题的解题步骤是什么？

**现学现用**

1．概念辨析(正确的打“√”，错误的打“×”)

(1) 若 $C_{12}^{n}=C_{12}^{2n-3}$，则 $n$ 等于 3.(　×　)

(2) $C_{3}^{0}+C_{4}^{1}+C_{5}^{2}+\cdots+C_{21}^{18}=7\,315$.(　√　)

(3) 10 个朋友聚会，每两人握手 1 次，共握手 45 次．(　√　)

2．$C_{6}^{2}=$__15__，$C_{18}^{17}=$__18__.

3．从 10 名学生中挑选出 3 名学生参加数学竞赛，不同的选法有(　C　)

A．$A_{10}^{3}$种　　　　B．3!

C．$C_{10}^{3}$种　　　　D．以上均不对

4．若 6 个人分 4 张无座的足球门票，每人至多分 1 张，而且票必须分完，那么不同分法的种数是__15__.

【解析】 因为是无座的足球门票，所以可以看成相同的元素，因此可以看成组合问题，有 $C_{6}^{4}=\dfrac{6!}{(6-4)!\cdot 4!}=\dfrac{6\times 5}{2}=15$ 种.

### 对点练▶先掌握基础

**类型 1　组合数的公式**

例 1　(1) 计算：$C_{100}^{98}+C_{200}^{199}$.

(2) 计算：$C_{3}^{3}+C_{4}^{3}+C_{5}^{3}+\cdots+C_{20}^{3}$.

(3) 解方程：①$C_{16}^{x^2+3x+2}=C_{16}^{5x+5}$；

②$3C_{x-3}^{x-7}=5A_{x-4}^{2}$.

【解答】 (1) $C_{100}^{98}+C_{200}^{199}=C_{100}^{2}+C_{200}^{1}=\frac{100\times 99}{2}+200=4\ 950+200=5\ 150$.

(2) $C_{3}^{3}+C_{4}^{3}+C_{5}^{3}+\cdots+C_{20}^{3}=C_{4}^{4}+C_{4}^{3}+C_{5}^{3}+\cdots+C_{20}^{3}=C_{5}^{4}+C_{5}^{3}+\cdots+C_{20}^{3}=\cdots=C_{21}^{4}=5\ 985$.

(3) ①因为 $C_{16}^{x^2+3x+2}=C_{16}^{5x+5}$,

所以 $x^2+3x+2=5x+5$ 或 $x^2+3x+2+5x+5=16$, 即 $x^2-2x-3=0$ 或 $x^2+8x-9=0$,

所以 $x=-1$ 或 $x=3$ 或 $x=-9$ 或 $x=1$.

经检验: $x=3$ 或 $x=-9$ 不合题意舍去.

故原方程的解是 $x_1=-1$, $x_2=1$.

②由排列数和组合数公式, 原方程可化为

$$3\cdot\frac{(x-3)!}{(x-7)!\cdot 4!}=5\cdot\frac{(x-4)!}{(x-6)!},$$

则 $\frac{3(x-3)}{4!}=\frac{5}{x-6}$, 即 $(x-3)(x-6)=40$,

所以 $x^2-9x-22=0$, 解得 $x=11$ 或 $x=-2$(舍去).

所以方程的根为 $x=11$.

【规律总结】 在利用组合数公式进行计算、化简时, 要灵活运用组合数的性质, 一般地, 计算 $C_n^m$ 时, 若 $m$ 比较大, 可利用性质 1, 不计算 $C_n^m$ 而改为计算 $C_n^{n-m}$, 在计算组合数之和时, 常利用性质 2.

**变式** 已知 $\frac{C_{n-1}^{5}+C_{n-3}^{3}}{C_{n-3}^{3}}=\frac{19}{5}$, 求 $n$ 的值.

【解答】 原方程可变形为 $\frac{C_{n-1}^{5}}{C_{n-3}^{3}}+1=\frac{19}{5}$, 所以 $C_{n-1}^{5}=\frac{14}{5}\cdot C_{n-3}^{3}$, 即

$\dfrac{(n-1)(n-2)(n-3)(n-4)(n-5)}{5!}=\dfrac{14}{5}\cdot\dfrac{(n-3)(n-4)(n-5)}{3!}$，化简并整理得 $n^2-3n-54=0$，

解得 $n=9$ 或 $n=-6$(不合题意，舍去)，所以 $n=9$.

类型 2　有限制条件的组合问题

**例2**　现有男运动员 6 名，女运动员 4 名，其中男、女队长各 1 名．选派 5 人外出参加比赛，则在下列情形中各有多少种选派方法？

(1) 男运动员 3 名，女运动员 2 名；

(2) 队长中至少有 1 人参加；

(3) 既要有队长，又要有女运动员．

【解答】 (1) 分两步完成：

第一步，选 3 名男运动员，有 $C_6^3$种选法；

第二步，选 2 名女运动员，有 $C_4^2$种选法．由分步乘法计数原理可得，共有 $C_6^3\cdot C_4^2=120$ 种选法．

(2) 方法一(直接法)：“只有男队长”的选法种数为 $C_8^4$；“只有女队长”的选法种数为 $C_8^4$；“男、女队长都入选”的选法种数为 $C_8^3$，所以共有 $2C_8^4+C_8^3=196$ 种选法．

方法二(间接法)：从 10 人中任选 5 人有 $C_{10}^5$种选法，其中不选队长的方法有 $C_8^5$种，所以“至少有 1 名队长”的选法有 $C_{10}^5-C_8^5=196$ 种．

(3) 当有女队长时，其他人任意选，共有 $C_9^4$种选法；当不选女队长时，必选男队长，共有 $C_8^4$种选法，其中不含女运动员的选法有 $C_5^4$种，所以不选女队长时的选法共有$(C_8^4-C_5^4)$种．故既要有队长又要有女运动员的选法共有 $C_9^4+C_8^4-C_5^4=191$ 种．

【规律总结】 组合问题常有以下两类题型：

(1) “含有”或“不含有”某些元素的组合题型：“含”，则先将这些元素取出，再由另外元素补足；“不含”，则先将这些元素剔除，再从剩下的元素中去选取．

(2) “至少”或“至多”含有几个元素的组合题型：解这类题必须十分重视“至少”与“至多”这两个关键词的含义，谨防重复与漏解．用直接法和间接法都可以求解，通常用直接法分类复杂时，考虑逆向思维，用间接法处理．

类型 3　组合中的分组、分配问题

**例3**　已知有 6 本不同的书，求按下列条件各有多少种不同的方法．

(1) 分给甲、乙、丙三人，每人两本；

(2) 分为三份，每份两本．

【解答】 (1) 根据分步乘法计数原理可得 $C_6^2C_4^2C_2^2=90$ 种．

(2) 分给甲、乙、丙三人，每人两本有 $C_6^2C_4^2C_2^2$种方法，这个过程可以分两步完成：第一步分为三份，每份两本，设有 $x$ 种方法；第二步再将这三份分给甲、乙、丙三名同学有 $A_3^3$种方法．由分步乘法计数原理可得 $C_6^2C_4^2C_2^2=xA_3^3$，所以 $x=\frac{C_6^2C_4^2C_2^2}{A_3^3}=15$.因此分为三份，每份两本一共有 15 种不同分配方法．

【规律总结】 (1) 解决这类问题的关键是分清分组问题还是分配问题．

(2) 分组问题属于“组合”问题，常见的分组问题有三种：①完全均匀分组，每组的元素个数均相等．②部分均匀分组，应注意不要重复，有 $n$ 组均匀，最后必须除以 $n$！.③完全非均匀分组，这种分组不考虑重复现象．

(3) 分配问题属于“排列”问题，分配问题可以按要求逐个分配，也可以分组后再分配．

**变式**　自 2019 年 12 月以来，在湖北省武汉市发现多起病毒性肺炎病例，研究表明，该新型冠状病毒具有很强的传染性．某社区按上级要求做好在鄂返乡人员体格检查登记，有 3 个不同的住户属在鄂返乡住户，负责该小区体格检查的社

区诊所共有 4 名医生，现要求这 4 名医生都要分配出去，且每个住户家里都要有医生去检查登记，则不同的分配方案共有多少种？

【解答】 先将 4 名医生分成 3 组，其中 1 组有 2 人，共有 $C_4^2$种方法，然后将这 3 组医生分配到 3 个不同的住户中去，有 $A_3^3$种方法．由分步原理可知共有 $C_4^2A_3^3=36$ 种不同分配方法．

课堂评价▶基础达标

1．下列计算结果是 21 的是(　D　)

A．$A_4^2+C_6^2$　　B．$C_7^3$

C．$A_7^2$　　D．$C_7^2$

【解析】 $A_4^2+C_6^2=\frac{4!}{2!}+\frac{6!}{2!\ 4!}=12+15=27$，$C_7^3=\frac{7!}{3!\ 4!}=35$，$A_7^2=\frac{7!}{5!}=42$，$C_7^2=\frac{7!}{2!\ 5!}=21$.

2．若 $3A_n^3-6A_n^2=4C_{n+1}^2$，则 $n$ 等于(　A　)

A．5　　B．8

C．7　　D．6

【解析】 因为 $3A_n^3-6A_n^2=4C_{n+1}^2$，所以 $3n(n-1)(n-2)-6n(n-1)=4\times\frac{(n+1)n}{2}$，即 $3(n-1)(n-2)-6(n-1)=2n+2$，解得 $n=5$ 或 $n=\frac{2}{3}$(舍去)．

3．若有 4 本不同的书，平均分给甲、乙 2 人，则不同的分法种数是(　B　)

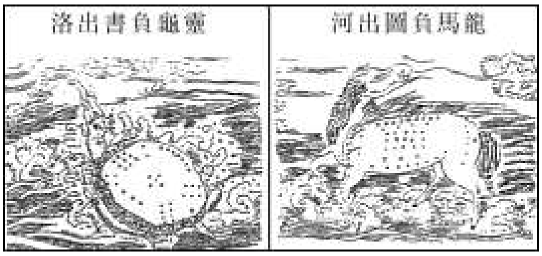
A．3　　B．6

C．12　　D．24

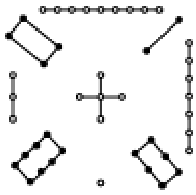
【解析】 根据题意，将 4 本不同的书平均分给甲、乙 2 人，每人分得 2 本，分 2 步进行分析：①在 4 本书中任选 2 本，分给甲，有 $C_4^2=6$ 种情况，②剩下的 2 本送给乙，有 1 种情况，则有 6 种不同的分法．

4．龙马负图、神龟载书数千年来被认为是中华传统文化的源头，其图象如图(1)所示；其中洛书有云，神龟出于洛水，甲壳上的图象如图(2)所示，其结构是戴

九履一，左三右七，二四为肩，六八为足，以五居中，五方白圈皆阳数，四角黑点为阴数．若从阳数和阴数中分别随机抽出 2 个和 1 个，则不同选法的种数是__40__．




(第 4 题(1))



洛书

(第 4 题(2))

**【解析】** 依题意，阳数为 1、3、5、7、9，阴数为 2、4、6、8，故所有的情况有 $C_5^2C_4^1=40$ 种．

5．已知 $C_n^x=C_n^{2x}$，$C_n^{x+1}=\frac{11}{3}C_n^{x-1}$，试求 $x$，$n$ 的值．

**【解答】** 由 $C_n^x=C_n^{2x}$，得 $x=2x$(舍去)或 $x+2x=n$，所以 $n=3x$，

所以 $C_n^{\frac{n}{3}+1}=\frac{11}{3}C_n^{\frac{n}{3}-1}$，

即 $\frac{n!}{\left(\frac{n}{3}+1\right)!\cdot\left(n-\frac{n}{3}-1\right)!}=\frac{11}{3}\cdot\frac{n!}{\left(\frac{n}{3}-1\right)!\cdot\left(n-\frac{n}{3}+1\right)!}$，

化简得 $11\cdot\left(\frac{n}{3}+1\right)\cdot\frac{n}{3}=3\cdot\left(\frac{2n}{3}+1\right)\cdot\frac{2n}{3}$，

解得 $n=15$，所以 $x=5$．

## 习题课　排列与组合的综合应用

### 题型 1　排列数公式和组合数公式

例1　(1) 解不等式 $A_n^2>C_n^3$；

(2) 解方程 $A_{2x+1}^4=140A_x^3$.

【解答】 (1) 由 $A_n^2>C_n^3$，得 $n(n-1)>\frac{n(n-1)(n-2)}{3\times2\times1}$，即 $n<8$.又 $n\geqslant3$，且 $n\in$ $\mathbf{N}^*$，所以 $n=3,4,5,6,7$，故不等式的解集为$\{3,4,5,6,7\}$.

(2) 因为$\begin{cases}2x+1\geqslant4,\\x\geqslant3,\end{cases}$ 所以 $x\geqslant3$，$x\in\mathbf{N}^*$.

由 $A_{2x+1}^4=140A_x^3$，得

$(2x+1)2x(2x-1)(2x-2)=140x(x-1)(x-2)$，

化简得 $4x^2-35x+69=0$，

解得 $x_1=3$，$x_2=\frac{23}{4}$(舍去)，所以方程的解为 $x=3$.

变式　解不等式 $A_8^x<6A_8^{x-2}$.

【解答】 由 $A_8^x<6A_8^{x-2}$，得$\frac{8!}{(8-x)!}<6\times\frac{8!}{(10-x)!}$，

化简得 $x^2-19x+84<0$，解得 $7<x<12$.①

又$\begin{cases}8\geqslant x,\\x-2>0,\end{cases}$ 所以 $2<x\leqslant8$.②

由①②及 $x\in\mathbf{N}^*$，得 $x=8$.

### 题型 2　排列问题

例2　一场小型晚会有 3 个唱歌节目和 2 个相声节目，要求排出一个节目单.

(1) 2 个相声节目要排在一起，有多少种排法？

(2) 第 1 个节目和最后 1 个节目都是唱歌节目，有多少种排法？

(3) 前 3 个节目中要有相声节目，有多少种排法？

【解答】 (1) 把 2 个相声节目捆绑在一起，作为一个节目与其他节目排列共有排法 $A_4^4A_2^2=48$ 种.

(2) 选 2 个唱歌节目排在首尾，剩下的 3 个节目在中间排列，排法为 $A_3^2A_3^3=36$ 种.

(3) 5 个节目全排列减去后两个都是相声的排法，共有排法 $A_5^5-A_3^3A_2^2=120-12=108$ 种.

变式 8 人围圆桌开会，其中正、副组长各 1 人，记录员 1 人.

(1) 若正、副组长相邻而坐，有多少种坐法?

(2) 若记录员坐于正、副组长之间(三者相邻)，有多少种坐法?

【解答】 (1) 若正、副组长相邻而坐，则可将此 2 人看作 1 人，即 7 人围一圆桌，有 $A_6^6$种坐法.

由于正、副组长 2 人可交换，故有 $A_2^2$种坐法.

所以共有 $A_6^6A_2^2=6\times5\times4\times3\times2\times1\times2=1\ 440$ 种坐法.

(2) 若记录员坐于正、副组长之间(三者相邻)，可将 3 人看作 1 人，即 6 人围一圆桌，有 $A_5^5$种坐法.

因为正、副组长 2 人可交换，所以有 $A_2^2$种坐法.

所以共有 $A_5^5A_2^2=5\times4\times3\times2\times1\times2=240$ 种坐法.

题型 3 组合问题

例3 在一次数学竞赛中，某学校有 12 人通过了初试，学校要从中选出 5 人参加市级培训，求在下列条件下，不同的选法种数.

(1) 任意选 5 人;

(2) 甲、乙、丙三人必须参加;

(3) 甲、乙、丙三人不能参加;

(4) 甲、乙、丙三人只能有 1 人参加.

【解答】 (1) 从中任选 5 人是组合问题，不同的选法种数为 $C_{12}^5=792$.

(2) 甲、乙、丙三人必须参加，则只需要从另外 9 人中选 2 人，是组合问题，不同的选法种数为 $C_9^2=36$.

(3) 甲、乙、丙三人不能参加，则只需要从另外的 9 人中选 5 人，不同的选法种数为 $C_9^5=126$.

(4) 甲、乙、丙三人只能有 1 人参加，可分两步：

第 1 步，从甲、乙、丙中选 1 人，有 $C_3^1$种选法；

第 2 步，从另外 9 人中选 4 人，有 $C_9^4$种选法.

根据分步乘法计数原理，可得不同的选法种数为 $C_3^1C_9^4=378$.

**变式** 某市工商局对 35 种商品进行抽样检查，已知其中有 15 种假货，现从 35 种商品中选取 3 种.

(1) 其中某一种假货必须在内，不同取法有多少种？

(2) 其中某一种假货不能在内，不同取法有多少种？

(3) 恰有 2 种假货在内，不同取法有多少种？

(4) 至少有 2 种假货在内，不同取法有多少种？

(5) 至多有 2 种假货在内，不同取法有多少种？

**【解答】** (1) 从余下的 34 种商品中，选取 2 种，有 $C_{34}^2=561$ 种取法，所以某一种假货必须在内的不同取法有 561 种.

(2) 从 34 种可选商品中，选取 3 种，有 $C_{34}^3=5\,984$ 种或 $C_{35}^3-C_{34}^2=C_{34}^3=5\,984$ 种取法，

所以某一种假货不能在内的不同取法有 5 984 种.

(3) 从 20 种真货中选取 1 种，从 15 种假货中选取 2 种，有

$C_{20}^1C_{15}^2=2\,100$ 种取法，

所以恰有 2 种假货在内的不同的取法有 2 100 种.

(4) 选取 2 种假货有 $C_{20}^1C_{15}^2$种，选取 3 种假货有 $C_{15}^3$种，共有选取方式 $C_{20}^1C_{15}^2$

$+C_{15}^{3}=2\ 100+455=2\ 555$ 种，

所以至少有 2 种假货在内的不同的取法有 2 555 种．

(5) 方法一：(间接法)选取 3 种商品的总数为 $C_{35}^{3}$，因此共有选取方式 $C_{35}^{3}-C_{15}^{3}=6\ 545-455=6\ 090$ 种，所以至多有 2 种假货在内的不同的取法有 6 090 种．

方法二：(直接法)共有选取方式 $C_{20}^{3}+C_{20}^{2}C_{15}^{1}+C_{20}^{1}C_{15}^{2}=6\ 090$ 种，所以至多有 2 种假货在内的不同的取法有 6 090 种．

题型 4　排列组合的综合问题

例4　已知 10 件不同的产品中有 4 件是次品，现对它们进行测试，直至找出所有的次品为止．

(1) 若恰在第 5 次测试后就找出了所有次品，则这样的不同测试方法数是多少？

(2) 若恰在第 2 次测试才测试到第 1 件次品，第 7 次才找到最后一件次品，则这样的不同测试方法数是多少？

【解答】 (1) 根据题意，若恰在第 5 次测试后就找出了所有次品，即第 5 次测试的产品恰为最后一件次品，另 3 件在前 4 次中出现，则前 4 次有一件正品出现，所以共有 $A_{4}^{1}\cdot(C_{6}^{1}C_{3}^{3})A_{4}^{4}=576$ 种不同的测试方法．

(2) 根据题意，分 3 步进行分析：先排第 1 次测试，只能取正品，有 6 种不同的测试方法；再从 4 件次品中选 2 件排在第 2 次和第 7 次的位置上测试，有 $A_{4}^{2}=12$ 种测试方法；最后排余下 4 件的测试位置，有 $C_{5}^{2}A_{4}^{4}=240$ 种测试方法，所以共有 $6\times12\times240=17\ 280$ 种不同的测试方法．

【规律总结】 解决排列、组合综合问题的方法：

(1) 仔细审题，判断是组合问题还是排列问题，要按元素的性质分类，按事件发生的过程进行分步．

(2) 以元素为主时，先满足特殊元素的要求，再考虑其他元素；以位置为主

时，先满足特殊位置的要求，再考虑其他位置.

(3) 对于有附加条件的比较复杂的排列、组合问题，要周密分析，设计出合理的方案，一般先把复杂问题分解成若干个简单的基本问题，然后应用分类加法计数原理或分步乘法计数原理来解决，一般遵循先选后排的原则.

# 6．3　二项式定理

## 第 1 课时　二项式定理

预习评价▶新知初探

**学习目标**

1．了解二项式定理的推导过程.

2．理解二项式定理并会应用.

3．掌握二项展开式的项数、系数、二项式系数、通项的特征及运用.

**问题导引**

1．二项式定理是什么？通项公式是什么？

2．二项式定理有何结构特征？

**现学现用**

1．概念辨析(正确的打“√”，错误的打“×”)

(1) $(a+b)^n$ 的展开式中共有 $n$ 项.（　×　）

(2) 在二项展开式的公式中，交换 $a$，$b$ 的顺序对各项没有影响.（　×　）

(3) $C_n^k a^{n-k} b^k$ 是 $(a+b)^n$ 展开式中的第 $k$ 项.（　×　）

2．$(2a+b)^5$ 的展开式的第 3 项是(　B　)

A．$2^3C_5^2$　　　　B．$2^3C_5^2a^3b^2$

C．$2^3C_5^3$　　　　D．$2^3C_5^3a^2b^3$

3．$(x+2)^6$ 的展开式中 $x^3$ 的系数为__<u>160</u>__.

4．$(a+2b)^4$ 的展开式为 <u>$a^4+8a^3b+24a^2b^2+32ab^3+16b^4$</u>.

**对点练▶先掌握基础**

### 类型 1　求二项展开式

**例1**　求 $\left(2\sqrt{x}+\frac{1}{\sqrt{x}}\right)^4$ 的展开式.

**【解答】**　方法一：$\left(2\sqrt{x}+\frac{1}{\sqrt{x}}\right)^4=C_4^0(2\sqrt{x})^4\cdot\left(\frac{1}{\sqrt{x}}\right)^0+C_4^1\cdot(2\sqrt{x})^3\cdot\left(\frac{1}{\sqrt{x}}\right)+C_4^2\cdot(2\sqrt{x})^2\cdot\left(\frac{1}{\sqrt{x}}\right)^2+C_4^3\cdot(2\sqrt{x})\cdot\left(\frac{1}{\sqrt{x}}\right)^3+C_4^4\cdot(2\sqrt{x})^0\cdot\left(\frac{1}{\sqrt{x}}\right)^4=16x^2+32x+24+\frac{8}{x}+\frac{1}{x^2}$.

方法二：$\left(2\sqrt{x}+\frac{1}{\sqrt{x}}\right)^4=\left(\frac{2x+1}{\sqrt{x}}\right)^4=\frac{1}{x^2}(2x+1)^4=\frac{1}{x^2}[C_4^0(2x)^4\cdot1^0+C_4^1\cdot(2x)^3\cdot1+$

$C_4^2\cdot(2x)^2\cdot1^2+C_4^3\cdot(2x)\cdot1^3+C_4^4\cdot(2x)^0\cdot1^4]=\frac{1}{x^2}(16x^4+32x^3+24x^2+8x+1)=16x^2+32x+24+\frac{8}{x}+\frac{1}{x^2}$.

【规律总结】 求形式简单的二项展开式时可直接由二项式定理展开，展开时注意二项展开式的特点：前一个字母是降幂，后一个字母是升幂．有时候在展开二项式之前，根据二项式的结构特征进行适当的变形，可使展开多项式的过程得到简化．

类型 2　求二项展开式中的特定项

**例2**　已知在$\left(\frac{1}{2}x^2-\frac{1}{\sqrt{x}}\right)^n$的展开式中，第 9 项为常数项．求：

(1) $n$ 的值；

(2) 展开式中 $x^5$ 的系数；

(3) 含 $x$ 的整数次幂的项的个数．

【解答】 二项展开式的通项

$T_{k+1}=C_n^k\left(\frac{1}{2}x^2\right)^{n-k}\cdot\left(-\frac{1}{\sqrt{x}}\right)^k=(-1)^k\left(\frac{1}{2}\right)^{n-k}C_n^kx^{2n-\frac{5}{2}k}$.

(1) 因为第 9 项为常数项，即当 $k=8$ 时，$2n-\frac{5}{2}k=0$，解得 $n=10$.

(2) 令 $2n-\frac{5}{2}k=5$，得 $k=\frac{2}{5}(2n-5)=6$，

所以 $x^5$ 的系数为$(-1)^6\left(\frac{1}{2}\right)^4C_{10}^6=\frac{105}{8}$.

(3) 要使 $2n-\frac{5}{2}k$，即$\frac{40-5k}{2}$为整数，只需 $k$ 为偶数，由于 $k=0,1,2,3$，…，$9,10$，故符合要求的有 6 项，分别为展开式的第 1,3,5,7,9,11 项．

【规律总结】 通项公式的主要作用是求展开式中的特定项，常见的题型有：①求第 $k$ 项；②求含 $x^r$(或 $x^py^q$)的项；③求常数项；④求有理项．其中求有理项时一般根据通项公式所得到的项，其所有的未知数的指数恰好都是整数的项．解这

类问题必须合并通项公式中同一字母的指数，根据具体要求，令其属于整数，再根据整数的整除性来求解．另外，若通项中含有根式，一般把根式化为分数指数幂，以减少计算中的错误．

类型 3　二项式定理的逆用

**例3**　设 $n\in\mathbf{N}^*$，化简 $C_n^1+C_n^2\cdot 6+C_n^3\cdot 6^2+\cdots+C_n^n\cdot 6^{n-1}=\underline{\frac{1}{6}(7^n-1)}$.

【解析】　$C_n^1+C_n^2\cdot 6+\cdots+C_n^n\cdot 6^{n-1}=\frac{1}{6}(C_n^1\cdot 6+C_n^2\cdot 6^2+\cdots+C_n^n\cdot 6^n)=\frac{1}{6}(C_n^0+C_n^1\cdot 6+C_n^2\cdot 6^2+\cdots+C_n^n\cdot 6^n-1)=\frac{1}{6}[(1+6)^n-1]=\frac{1}{6}(7^n-1)$.

【规律总结】　逆用二项式定理可将多项式化简，对于这类问题的求解，要熟悉公式的特点、项数、各项幂指数的规律以及各项的系数．

**变式**　化简：$1-2C_n^1+4C_n^2-8C_n^3+\cdots+(-2)^n\cdot C_n^n=\underline{(-1)^n}$.

【解析】　原式$=C_n^0+C_n^1(-2)^1+C_n^2(-2)^2+C_n^3(-2)^3+\cdots+C_n^n(-2)^n=(1-2)^n=(-1)^n$.

课堂评价▶基础达标

1．若$(x+2)^n$的展开式共有 11 项，则 $n$ 等于(　B　)

A．9　　B．10

C．11　　D．8

2．二项式$(1-x)^{2\,020}$展开式中的第 2 020 项是(　C　)

A．1　　B．$2\,020x^{2\,019}$

C．$-2\,020x^{2\,019}$　　D．$x^{2\,020}$

【解析】　二项式$(1-x)^{2\,020}$的展开式的通项为 $T_{r+1}=C_{2\,020}^r(-x)^r$，所以展开式中的第 2 020 项为 $T_{2\,020}=C_{2\,020}^{2\,019}(-x)^{2\,019}=-2\,020x^{2\,019}$.

3．在$(x-\sqrt{2})^4$的展开式中 $x^2$ 的系数为(　B　)

A．6　　B．12

C．24　　D．48

【解析】　$(x-\sqrt{2})^4$的展开式的通项为 $C_4^r x^{4-r}(-\sqrt{2})^r$，由 $4-r=2$，解得 $r=2$，

则 $x^2$ 的系数为 $C_4^2(-\sqrt{2})^2=6\times2=12$.

4．$1.95^7$ 的计算结果精确到个位的近似值为(　B　)

A．106　　　　B．107

C．108

**【解析】** 因为 $1.95^7=(2-0.05)^7=2^7-C_7^1\times2^6\times0.05+C_7^2\times2^5\times0.05^2-\cdots-0.05^7\approx107.2$，所以 $1.95^7\approx107$.

5．若 $\left(x-\frac{a}{x}\right)^9$ 的展开式中 $x^3$ 的系数是－84，则 $a=$__1__.

**【解析】** $\left(x-\frac{a}{x}\right)^9$ 的展开式的通项为 $T_{r+1}=C_9^rx^{9-r}\cdot\left(-\frac{a}{x}\right)^r=C_9^rx^{9-2r}(-a)^r$，令 $9-2r=3$，得 $r=3$，则 $\left(x-\frac{a}{x}\right)^9$ 的展开式中 $x^3$ 的系数为 $C_9^3(-a)^3=-84$，得 $a=1$.

## 第 2 课时　二项式系数的性质

### 预习评价▶新知初探

**学习目标**

1．理解二项式系数的性质并会应用．

2．掌握应用“赋值法”．

**问题导引**

1．什么是“杨辉三角”？它具有哪些特点？

2．二项式系数的性质有哪些？

**现学现用**

1．概念辨析(正确的打“√”，错误的打“×”)

(1) 杨辉三角的每一斜行数字的差成一个等差数列．(　×　)

(2) 二项展开式中系数最大项与二项式系数最大项是相同的．(　×　)

(3) 二项展开式的二项式系数和为 $C_n^1+C_n^2+\cdots+C_n^n$．(　×　)

2．$\left(x-\frac{1}{x}\right)^{11}$ 的展开式中二项式系数最大的项是(　D　)

A．第 6 项　　　　B．第 8 项

C．第 5,6 项　　　　D．第 6,7 项

3．已知$(ax+1)^n$ 的展开式中，二项式系数和为 32，则 $n$ 等于(　A　)

A．5　　　　B．6

C．7　　　　D．8

4．已知$(a+b)^n$ 的二项展开式中只有第 5 项的二项式系数最大，则 $n$ 等于(　D　)

A．11　　　　B．10

C．9　　　　D．8

### 对点练▶先掌握基础

**类型 1　求二项式系数或系数最大的项**

**例 1**　在$\left(\sqrt{x}-\frac{2}{x^2}\right)^8$ 的展开式中：

(1) 系数的绝对值最大的项是第几项？

(2) 求二项式系数最大的项；

(3) 求系数最大的项.

【解答】 $T_{r+1}=C_8^r\cdot(\sqrt{x})^{8-r}\cdot\left(\frac{-2}{x^2}\right)^r=(-1)^r\cdot C_8^r\cdot 2^r\cdot x^{4-\frac{5}{2}r}(r=0,1,2,\cdots,8)$.

(1) 设第 $r+1$ 项系数的绝对值最大，则$\begin{cases}C_8^r\cdot 2^r\geqslant C_8^{r+1}\cdot 2^{r+1},\\ C_8^r\cdot 2^r\geqslant C_8^{r-1}\cdot 2^{r-1},\end{cases}$

所以$\begin{cases}\frac{1}{8-r}\geqslant\frac{2}{r+1},\\ \frac{2}{r}\geqslant\frac{1}{9-r},\end{cases}$ 解得 $5\leqslant r\leqslant 6$.

又因为 $0\leqslant r\leqslant 8$，$r\in\mathbf{N}$，所以 $r=5$ 或 $r=6$，

故系数的绝对值最大的项是第 6 项和第 7 项.

(2) 二项式系数最大的项为中间项，即第 5 项，

$T_5=C_8^4\cdot 2^4\cdot x^{4-\frac{20}{2}}=1\ 120x^{-6}$.

(3) 由(1)知展开式中第 6 项和第 7 项的系数的绝对值最大，而第 6 项的系数为负，第 7 项的系数为正，所以系数最大的项为 $T_7=C_8^6\cdot 2^6\cdot x^{-11}=1\ 792x^{-11}$.

**变式** 若$(1+2x)^n$ 的展开式中第 6 项与第 7 项的系数相等，求展开式中二项式系数最大的项和系数最大的项.

【解答】 $T_6=C_n^5(2x)^5$，$T_7=C_n^6(2x)^6$，依题意有 $C_n^5 2^5=C_n^6 2^6$，解得 $n=8$.故$(1+2x)^8$ 的展开式中，二项式系数最大的项为 $T_5=C_8^4\cdot(2x)^4=1\ 120x^4$.

设第 $k+1$ 项系数最大，则有

$\begin{cases}C_8^k\cdot 2^k\geqslant C_8^{k-1}\cdot 2^{k-1},\\ C_8^k\cdot 2^k\geqslant C_8^{k+1}\cdot 2^{k+1}\end{cases}$ $\Rightarrow 5\leqslant k\leqslant 6$，又 $k\in\{0,1,2,\cdots,8\}$，故 $k=5$ 或 $k=6$.

故系数最大的项为 $T_6=1\ 792x^5$，$T_7=1\ 792x^6$.

【规律总结】 (1) 注意“系数最大”“二项式系数最大”及“系数绝对值最大”的区别.

(2) 要求展开式的系数的最大值，在系数均为正数的前提下，只需比较相邻两个系数的大小，即设第 $r+1$ 项的系数最大，则 $\begin{cases} T_{r+1}\text{的系数} \geqslant T_r\text{的系数}, \\ T_{r+1}\text{的系数} \geqslant T_{r+2}\text{的系数}. \end{cases}$

类型 2　整除或余数问题

例2　(1) 求证：$1+2+2^2+\cdots+2^{5n-1}$ 能被 31 整除($n\in\mathbf{N}^*$)；

(2) 求 $S=C_{27}^1+C_{27}^2+\cdots+C_{27}^{27}$ 除以 9 的余数.

【解答】 (1) $1+2+2^2+\cdots+2^{5n-1}=\dfrac{2^{5n}-1}{2-1}=2^{5n}-1=32^n-1=(31+1)^n-1=C_n^0\times31^n+C_n^1 31^{n-1}+\cdots+C_n^{n-1}\times31+C_n^n-1=31(C_n^0\times31^{n-1}+C_n^1\times31^{n-2}+\cdots+C_n^{n-1})$，

显然上式括号内的数为整数，所以原式能被 31 整除.

(2) $S=C_{27}^1+C_{27}^2+\cdots+C_{27}^{27}=2^{27}-1=8^9-1=(9-1)^9-1=C_9^0\times9^9-C_9^1\times9^8+\cdots+C_9^8\times9-C_9^9-1=9(C_9^0\times9^8-C_9^1\times9^7+\cdots+C_9^8)-2=9(C_9^0\times9^8-C_9^1\times9^7+\cdots+C_9^8-1)+7$，

显然上式括号内的数是正整数，故 $S$ 除以 9 的余数是 7.

【规律总结】 在利用二项式定理证明整除问题或求余数问题时，要进行合理的变形，常用的变形方法是拆数，往往是将幂底数写成两数和或差的形式，其中的一个数是除数或其正整数倍.

类型 3　“赋值法”的应用

例3　已知 $(1-2x)^7=a_0+a_1x+a_2x^2+\cdots+a_7x^7$，求下列各式的值.

(1) $a_1+a_2+\cdots+a_7$；

(2) $a_1+a_3+a_5+a_7$；

(3) $a_0+a_2+a_4+a_6$.

【解答】 令 $x=1$，

则 $a_0+a_1+a_2+a_3+\cdots+a_7=-1$.①

令 $x=-1$，则 $a_0-a_1+a_2-\cdots-a_7=3^7$.②

(1) 令 $x=0$，得 $a_0=1$，代入①中得 $a_1+a_2+a_3+\cdots+a_7=-2$.

(2) 由①－②得 $2a_1+2a_3+2a_5+2a_7=-1-3^7$,所以 $a_1+a_3+a_5+a_7=\dfrac{-1-3^7}{2}$

$=-1\ 094$.

(3) 由①＋②得 $2a_0+2a_2+2a_4+2a_6=-1+3^7$,所以 $a_0+a_2+a_4+a_6=\dfrac{-1+3^7}{2}$

$=1\ 093$.

**变式** 本例条件下，求$|a_0|+|a_1|+|a_2|+\cdots+|a_7|$.

【解答】 方法一：因为$(1-2x)^7$的展开式中，$a_0$，$a_2$，$a_4$，$a_6$大于零，而 $a_1$，$a_3$，$a_5$，$a_7$ 小于零，所以$|a_0|+|a_1|+|a_2|+\cdots+|a_7|=(a_0+a_2+a_4+a_6)-(a_1+a_3+a_5+a_7)=1\ 093-(-1\ 094)=2\ 187$.

方法二：$|a_0|+|a_1|+|a_2|+\cdots+|a_7|$是$(1+2x)^7$展开式中各项的系数和，令 $x=1$，所以$|a_0|+|a_1|+\cdots+|a_7|=3^7=2\ 187$.

【规律总结】 “赋值法”是解决二项式系数问题的常用的方法，根据题目要求，灵活赋给字母所取的不同值．一般地，要使展开式中项的关系变为系数的关系，令 $x=0$ 可得常数项，令 $x=1$ 可得所有项系数之和，令 $x=-1$ 可得奇次项系数之和与偶次项系数之和的差，而当二项展开式中含负数时，令 $x=-1$ 则可得各项系数绝对值之和．

课堂评价▸基础达标

1．在$(1-x)^{201}$的展开式中，系数的最大值是(　B　)

A．$C_{201}^{99}$　　B．$C_{201}^{100}$

C．$C_{201}^{101}$　　D．$C_{201}^{102}$

【解析】 二项式系数最大的项是第 101 项和第 102 项，因为 $T_{101}=C_{201}^{100}x^{100}$，$T_{102}=-C_{201}^{101}x^{101}=-C_{201}^{100}x^{101}$，且第 102 项的系数为负，所以第 101 项的系数最大，

为 $C_{201}^{100}$.

2．已知$\left(\sqrt{x}+\dfrac{3}{\sqrt[3]{x}}\right)^n$的展开式中，各项系数的和与其各项二项式系数的和之比为 64，则 $n$ 等于(　C　)

A．4　　　　B．5

C．6　　　　D．7

【解析】 令 $x=1$，得各项系数的和为 $4^n$，二项式系数的和为 $2^n$，故有$\dfrac{4^n}{2^n}=64$，所以 $n=6$.

3．在$(x-2)^6$ 的展开式中，二项式系数的最大值为 $a$，$x^5$ 的系数为 $b$，则$\dfrac{a}{b}$等于(　B　)

A．$\dfrac{5}{3}$　　　　B．$-\dfrac{5}{3}$

C．$\dfrac{3}{5}$　　　　D．$-\dfrac{3}{5}$

【解析】 在$(x-2)^6$的展开式中，二项式系数的最大值为 $C_6^3=20$，即 $a=20$，其展开式的通项为 $T_{k+1}=C_6^kx^{6-k}\cdot(-2)^k$.令 $6-k=5$，则 $k=1$，可得 $x^5$ 的系数 $b=C_6^1\times(-2)^1=-12$，所以$\dfrac{a}{b}=\dfrac{20}{-12}=-\dfrac{5}{3}$.

4．设 $a\in\mathbf{Z}$，且 $0\leqslant a\leqslant 13$，若 $51^{2\,012}+a$ 能被 13 整除，则 $a$ 等于(　D　)

A．0　　　　B．1

C．11　　　　D．12

【解析】 由题意，因为 $51=52-1$，所以 $51^{2\,012}=(52-1)^{2\,012}=C_{2\,012}^{0}52^{2\,012}-C_{2\,012}^{1}52^{2\,011}+\cdots-C_{2\,012}^{2\,011}52+1$.又因为 52 能被 13 整除，所以只需 $1+a$ 能被 13 整除．因为 $a\in\mathbf{Z},0\leqslant a\leqslant 13$，所以 $a=12$.

5．若$(x^2+1)(x-2)^9=a_0+a_1(x-1)+a_2(x-1)^2+a_3(x-1)^3+\cdots+a_{11}(x-1)^{11}$，则 $a_1+a_2+a_3+\cdots+a_{11}=$__2__.

【解析】 令 $x=1$，得 $a_0=-2$.令 $x=2$，得 $a_0+a_1+a_2+\cdots+a_{11}=0$，所以 $a_1$

$+ a_2 + a_3 + \cdots + a_{11} = 2.$

## 章复习　能力整合与素养提升

### 要点回顾▶连点成面

1．排列与排列数

(1) 排列数的定义：从 $n$ 个不同元素中取出 $m(m\leqslant n)$个元素的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，用符号 $A_n^m$ 表示．

(2) 排列数公式：$A_n^m=n(n-1)(n-2)\cdot\cdots\cdot(n-m+1)$，这里 $n$，$m\in\mathbf{N}^*$，且 $m\leqslant n$.

2．组合与组合数

(1) 组合数的定义：从 $n$ 个不同元素中取出 $m(m\leqslant n)$个元素的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，用符号 $C_n^m$ 表示．

(2) 组合数公式：$C_n^m=\dfrac{A_n^m}{A_m^m}=\dfrac{n(n-1)(n-2)\cdot\cdots\cdot(n-m+1)}{m!}=\dfrac{n!}{m!(n-m)!}$.这里 $n\in\mathbf{N}^*$，$m\in\mathbf{N}$，并且 $m\leqslant n$.

(3) 组合数的两个性质

①$C_n^m=$ $C_n^{n-m}$ ；

②$C_{n+1}^m=$ $C_n^m$ ＋ $C_n^{m-1}$ .

3．二项式定理

$(a+b)^n=C_n^0a^n+C_n^1a^{n-1}b+\cdots+C_n^ka^{n-k}b^k+\cdots+C_n^nb^n(n\in\mathbf{N}^*)$，$(a+b)^n$ 的二项展开式共有 $n+1$ 项，其中各项的系数 $C_n^k$ $(k\in\{0,1,2,\cdots,n\})$叫做二项式系数．

4．二项式系数的性质

(1) 对称性

在二项展开式中，与首末两端“等距离”的两个二项式系数相等，即 $C_n^0=C_n^n$，$C_n^1=C_n^{n-1}$，$C_n^2=C_n^{n-2}$，…，$C_n^k=C_n^{n-k}$，…，$C_n^n=C_n^0$.

(2) 增减性与最大值

二项式系数 $C_n^k$，当 $k<\dfrac{n+1}{2}$时，二项式系数是递增的；当 $k>\dfrac{n+1}{2}$时，二项式系数是递减的．

当 $n$ 是偶数时，中间的一项 $C_n^{\frac{n}{2}}$ 取得最大值．

当 $n$ 是奇数时，中间的两项 $C_n^{\frac{n-1}{2}}$ 和 $C_n^{\frac{n+1}{2}}$ 相等，且同时取得最大值．

(3) 各二项式系数的和

$(a+b)^n$ 的展开式的各个二项式系数的和等于 $2^n$ .二项展开式中，偶数项的

二项式系数的和等于奇数项的二项式系数的和，即 $C_n^1+C_n^3+C_n^5+\cdots=C_n^0+C_n^2+C_n^4+\cdots=\underline{2^{n-1}}$.

考法聚焦▶核心突破

考法1　排列、组合的应用

例1　(1) 如果一个四位数的各位数字互不相同，且各位数字之和等于 10，则称此四位数为“完美四位数”(如 1 036)，则由数字 0,1,2,3,4,5,6,7 构成的“完美四位数”中，奇数的个数为(　B　)

A．12　　B．44

C．58　　D．76

【解析】 尾数为 1：则前三位的数字可能为 027,036,045，共 $C_2^1\cdot A_2^2\cdot 3=12$ 种，还可能为 234，有 $A_3^3=6$ 种．尾数为 3：则前三位的数字可能为 016,025，共 $C_2^1\cdot A_2^2\cdot 2=8$ 种，还可能为 124，有 $A_3^3=6$ 种．尾数为 5：则前三位的数字可能为 014,023，共 $C_2^1\cdot A_2^2\cdot 2=8$ 种．尾数为 7：则前三位的数字可能为 012，共 $C_2^1\cdot A_2^2=4$ 种．综上所述，共有 $12+6+8+6+8+4=44$ 种．

(2) 为了宣传 2022 年北京冬奥会和冬残奥会，某学校决定派小明和小李等 5 名志愿者将两个吉祥物安装在学校的体育广场．若小明和小李必须安装同一个吉祥物，且每个吉祥物都至少由两名志愿者安装，则不同的安装方案种数为(　A　)

A．8　　B．10

C．12　　D．14

【解析】 由题意可知应将志愿者分为三人组和两人组，当三人组中包含小明和小李时，安装方案有 $C_3^1A_2^2=6$ 种；当三人组中不包含小明和小李时，安装方案有 $A_2^2=2$ 种，共计有 $6+2=8$ 种．

【类题固法】

1．5 个人排成一排，其中甲与乙不相邻，而丙与丁必须相邻，则不同的排法种数为(　C　)

A．72　　B．48

C．24　　D．60

【解析】先将丙与丁捆绑，形成一个“大元素”与戊进行排列，然后再将甲、乙插空．由分步乘法计数原理可知，不同的排法种数为 $A_2^2A_2^2A_3^2=24$.

2．有四位朋友于七夕那天乘坐高铁 G77 从武汉出发(G77 只会在长沙、广州、深圳停)，分别在每个停的站点至少下一个人，则不同的下车方案有(　B　)

A．24 种　　B．36 种

C．81 种　　D．256 种

【解析】依据题意每个停的站点至少下一个人，先按 $2+1+1$ 分成三组，有 $C_4^2$种分法，再分配到三个站点，有 $A_3^3$种分法，所以一共有 $C_4^2A_3^3=36$ 种不同的下车方案．

3．现“学习强国”平台设有“阅读文章”“视听学习”等多个栏目．在某时段时，更新了 2 篇文章和 4 个视频，一位学习者准备学习这 2 篇文章和其中 2 个视频，则这 2 篇文章学习顺序不相邻的学法有(　C　)

A．24 种　　B．36 种

C．72 种　　D．144 种

【解析】根据题意，分 2 步进行分析：①在 4 个视频中任选 2 个进行学习，有 $C_4^2=6$ 种情况；②将选出的 2 个视频与 2 篇文章依次进行学习，共有 $A_4^4=24$ 种情况，其中 2 篇文章学习顺序相邻的情况有 $A_2^2A_3^3=12$ 种情况，故 2 篇文章学习顺序不相邻的情况有 12 种，则这 2 篇文章学习顺序不相邻的学法有 $6\times12=72$ 种．

4．某会议期间举办了一场“互动沙龙”，要求从 6 位男嘉宾，2 位女嘉宾中随机选出 4 位嘉宾进行现场演讲，且女嘉宾至少要选中 1 位．如果 2 位女嘉宾同时被选中，她们的演讲顺序不能相邻，那么不同演讲顺序的种数是(　C　)

A．1 860　　B．1 320

C．1 140　　D．1 020

【解析】由题知可分为两类：第一类，2 位女嘉宾只有 1 位被选中，则还需从 6 位男嘉宾里选出 3 位，然后全排列，所以不同的演讲顺序有 $C_2^1\cdot C_6^3\cdot A_4^4$种；第

二类，2 位女嘉宾同时被选中，则还需从 6 位男嘉宾里选出 2 位，所以 2 位女嘉宾的演讲顺序不相邻的不同演讲顺序有 $C_2^2 \cdot C_6^2 \cdot A_2^2 \cdot A_3^2$种．综上，不同演讲顺序的种数是 $C_2^1 \cdot C_6^3 \cdot A_4^4 + C_2^2 \cdot C_6^2 \cdot A_2^2 \cdot A_3^2 = 1\ 140$.

考法 2　二项式定理

**例2**　$\left(\sqrt{x}+\dfrac{1}{a\sqrt[4]{x}}\right)^n$的展开式中，奇数项的二项式系数之和为 128，且前三项系数成等差数列．

(1) 求 $a$ 的值；

(2) 若 $a<3$，则展开式有多少有理项？写出所有有理项．

**【解答】** 因为奇数项的二项式系数之和为 128，所以 $2^{n-1}=128$，解得 $n=8$，所以二项式为$\left(\sqrt{x}+\dfrac{1}{a\sqrt[4]{x}}\right)^8$.

第一项：$T_1=T_{0+1}=C_8^0(\sqrt{x})^8\left(\dfrac{1}{a\sqrt[4]{x}}\right)^0=x^4$，系数为 1；

第二项：$T_2=T_{1+1}=C_8^1(\sqrt{x})^7\left(\dfrac{1}{a\sqrt[4]{x}}\right)^1=\dfrac{8}{a}x^{\frac{13}{4}}$，系数为$\dfrac{8}{a}$；

第三项：$T_3=T_{2+1}=C_8^2(\sqrt{x})^6\left(\dfrac{1}{a\sqrt[4]{x}}\right)^2=\dfrac{28}{a^2}x^{\frac{5}{2}}$，系数为$\dfrac{28}{a^2}$.

由前三项系数成等差数列，得 $2\times\dfrac{8}{a}=1+\dfrac{28}{a^2}$，

解得 $a=2$ 或 $a=14$.

(2) 若 $a<3$，由(1)得二项式为$\left(\sqrt{x}+\dfrac{1}{2\sqrt[4]{x}}\right)^8$，其通项为 $T_{r+1}=C_8^r(\sqrt{x})^{8-r}\left(\dfrac{1}{2\sqrt[4]{x}}\right)^r$

$=\dfrac{C_8^r}{2^r}x^{\frac{16-3r}{4}}$，其中 $r=0,1,\cdots,8$.

所以当 $r=0,4,8$ 时，为有理项，

当 $r=0$ 时，$T_1=C_8^0x^4=x^4$；

当 $r=4$ 时，$T_5=\frac{C_8^4}{2^4}x=\frac{35}{8}x$；

当 $r=8$ 时，$T_9=\frac{C_8^8}{2^8}x^{-2}=\frac{1}{256}x^{-2}$.

综上，有 3 项有理项，分别是 $T_1=x^4$，$T_5=\frac{35}{8}x$，$T_9=\frac{1}{256}x^{-2}$.

**【类题固法】**

1．若二项式$\left(2x^4-\frac{1}{3x^3}\right)^n$的展开式中含有非零常数项，则正整数 $n$ 的最小值为( B )

A．8　　B．7

C．6　　D．5

【解析】 由$\left(2x^4-\frac{1}{3x^3}\right)^n$，得展开式的通项为 $T_{r+1}=\left(-\frac{1}{3}\right)^r 2^{n-r}C_n^r x^{4n-7r}$.令 $4n-7r=0$，据题意此方程有解，所以 $n=\frac{7r}{4}$.当 $r=4$ 时，$n$ 的最小值为 7.

2．若 $a=2-\sqrt{2}$，则 $a^{10}-2C_{10}^1a^9+2^2C_{10}^2a^8-\cdots+2^{10}$ 等于( A )

A．32　　B．$-32$

C．1 024　　D．512

【解析】 由二项式定理，得 $a^{10}-2C_{10}^1a^9+2^2C_{10}^2a^8-\cdots+2^{10}=C_{10}^0(-2)^0a^{10}+C_{10}^1(-2)^1a^9+C_{10}^2(-2)^2a^8+\cdots+C_{10}^{10}(-2)^{10}=(a-2)^{10}=(-\sqrt{2})^{10}=2^5=32$.

3．$(1+ax)^n=a_0+a_1x+a_2x^2+\cdots+a_nx^n(n\in\mathbf{N}^*)$，当 $n=5$ 时，$a_1+a_2+a_3+a_4+a_5=242$，则当 $n=6$ 时，$a_1+3a_3+5a_5$ 的值为( B )

A．$-1\ 452$　　B．1 452

C．$-726$　　D．726

【解析】 $(1+ax)^n=a_0+a_1x+a_2x^2+\cdots+a_nx^n(n\in\mathbf{N}^*)$，令 $x=0$，则 $a_0=1$；令 $x=1$，$n=5$，则$(1+a)^5=a_0+a_1+a_2+\cdots+a_5$.又 $a_1+a_2+a_3+a_4+a_5=242$，所以$(1+a)^5=243=3^5$，所以 $a=2$，$(1+2x)^n=a_0+a_1x+a_2x^2+\cdots+a_nx^n$.当 $n=6$ 时，$(1+2x)^6=C_6^0\times(2x)^0+C_6^1\times(2x)^1+C_6^2\times(2x)^2+C_6^3\times(2x)^3+C_6^4\times(2x)^4+C_6^5\times(2x)^5+C_6^6\times(2x)^6=1+12x+60x^2+160x^3+240x^4+192x^5+64x^6$，则 $a_1=12$，$a_3=160$，$a_5=$

192，$a_1+3a_3+5a_5=1\ 452$.

4．已知$\left(1+\frac{1}{2}x\right)^m=a_0+a_1x+a_2x^2+a_3x^3+\cdots+a_mx^m$，$a_0$，$a_1$，$a_2$成等差数列．

(1) 求$\left(1+\frac{1}{2}x\right)^m$的展开式的中间项；

(2) 求$\left(1+\frac{1}{2}x\right)^m$的展开式中所有含$x$奇次幂的系数和．

**【解答】** (1) 依题意$a_0=1$，$a_1=\frac{m}{2}$，$a_2=C_m^2\left(\frac{1}{2}\right)^2$.

由$2a_1=a_0+a_2$，得$m=8$或$m=1$(舍去).

所以$\left(1+\frac{1}{2}x\right)^m$展开式的中间项是第五项，

$T_5=C_8^4\left(\frac{1}{2}x\right)^4=\frac{35}{8}x^4$.

(2) 因为$\left(1+\frac{1}{2}x\right)^m=a_0+a_1x+a_2x^2+\cdots+a_mx^m$，$m=8$，

所以$\left(1+\frac{1}{2}x\right)^8=a_0+a_1x+a_2x^2+\cdots+a_8x^8$.

令$x=1$，则$a_0+a_1+a_2+a_3+\cdots+a_8=\left(\frac{3}{2}\right)^8$；

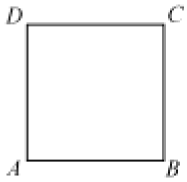
令$x=-1$，则$a_0-a_1+a_2-a_3+\cdots+a_8=\left(\frac{1}{2}\right)^8$，

所以$a_1+a_3+a_5+a_9=\frac{3^8-1}{2^9}=\frac{205}{16}$，

所以展开式中$x$的奇次幂的系数和为$\frac{205}{16}$.

考法3　排列组合、二项式定理创新运用

**例3**　某单人游戏规则如下：先将一棋子放在如图所示正方形$ABCD$(边长为2个单位长度)的顶点$A$处，然后通过掷骰子来确定棋子沿正方形的边按逆时针方向行走的单位长度，如果掷出的点数为$i(i=1,2,\cdots,6)$，则棋子就按逆时针方向行走$i$个单位长度，一直循环下去．则某人抛掷三次骰子后棋子恰好又回到点$A$处的所有不同走法共有(　D　)




(例 3)

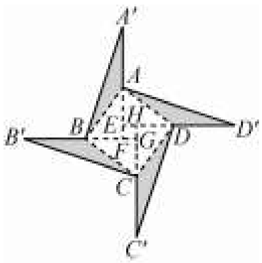
A．22 种　　B．24 种

C．25 种　　D．27 种

【解析】 抛掷三次骰子后棋子恰好又回到点 $A$ 处，表示三次骰子的点数之和是 8 或 16，列举出在点数中三个数字能够使得和为 8 或 16 的有 125；134；116；224；233；466；556，共有 7 种组合．组合 125；134，每种情况可以排列出 $A_3^3=6$ 种结果，共有 $2A_3^3=2\times6=12$ 种结果；116；224；233；466；556 各有 3 种结果，共有 $5\times3=15$ 种结果，根据分类计数原理知共有 $12+15=27$ 种结果．

【类题固法】

1．如图是在“赵爽弦图”的基础上创作出的一个“数学风车”，其中正方形 $ABCD$ 内部为“赵爽弦图”，它是由四个全等的直角三角形和一个小正方形组成的．我们将图中阴影所在的四个三角形称为“风叶”，若从该“数学风车”的八个顶点中任取两点，则两点取自同一片“风叶”的种数为(　A　)




(第 1 题)

A．12　　B．16

C．6　　D．22

【解析】 从“数学风车”的八个顶点中任取两个顶点，其中这两个顶点取自同一片“风叶”的种数有 $4C_3^2=12$ 种．

2．北京 2022 年冬奥会和冬残奥会色彩系统的主色包括霞光红、迎春黄、天雾蓝、长城灰、瑞雪白；间色包括天青、梅红、竹绿、冰蓝、吉柿；辅助色包括

墨、金、银．若各赛事纪念品的色彩设计要求：主色至少一种、至多两种，间色两种、辅助色一种，则某个纪念品的色彩搭配中包含有瑞雪白、冰蓝、银色这三种颜色的概率为(　B　)

A．$\frac{8}{225}$　　B．$\frac{2}{45}$

C．$\frac{1}{15}$　　D．$\frac{2}{15}$

【解析】 当主色只选一种时，共有 $C_5^1C_5^2C_3^1=150$ 种；当主色选两种时，共有 $C_5^2C_5^2C_3^1=300$ 种．其中，若主色只选一种时，某个纪念品的色彩搭配中包含有瑞雪白、冰蓝、银色这三种颜色的共有 $C_4^1=4$ 种；若主色选两种时，某个纪念品的色彩搭配中包含有瑞雪白、冰蓝、银色这三种颜色的共有 $C_4^1C_4^1=16$ 种．故某个纪念品的色彩搭配中包含有瑞雪白、冰蓝、银色这三种颜色的概率为$\frac{4+16}{150+300}=\frac{2}{45}$.

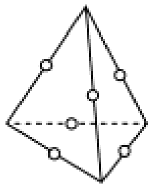
3．李老师给 6 位小朋友布置了一项搜寻空投食物的任务．已知：①食物投掷地点有远、近两处；②因为小明年纪尚小，所以要么不参与该项任务，但此时另需一位小朋友在大本营陪同，要么参与搜寻近处投掷点的食物；③所有参与搜寻任务的小朋友须被均分成两组，一组去远处，一组去近处．那么不同的搜寻方案有(　B　)

A．10 种　　B．40 种

C．70 种　　D．80 种

【解析】 若小明不参与任务，则需要从剩下的 5 位小朋友中任意挑出 1 位陪同，有 $C_5^1$种挑法，再从剩下的 4 位小朋友中挑出 2 位搜寻远处，有 $C_4^2$种挑法，最后剩下的 2 位小朋友搜寻近处，因此一共有 $C_5^1C_4^2=30$ 种搜寻方案；若小明参与任务，则其只能去近处，需要从剩下的 5 位小朋友中挑出 2 位搜寻近处，有 $C_5^2$种挑法，剩下 3 位小朋友去搜寻远处，因此共有 $C_5^2=10$ 种搜寻方案．综上，共有 $30+10=40$ 种搜寻方案．

4．四面体的顶点和各棱中点共 10 个点，在其中取 4 个不共面的点，不同的取法共有__141__种．



(第 4 题)

**【解析】** 如图，从 10 个点中任取 4 个点的组合数为 $C_{10}^{4}=210$，其中四点共面的情况可分为三类：①4 点在同一个侧面或底面的共 4 组，即 $4\times C_{6}^{4}=60$(种)；②每条棱的中点与它对棱的三点共面的有 6 种；③在 6 个中点中，四点共面的有 3 种．则 4 点不共面的取法共有 $210-(60+6+3)=141$ 种．

# 第七章　随机变量及其分布

重点与难点

1．条件概率与乘法公式

2．全概率公式

3．分布列

4．期望与方差

5．二项分布

6．超几何分布

7．正态分布

## 7．1　条件概率与全概率公式

### 第 1 课时　条件概率

预习评价▸新知初探

**学习目标**

1．结合古典概型，了解条件概率，能计算简单随机事件的条件概率．

2．掌握条件概率的性质并能解决复杂的条件概率．

**问题导引**

1．如何判断条件概率？

2．$P(B|A)$与 $P(A|B)$的区别是什么？

**现学现用**

1．概念辨析(正确的打“√”，错误的打“×”)

(1) 若事件 $A$，$B$ 互斥，则 $P(B|A)=1$.(　×　)

(2) 事件 $A$ 发生的条件下，事件 $B$ 发生，相当于 $A$，$B$ 同时发生．(　×　)

(3) $P(B|A)\neq P(AB)$．(　√　)

2．设 $A$，$B$ 为两个事件，且 $P(A)>0$，若 $P(AB)=\frac{1}{3}$，$P(A)=\frac{2}{3}$，则 $P(B|A)$等于(　A　)

A．$\frac{1}{2}$　　　　B．$\frac{2}{9}$

C. $\frac{1}{9}$　　　　　　　　　　D. $\frac{4}{9}$

3．设某动物由出生算起活到 20 岁的概率为 0.8，活到 25 岁的概率为 0.4，现有一个 20 岁的这种动物，则它活到 25 岁的概率是$\underline{\quad 0.5 \quad}$．

## 对点练▶先掌握基础

### 类型 1　定义法求条件概率

**例1**　现有 6 个节目准备参加比赛，其中 4 个舞蹈节目，2 个语言类节目，如果不放回地依次抽取 2 个节目，求：

(1) 第 1 次抽到舞蹈节目的概率；

(2) 第 1 次和第 2 次都抽到舞蹈节目的概率；

(3) 在第 1 次抽到舞蹈节目的条件下，第 2 次抽到舞蹈节目的概率．

【解答】 设第 1 次抽到舞蹈节目为事件 $A$，第 2 次抽到舞蹈节目为事件 $B$，则第 1 次和第 2 次都抽到舞蹈节目为事件 $AB$.

(1) 从 6 个节目中不放回地依次抽取 2 个的样本点总数为 $n(\Omega)=A_6^2=30$，根据分步乘法计数原理 $n(A)=A_4^1A_5^1=20$，于是 $P(A)=\frac{n(A)}{n(\Omega)}=\frac{20}{30}=\frac{2}{3}$.

(2) 因为 $n(AB)=A_4^2=12$，所以 $P(AB)=\frac{n(AB)}{n(\Omega)}=\frac{12}{30}=\frac{2}{5}$.

(3) 方法一：由(1)(2)可得，在第 1 次抽到舞蹈节目的条件下，第 2 次抽到舞蹈节目的概率为 $P(B|A)=\frac{P(AB)}{P(A)}=\frac{\frac{2}{5}}{\frac{2}{3}}=\frac{3}{5}$.方法二：因为 $n(AB)=12$，$n(A)=20$，所以 $P(B|A)=\frac{n(AB)}{n(A)}=\frac{12}{20}=\frac{3}{5}$.

【规律总结】 利用定义计算条件概率的步骤：

(1) 分别计算概率 $P(AB)$和 $P(A)$.

(2) 将它们相除得到条件概率 $P(B|A)=\frac{P(AB)}{P(A)}$，这个公式适用于一般情形，其中 $AB$ 表示 $A$，$B$ 同时发生.

类型 2　缩小样本空间范围求条件概率

例2　已知集合 $A=\{1,2,3,4,5,6\}$，甲、乙两人各从 $A$ 中任取一个数，若甲先取(不放回)，乙后取，在甲抽到奇数的条件下，求乙抽到的数比甲抽到的数大的概率．

【解答】　将甲抽到数字 $a$，乙抽到数字 $b$，记作$(a，b)$，甲抽到奇数的情形有(1,2)，(1,3)，(1,4)，(1,5)，(1,6)，(3,1)，(3,2)，(3,4)，(3,5)，(3,6)，(5,1)，(5,2)，(5,3)，(5,4)，(5,6)，共 15 个．在这 15 个中，乙抽到的数比甲抽到的数大的有(1,2)，(1,3)，(1,4)，(1,5)，(1,6)，(3,4)，(3,5)，(3,6)，(5,6)，共 9 个，所以所求概率 $P=\frac{9}{15}=\frac{3}{5}$.

【规律总结】　缩小样本空间法求概率的方法：

将原来的样本空间 $\Omega$ 缩小为已知的条件事件 $A$，原来的事件 $B$ 缩小为 $AB$.而 $A$ 中仅包含有限个样本点，每个样本点发生的概率相等，从而可以在缩小的概率空间上利用古典概型公式计算条件概率，即 $P(B|A)=\frac{n(AB)}{n(A)}$，这里 $n(A)$和 $n(AB)$的计数是基于缩小的样本点范围的．

类型 3　条件概率的性质及应用

例3　有外形相同的球分装三个盒子，每盒 10 个．其中，第一个盒子中有 7 个球标有字母 $A$,3 个球标有字母 $B$；第二个盒子中有红球和白球各 5 个；第三个盒子中有红球 8 个，白球 2 个．试验按如下规则进行：先在第一个盒子中任取一个球，若取得标有字母 $A$ 的球，则在第二个盒子中任取一个球；若第一次取得标有字母 $B$ 的球，则在第三个盒子中任取一个球．如果第二次取出的是红球，则称试验为成功，求试验成功的概率．

【解答】　设 $A=\{$从第一个盒子中取得标有字母 $A$ 的球$\}$，$B=\{$从第一个盒子中取得标有字母 $B$ 的球$\}$，$C=\{$第二次取出的球是红球$\}$，$D=\{$第二次取出的球是白球$\}$，则易得 $P(A)=\frac{7}{10}$，$P(B)=\frac{3}{10}$，$P(C|A)=\frac{1}{2}$，$P(D|A)=\frac{1}{2}$，$P(C|B)=\frac{4}{5}$，$P(D|B)$

$=\frac{1}{5}$.事件“试验成功”表示为 $CA\cup CB$，又事件 $CA$ 与事件 $CB$ 互斥，故由概率的加法公式，得 $P(CA\cup CB)=P(CA)+P(CB)=P(C|A)P(A)+P(C|B)P(B)=\frac{1}{2}\times\frac{7}{10}+\frac{4}{5}\times\frac{3}{10}=\frac{59}{100}$.

**变式** 已知男人中有 5%患色盲，女人中有 0.25%患色盲，从 100 个男人和 100 个女人中任选一人.

(1) 求此人患色盲的概率；

(2) 如果此人是色盲，求此人是男人的概率.

【解答】 设“任选一人是男人”为事件 $A$，“任选一人是女人”为事件 $B$，“任选一人是色盲”为事件 $C$.

(1) 此人患色盲的概率 $P(C)=P(AC)+P(BC)=P(A)\cdot P(C|A)+P(B)P(C|B)=\frac{100}{200}\times\frac{5}{100}+\frac{100}{200}\times\frac{0.25}{100}=\frac{21}{800}$.

(2) $P(A|C)=\frac{P(AC)}{P(C)}=\frac{\frac{5}{200}}{\frac{21}{800}}=\frac{20}{21}$.

【规律总结】 条件概率的解题策略：

(1) 应用概率加法公式的前提是事件互斥.

(2) 分解计算，代入求值：为了求比较复杂事件的概率，一般先把它分解成两个(或若干个)互不相容的较简单的事件之和，求出这些简单事件的概率，再利用加法公式即得所求的复杂事件的概率.

课堂评价▶基础达标

1．下面是条件概率的是(　B　)

A．甲、乙二人投篮命中率分别为 0.6,0.7，各投篮一次都投中的概率

B．甲、乙二人投篮命中率分别为 0.6,0.7，在甲投中的条件下乙投篮一次命中的概率

C．有 10 件产品，其中 3 件次品，抽 2 件产品进行检验，恰好抽到一件次品的概率

D．小明上学路上要过四个路口，每个路口遇到红灯的概率都是$\frac{2}{5}$，则小明在一次上学中遇到红灯的概率

2．若 $P(B|A)=\frac{1}{2}$，$P(A)=\frac{3}{5}$，则 $P(AB)$等于(　C　)

A．$\frac{5}{6}$　　　　B．$\frac{9}{10}$

C．$\frac{3}{10}$　　　　D．$\frac{1}{10}$

3．某班学生的考试成绩中，数学不及格的占 15%，语文不及格的占 5%，两门都不及格的占 3%.已知某学生数学不及格，则他语文也不及格的概率是(　A　)

A．$\frac{1}{5}$　　　　B．$\frac{3}{10}$

C．$\frac{1}{2}$　　　　D．$\frac{3}{5}$

**【解析】** 设 $A$ 为事件“数学不及格”，$B$ 为事件“语文不及格”，则 $P(B|A)=\frac{P(AB)}{P(A)}=\frac{0.03}{0.15}=\frac{1}{5}$.

4．设 $A$，$B$ 为两个事件，若事件 $A$ 和 $B$ 同时发生的概率为$\frac{3}{10}$，在事件 $A$ 发生的条件下，事件 $B$ 发生的概率为$\frac{1}{2}$，则事件 $A$ 发生的概率为__$\frac{3}{5}$__.

**【解析】** 由题意知，$P(A\cap B)=\frac{3}{10}$，$P(B|A)=\frac{1}{2}$.由 $P(B|A)=\frac{P(A\cap B)}{P(A)}$，得 $P(A)=\frac{P(A\cap B)}{P(B|A)}=\frac{3}{5}$.

5．某个班级有学生 40 人，其中有共青团员 15 人．全班分成四个小组，第一小组有学生 10 人，其中共青团员 4 人．现在要在班内任选一名共青团员当团员代表，则这个代表恰好在第一组内的概率为__$\frac{2}{5}$__.

**【解析】** 把 40 名学生看成 40 个样本点，其中第一小组所包含的样本点个数为 10，第一小组的团员所包含的样本点个数为 4.记“代表恰好在第一组”为事件

$A$，记“代表为团员代表”记为事件 $B$，所以 $n(A)=10$，$n(AB)=4$，所以 $P(B|A)$ $=\frac{n(AB)}{n(A)}=\frac{4}{10}=\frac{2}{5}$，即这个团员代表恰好在第一组内的概率为$\frac{2}{5}$.

## 第 2 课时　全概率公式

### 预习评价▶新知初探

### 学习目标

1．理解并掌握乘法公式.

2．熟练掌握全概率公式.

3．会用乘法公式、全概率公式解决一些实际问题.

### 问题导引

1. 在 $P(B|A)$，$P(BA)$, $P(A)$这三者中，如果已知 $P(A)$与 $P(B|A)$，能否求出 $P(BA)$？

2．盒中有 $a$ 个红球，$b$ 个白球，现随机地从中取出一个球，观察其颜色后，并加上同色球 $c$ 个，再从盒中第二次取一个球，你能计算出第二次取出的是黑球的概率吗？

### 现学现用

1．概念辨析(正确的打“√”，错误的打“×”)

(1) 全概率公式 $P(B)=\sum_{i=1}^{n}P(A_i)P(B|A_i)$中的事件 $B$ 只能是单一的事件．(　×　)

(2) $P(C|A)$和 $P(A|C)$是两个相同的概念．(　×　)

2. 甲、乙、丙三个车间生产同一种产品，其产量分别是总量的 25%,35%,40%，次品率分别为 5%,4%,2%.从这批产品中任取一件，则它是次品的概率为(　C　)

A．0.012 3　　　　B．0.023 4

C．0.034 5　　　　D．0.045 6

3．播种用的一等小麦种子中混有 2%的二等种子，1.5%的三等种子，1%的四等种子．一、二、三、四等种子长出的穗含 50 颗以上麦粒的概率分别为 0.5,0.15,0.1,0.05，则这批种子所结的穗含 50 颗以上麦粒的概率为(　D　)

A．0.8　　　　B．0.832 5

C．0.532 5　　　　D．0.482 5

**【解析】** 设从这批种子中任选一颗是一、二、三、四等种子的事件分别是 $A_1$，$A_2$，$A_3$，$A_4$，$B$＝“从这批种子中任选一颗，所结的穗含 50 颗以上麦粒”，则 $P(B)=\sum_{i=1}^{4}P(A_i)P(B|A_i)=95.5\%\times0.5+2\%\times0.15+1.5\%\times0.1+1\%\times0.05=0.4825$.

## 对点练▶先掌握基础

类型 1　乘法公式的应用

例1　在某大型商场促销抽奖活动中有 60 张奖券，其中有 6 张中奖奖券．假设抽完的奖券不放回，甲抽完以后乙再抽，求：

(1) 甲中奖且乙也中奖的概率；

(2) 甲没中奖且乙中奖的概率．

【解答】 方法一：设 $A$：甲中奖，$B$：乙中奖，则 $P(A)=\frac{6}{60}=\frac{1}{10}$，$P(B|A)=\frac{5}{59}$，$P(\overline{A})=\frac{9}{10}$，$P(B|\overline{A})=\frac{6}{59}$.

(1) 甲中奖而且乙也中奖的概率 $P(BA)=P(A)P(B|A)=\frac{1}{10}\times\frac{5}{59}=\frac{1}{118}$.

(2) 甲没中奖且乙中奖的概率 $P(B\overline{A})=P(\overline{A})P(B|\overline{A})=\frac{9}{10}\times\frac{6}{59}=\frac{27}{295}$.

方法二：(1) 甲中奖且乙也中奖的概率为$\frac{A_6^2}{A_{60}^2}=\frac{6\times5}{60\times59}=\frac{1}{118}$.

(2) 甲没中奖且乙中奖的概率为$\frac{C_{54}^1A_6^1}{A_{60}^2}=\frac{54\times6}{60\times59}=\frac{27}{295}$.

【规律总结】 在 $P(B|A)$，$P(BA)$，$P(A)$这三者中，如果已知 $P(A)$，$P(B|A)$，那么可以由 $P(BA)=P(A)P(B|A)$求出 $P(BA)$.

类型 2　全概率公式的运用

例2　设有两箱同一种商品：第一箱内装 50 件，其中 10 件优质品；第二箱内装 30 件，其中 18 件优质品．现在随意的打开一箱，然后从中随意取出一件，求取到是优质品的概率．

【解答】 设 $A=\{$取到的是优质品$\}$，$B_i=\{$打开的是第 $i$ 箱$\}(i=1,2)$，$P(B_1)=P(B_2)=\frac{1}{2}$，$P(A|B_1)=\frac{10}{50}=\frac{1}{5}$，$P(A|B_2)=\frac{18}{30}=\frac{3}{5}$.由全概率公式得 $P(A)=P(B_1)P(A|B_1)+P(B_2)P(A|B_2)=\frac{1}{2}\times\frac{1}{5}+\frac{1}{2}\times\frac{3}{5}=\frac{2}{5}$.

【规律总结】 全概率公式针对的是某一个过程中已知条件求出最后结果的

概率，解题步骤如下：(1) 找出条件事件里的某一个完备事件组，分别命名为 $A_i$；

(2) 命名目标的概率事件为事件 $B$；

(3) 代入全概率公式求解．

课堂评价▶基础达标

1．设 $P(A|B)=P(B|A)=\frac{1}{2}$，$P(A)=\frac{1}{3}$，则 $P(B)$等于(　B　)

A．$\frac{1}{2}$　　B．$\frac{1}{3}$

C．$\frac{1}{4}$　　D．$\frac{1}{6}$

**【解析】** 由 $P(AB)=P(A)P(B|A)=\frac{1}{3}\times\frac{1}{2}=\frac{1}{6}$，$P(A|B)=\frac{P(AB)}{P(B)}$，得 $P(B)=\frac{P(AB)}{P(A|B)}=\frac{1}{6}\times 2=\frac{1}{3}$.

2．市场上供应的灯泡中，甲厂产品占 70%，乙厂产品占 30%，甲厂产品的合格率是 95%，乙厂产品的合格率是 80%，则从市场上买到的一个甲厂的合格灯泡的概率是(　A　)

A．0.665　　B．0.564

C．0.245　　D．0.285

**【解析】** 记事件 $A$ 为“甲厂产品”，事件 $B$ 为“合格产品”，则 $P(A)=0.7$，$P(B|A)=0.95$，所以 $P(AB)=P(A)\cdot P(B|A)=0.7\times 0.95=0.665$.

3．设甲乘汽车、火车前往某目的地的概率分别为 0.6,0.4，汽车和火车正点到达目的地的概率分别为 0.9,0.8.则甲正点到达目的地的概率为(　C　)

A．0.72　　B．0.96

C．0.86　　D．0.84

**【解析】** 设事件 $A$ 表示甲正点到达目的地．事件 $B$ 表示甲乘火车到达目的地，事件 $C$ 表示甲乘汽车到达目的地，由题意知 $P(B)=0.4$，$P(C)=0.6$，$P(A|B)=0.8$，$P(A|C)=0.9$.由全概率公式得 $P(A)=P(B)P(A|B)+P(C)P(A|C)=0.4\times 0.8+0.6\times 0.9=0.32+0.54=0.86$.

4．设某医院仓库中有 10 盒同样规格的 $X$ 光片，其中有 5 盒、3 盒、2 盒依次是甲厂、乙厂、丙厂生产的，且甲、乙、丙三厂生产该种 $X$ 光片的次品率依次为$\frac{1}{10}$，$\frac{1}{15}$，$\frac{1}{20}$.现从这 10 盒中任取一盒，再从这盒中任取一张 $X$ 光片，则取得的 $X$ 光片是次品的概率为(　A　)

A．0.08　　　　B．0.1

C．0.15　　　　D．0.2

**【解析】** 设 $A_1$，$A_2$，$A_3$ 分别表示取得的这盒 $X$ 光片是由甲厂、乙厂、丙厂生产的，$B$ 表示取得的 $X$ 光片为次品，则 $P(A_1)=\frac{5}{10}$，$P(A_2)=\frac{3}{10}$，$P(A_3)=\frac{2}{10}$，$P(B|A_1)=\frac{1}{10}$，$P(B|A_2)=\frac{1}{15}$，$P(B|A_3)=\frac{1}{20}$.由全概率公式得 $P(B)=P(A_1)P(B|A_1)+P(A_2)P(B|A_2)+P(A_3)P(B|A_3)=\frac{5}{10}\times\frac{1}{10}+\frac{3}{10}\times\frac{1}{15}+\frac{2}{10}\times\frac{1}{20}=0.08$.

5．甲袋中有 2 个白球和 4 个红球，乙袋中有 1 个白球和 2 个红球，现随机地从甲袋中取出一球放入乙袋，然后从乙袋中随机地取出一球，则从乙袋中取出白球的概率是__$\frac{1}{3}$__.

**【解析】** 设 $A$ 表示事件“从甲袋中移入乙袋中的球是白球”，$B$ 表示事件“最后从乙袋中取出的是白球”，则 $P(A)=\frac{2}{6}$，$P(\overline{A})=\frac{4}{6}$，$P(B|A)=\frac{2}{4}$，$P(B|\overline{A})=\frac{1}{4}$，所以 $P(B)=P(A)P(B|A)+P(\overline{A})P(B|\overline{A})=\frac{2}{6}\times\frac{2}{4}+\frac{4}{6}\times\frac{1}{4}=\frac{1}{3}$.

# 7．2　离散型随机变量及其分布列

预习评价▶新知初探

## 学习目标

1．理解取有限值的离散型随机变量及其分布列的概念与性质；会求出某些简单的离散型随机变量的分布列．

2．理解两点分布，并能简单地运用．

## 问题导引

1．随机变量和函数有什么联系和区别？

2．离散型随机变量 $X$ 的每一个可能取值为实数，其实质代表的是什么？

3．如何判断所求离散型随机变量的分布列是否正确？

## 现学现用

1．概念辨析(正确的打“√”，错误的打“×”)

(1) 离散型随机变量的取值是任意的实数．(　×　)

(2) 随机变量的取值可以是有限个，也可以是无限个．(　√　)

(3) 设 $a$，$b\in\mathbf{R}$，对随机变量 $X$，事件 $X=a$ 与事件 $X=b$ 互斥．(　×　)

2．若随机变量 $X$ 的分布列如下表，则 $a$ 的值为(　D　)

| $X$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $P$ | $\frac{1}{2}$ | $\frac{1}{12}$ | $\frac{1}{4}$ | $a$ |

A．1　　　　B．$\frac{1}{2}$

C．$\frac{1}{3}$　　　　D．$\frac{1}{6}$

【解析】 由分布列的性质，得$\frac{1}{2}+\frac{1}{12}+\frac{1}{4}+a=1$，解得 $a=\frac{1}{6}$.

3．设随机变量 $X$ 的分布列为 $P(X=k)=m\left(\frac{2}{3}\right)^k$，$k=1,2,3$，则 $m$ 的值为__$\frac{27}{38}$__.

【解析】 $P(X=1)=\frac{2m}{3}$，$P(X=2)=\frac{4m}{9}$，$P(X=3)=\frac{8m}{27}$，由离散型随机变量的分布列的性质知 $P(X=1)+P(X=2)+P(X=3)=1$，即$\frac{2m}{3}+\frac{4m}{9}+\frac{8m}{27}=1$，解得 $m=\frac{27}{38}$.

## 对点练▶先掌握基础

类型 1　离散型随机变量的判定

例1　指出下列随机变量是否是离散型随机变量，并说明理由.

(1) 某座大桥一天经过的车辆数 $X$;

(2) 某超市 5 月份每天的销售额;

(3) 某加工厂加工的一批某种钢管的外径与规定的外径尺寸之差 $\xi$;

(4) 九江市长江水位监测站所测水位在(0,29]这一范围内变化，该水位站所测水位 $\xi$.

【解答】 (1) 车辆数 $X$ 的取值可以一一列出，故 $X$ 为离散型随机变量.

(2) 某超市 5 月份每天销售额可以一一列出，故为离散型随机变量.

(3) 实际测量值与规定值之间的差值无法一一列出，不是离散型随机变量.

(4) 水位在(0,29]这一范围内变化，不能按次序一一列举，不是离散型随机变量.

【规律总结】 判断一个随机变量 $X$ 是否为离散型随机变量的具体方法:

(1) 明确随机试验的所有可能结果;

(2) 将随机试验的试验结果数量化;

(3) 确定试验结果所对应的实数是否可按一定次序一一列出，如果能一一列出，则该随机变量是离散型随机变量，否则不是.

类型 2　离散型随机变量的性质

例2　(1) 设离散型随机变量 $X$ 的分布列如下

| $X$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $P$ | $\frac{1}{6}$ | $\frac{1}{3}$ | $\frac{1}{6}$ | $p$ |

则 $p$ 的值为(　C　)

A. $\frac{1}{2}$　　　　B. $\frac{1}{6}$

C．$\dfrac{1}{3}$　　　　　　　　　　　　　　D．$\dfrac{1}{4}$

(2) 已知随机变量$\xi$只能取三个值$x_1$，$x_2$，$x_3$，其概率依次成等差数列，则该等差数列的公差的取值范围是(　B　)

A．$\left[0，\dfrac{1}{3}\right]$　　　　　　　　　　B．$\left[-\dfrac{1}{3}，\dfrac{1}{3}\right]$

C．$[-3,3]$　　　　　　　　　　　　D．$[0,1]$

**【解析】** 设随机变量$\xi$取$x_1$，$x_2$，$x_3$的概率分别为$a-d$，$a$，$a+d$，则由分布列的性质得$(a-d)+a+(a+d)=1$，故$a=\dfrac{1}{3}$.由$\begin{cases}\dfrac{1}{3}-d\geqslant 0,\\ \dfrac{1}{3}+d\geqslant 0,\end{cases}$解得$-\dfrac{1}{3}\leqslant d\leqslant\dfrac{1}{3}$.

类型 3　求离散型随机变量的分布列

例3　某射手有 5 发子弹，射击一次命中率为 0.8，若命中就停止射击，否则一直到子弹用尽，求耗用子弹数$X$的分布列.

**【解答】** $X$的取值为 1,2,3,4,5.当$X=1$时，即第一枪就中了，故$P(X=1)=0.8$；当$X=2$时，即第一枪未中，第二枪中了，故$P(X=2)=0.2\times 0.8=0.16$；同理，$P(X=3)=0.2^2\times 0.8=0.032$；$P(X=4)=0.2^3\times 0.8=0.006\ 4$；$P(X=5)=0.2^4=0.001\ 6$.则$X$的分布列为

| $X$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $P$ | 0.8 | 0.16 | 0.032 | 0.006 4 | 0.001 6 |

变式　袋中有 8 个大小相同的小球，其中 1 个黑球，3 个白球，4 个红球.

(1) 若从袋中一次摸出 2 个小球，求恰为异色球的概率；

(2) 若从袋中一次摸出 3 个小球，且 3 个球中，黑球与白球的个数都没有超过红球的个数，记此时红球的个数为$\xi$，求$\xi$的分布列.

**【解答】** (1) 摸出的 2 个小球为异色球的种数为$C_1^1C_7^1+C_3^1C_4^1=19$，从 8 个球中摸出 2 个小球的种数为$C_8^2=28$，故所求概率为$P=\dfrac{19}{28}$.

(2) 符合条件的摸法包括以下三种：①有 1 个红球，1 个黑球，1 个白球，共

有 $C_1^1C_4^1C_3^1=12$(种)．②有 2 个红球，1 个其他颜色球，共有 $C_4^2C_4^1=24$(种)．③所摸得的 3 个球均为红球，共有 $C_4^3=4$ 种不同摸法，故符合条件的不同摸法共有 40 种．故随机变量 $\xi$ 的取值为 1,2,3，且 $P(\xi=1)=\frac{12}{40}=\frac{3}{10}$，$P(\xi=2)=\frac{24}{40}=\frac{3}{5}$，$P(\xi=3)=\frac{4}{40}=\frac{1}{10}$，

其分布列为

| $\xi$ | 1 | 2 | 3 |
| --- | --- | --- | --- |
| $P$ | $\frac{3}{10}$ | $\frac{3}{5}$ | $\frac{1}{10}$ |

【规律总结】 求离散型随机变量的分布列的步骤：

(1) 求出随机变量 $\xi$ 的所有可能的取值 $x_i(i=1,2,\cdots,n)$．

(2) 求出取每一个值的概率 $P(\xi=x_i)=p_i$.

(3) 列出表格．

课堂评价▶基础达标

1．(多选)下列是离散型随机变量的是(　ABD　)

A．某机场候机室中一天的旅客数量 $X$

B．连续投掷一枚均匀硬币 4 次，正面向上的次数 $X$

C．某篮球下降过程中离地面的距离 $X$

D．某道路斑马线一天经过的人数 $X$

【解析】 A，B，D 中的随机变量 $X$ 可能取的值，我们都可以按一定次序一一列出，因此，它们都是离散型随机变量；C 中的 $X$ 可以取某一区间内的一切值，无法按一定次序一一列出，故 C 中的 $X$ 不是离散型随机变量．

2．(多选)下列问题中的随机变量服从两点分布的是(　BCD　)

A．抛掷一枚骰子，所得点数为随机变量 $X$

B．某射手射击一次，击中目标的次数为随机变量 $X$

C．从装有 5 个红球，3 个白球的袋中取 1 个球，令随机变量 $X=$

$\begin{cases}1，取出白球，\\0，取出红球\end{cases}$

D．某医生做一次手术，手术成功的次数为随机变量 $X$

3．抛掷两枚质地均匀的骰子，记第一枚骰子掷出的点数与第二枚骰子掷出的点数之差为 $\xi$，则“$\xi>4$”表示的试验结果是(　D　)

A．第一枚掷出 6 点，第二枚掷出 2 点

B．第一枚掷出 5 点，第二枚掷出 1 点

C．第一枚掷出 2 点，第二枚掷出 6 点

D．第一枚掷出 6 点，第二枚掷出 1 点

【解析】 只有 D 中的点数差为 $6-1=5>4$，其余均不是．

4．从 4 张编号为 1～4 的卡片中任意取出 2 张，取出的卡片号码数之和为 $X$，求随机变量 $X$ 的概率分布．

【解答】 $X$ 可取 3,4,5,6,7．其中 $X=3$ 表示取出分别标有 1,2 的 2 张卡片，$P(X=3)=\frac{1}{C_4^2}=\frac{1}{6}$；

$X=4$ 表示取出分别标有 1,3 的 2 张卡片，$P(X=4)=\frac{1}{C_4^2}=\frac{1}{6}$；

$X=5$ 表示取出分别标有 1,4 或 2,3 的 2 张卡片，$P(X=5)=\frac{2}{C_4^2}=\frac{1}{3}$；

$X=6$ 表示取出分别标有 2,4 的 2 张卡片，$P(X=6)=\frac{1}{6}$；

$X=7$ 表示取出分别标有 3,4 的 2 张卡片，$P(X=7)=\frac{1}{6}$.

所以随机变量 $X$ 的概率分布为

| $X$ | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|
| $P$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{3}$ | $\frac{1}{6}$ | $\frac{1}{6}$ |

# 7．3　离散型随机变量的数字特征

## 第 1 课时　离散型随机变量的均值

### 预习评价▶新知初探

### 学习目标

1．理解离散型随机变量的均值的意义和性质，会根据离散型随机变量的分布列求出均值．

2．掌握两点分布的均值．

3．会利用离散型随机变量的均值，解决一些相关问题．

### 问题导引

1．已知随机变量 $X$ 服从参数为 $p$ 的两点分布，求 $E(X)$．

2．已知 $X$ 是一个随机变量，设 $a$，$b$ 都是实数且 $a\neq0$，则 $Y=aX+b$ 也是一个随机变量，那么，这两个随机变量的均值之间有什么联系呢？

### 现学现用

1．概念辨析(正确的打“√”，错误的打“×”)

(1) 随机变量 $X$ 的数学期望 $E(X)$是个变量，其随 $X$ 的变化而变化．(　×　)

(2) 随机变量的均值与样本的平均值相同．(　×　)

(3) 若随机变量 $X$ 的数学期望 $E(X)=2$，则 $E(2X)=4$.(　√　)

(4) 若某人投篮的命中率为 0.8，那么他投篮 10 次一定会进 8 个球．(　×　)

2．若离散型随机变量 $X$ 的分布列如下表，则 $X$ 的数学期望 $E(X)$等于(　A　)

| $X$ | 1 | 2 | 3 |
|---|---|---|---|
| $P$ | $\frac{3}{5}$ | $\frac{3}{10}$ | $\frac{1}{10}$ |

A．$\frac{3}{2}$　　　　B．2

C．$\frac{5}{2}$　　　　D．3

3．设 $E(X)=10$，则 $E(3X+5)=$__35__．

4．篮球运动员在比赛中每次罚球命中得 1 分，不命中得 0 分．已知他命中的概率为 0.8，则罚球一次得分 $X$ 的期望是__0.8__．

【解析】 由题意知，$X$ 服从两点分布，所以 $E(X)=1\times0.8=0.8$.

### 对点练▶先掌握基础

类型 1　离散型随机变量的均值

**例1**　某地最近出台一项机动车驾照考试规定：每位考试者一年之内最多有 4 次参加考试的机会，一旦某次考试通过，即可领取驾照，不再参加以后的考试，否则就一直考到第 4 次为止．如果李明决定参加驾照考试，设他每次参加考试通过的概率依次为 0.6,0.7,0.8,0.9，求在一年内李明参加驾照考试次数 $X$ 的分布列和 $X$ 的均值．

【解答】　$X$ 的取值分别为 1,2,3,4.$X=1$，表明李明第一次参加驾照考试就通过了，故 $P(X=1)=0.6$.

$X=2$，表明李明第一次考试未通过，第二次通过了，故 $P(X=2)=(1-0.6)\times0.7=0.28$.

$X=3$，表明李明第一、二次考试未通过，第三次通过了，故 $P(X=3)=(1-0.6)\times(1-0.7)\times0.8=0.096$.

$X=4$，表明李明第一、二、三次考试都未通过，故 $P(X=4)=(1-0.6)\times(1-0.7)\times(1-0.8)=0.024$.故 $X$ 的分布列为

| $X$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $P$ | 0.6 | 0.28 | 0.096 | 0.024 |

所以 $X$ 的均值 $E(X)=1\times0.6+2\times0.28+3\times0.096+4\times0.024=1.544$.

【规律总结】　求离散型随机变量 $X$ 的均值的步骤：

(1) 根据 $X$ 的实际意义，写出 $X$ 的全部取值；

(2) 求出 $X$ 的每个值的概率；

(3) 写出 $X$ 的分布列；

(4) 利用定义求出均值．

其中第(1)、(2)两条是解答此类题目的关键，在求解过程中应注重分析概率的相关知识．

类型 2　离散型随机变量均值公式及性质

例2　已知随机变量 $X$ 的分布列为

| $X$ | $-2$ | $-1$ | $0$ | $1$ | $2$ |
|---|---|---|---|---|---|
| $P$ | $\frac{1}{4}$ | $\frac{1}{3}$ | $\frac{1}{5}$ | $m$ | $\frac{1}{20}$ |

则 $E(X)=$<u>$-\frac{17}{30}$</u>.若 $\xi=aX+3$，且 $E(\xi)=-\frac{11}{2}$，则 $a$ 的值为<u>　15　</u>.

【解析】 由随机变量分布列的性质，得$\frac{1}{4}+\frac{1}{3}+\frac{1}{5}+m+\frac{1}{20}=1$，解得 $m=\frac{1}{6}$，所以 $E(X)=(-2)\times\frac{1}{4}+(-1)\times\frac{1}{3}+0\times\frac{1}{5}+1\times\frac{1}{6}+2\times\frac{1}{20}=-\frac{17}{30}$.由 $E(\xi)=E(aX+3)=aE(X)+3=-\frac{17}{30}a+3=-\frac{11}{2}$，所以 $a=15$.

【规律总结】 与离散型随机变量性质有关问题的解题思路：若 $\xi=aX+b$，$a\neq0$，$a$，$b$ 为常数，求 $E(\xi)$，一般思路是先求出 $E(X)$，再利用公式 $E(aX+b)=aE(X)+b$ 求 $E(\xi)$．也可以利用 $X$ 的分布列得到 $\xi$ 的分布列，关键由 $X$ 的取值计算 $\xi$ 的取值，对应的概率相等，再由定义法求得 $E(\xi)$．

类型 3　离散型随机变量均值的应用

例3　某中药种植基地有两处种植区的药材需在下周一、周二两天内采摘完毕，基地员工一天可以完成一处种植区的采摘．由于下雨会影响药材品质，基地收益如下表所示：

| | | | | |
|---|---|---|---|---|
| 周一 | 无雨 | 无雨 | 有雨 | 有雨 |
| 周二 | 无雨 | 有雨 | 无雨 | 有雨 |
| 收益 | 20 万元 | 15 万元 | 10 万元 | 7.5 万元 |

若基地额外聘请工人，可在周一当天完成全部采摘任务．无雨时收益为 20 万元；有雨时收益为 10 万元．额外聘请工人的成本为 $a$ 万元．已知下周一和下周二有雨的概率相同，两天是否下雨互不影响，基地收益为 20 万元的概率为 0.36.

(1) 若不额外聘请工人，写出基地收益 $X$ 的分布列及基地的预期收益；

(2) 该基地是否应该外聘工人，请说明理由．

【解答】 (1) 设下周一无雨的概率为 $p$，由题意知，$p^2=0.36$，$p=0.6$，基地

收益 $X$ 的可能取值为 20,15,10,7.5，则 $P(X=20)=0.36$，$P(X=15)=0.24$，$P(X=10)=0.24$，$P(X=7.5)=0.16$，

故 $X$ 的分布列为

| $X$ | 20 | 15 | 10 | 7.5 |
|---|---|---|---|---|
| $P$ | 0.36 | 0.24 | 0.24 | 0.16 |

所以 $E(X)=20\times0.36+15\times0.24+10\times0.24+7.5\times0.16=14.4$，所以基地的预期收益为 14.4 万元.

(2) 设基地额外聘请工人时的收益为 $Y$ 万元，则其预期收益 $E(Y)=20\times0.6+10\times0.4-a=16-a$(万元)，

$E(Y)-E(X)=1.6-a$.

综上，当额外聘请工人的成本高于 1.6 万元时，不外聘工人；成本低于 1.6 万元时，外聘工人；成本恰为 1.6 万元时，是否外聘工人均可以.

**变式** 某种产品共 200 件，其中一等品 126 件，二等品 50 件，三等品 20 件，次品 4 件．已知生产 1 件一、二、三等品获得的利润分别为 6 万元、2 万元、1 万元，而 1 件次品亏损 2 万元，设 1 件产品的利润(单位：万元)为 $X$.

(1) 求 $X$ 的分布列；

(2) 求 1 件产品的平均利润(即 $X$ 的数学期望)；

(3) 经技术革新后，仍有四个等级的产品，但次品率降为 1%，一等品率提高为 70%，如果此时要求 1 件产品的平均利润不小于 4.73 万元，则三等品率最多是多少？

【解答】 (1) $X$ 的所有可能取值有 6,2,1，$-2$.

$P(X=6)=\dfrac{126}{200}=0.63$，$P(X=2)=\dfrac{50}{200}=0.25$，

$P(X=1)=\dfrac{20}{200}=0.1$，$P(X=-2)=\dfrac{4}{200}=0.02$.

故 $X$ 的分布列为

| $X$ | 6 | 2 | 1 | $-2$ |
|---|---|---|---|---|
| $P$ | 0.63 | 0.25 | 0.1 | 0.02 |

(2) $E(X)=6\times0.63+2\times0.25+1\times0.1+(-2)\times0.02=4.34$.

(3) 设技术革新后的三等品率为 $x$，则此时 1 件产品的平均利润为 $E(X)=6\times0.7+2\times(1-0.7-0.01-x)+1\times x+(-2)\times0.01=4.76-x(0\leqslant x\leqslant0.29)$．依题意，$E(X)\geqslant4.73$，即 $4.76-x\geqslant4.73$，解得 $x\leqslant0.03$，所以三等品率最多为 3%.

【规律总结】 1．实际问题中的均值问题：

均值在实际生活中有着广泛的应用，如对体育比赛的成绩预测、消费预测、工程方案的预测、产品合格率的预测、投资收益的预测等，都可以通过随机变量的均值来进行估计.

2．概率模型的三个解答步骤：

(1) 审题，确定实际问题是哪一种概率模型，可能用到的事件类型，所用的公式有哪些.

(2) 确定随机变量的分布列，计算随机变量的均值.

(3) 对照实际意义，回答概率、均值等所表示的结论.

课堂评价▶基础达标

1．若随机变量 $\xi$ 的分布列如下表，则 $E(5\xi+4)$等于(　A　)

| $\xi$ | 1 | 2 | 4 |
|---|---|---|---|
| $P$ | 0.4 | 0.3 | 0.3 |

A．15　　　　B．11

C．2.2　　　　D．2.3

【解析】由已知得 $E(\xi)=1\times0.4+2\times0.3+4\times0.3=2.2$，所以 $E(5\xi+4)=5E(\xi)+4=5\times2.2+4=15$.

2．已知离散型随机变量 $\xi$ 的分布列如下表，则 $E(\xi)$等于(　D　)

| $\xi$ | 1 | 3 | 5 |
|---|---|---|---|
| $P$ | 0.5 | $m$ | 0.2 |

A．1　　　　B．0.6

C．2.44　　　　D．2.4

【解析】 因为分布列中所有的概率之和等于 1，所以 $0.5+m+0.2=1$，所以 $m=0.3$，所以 $E(\xi)=1\times0.5+3\times0.3+5\times0.2=2.4$.

3．设某项试验的成功率是失败率的 2 倍，用随机变量 $X$ 描述 1 次试验的成功次数，则 $X$ 的值可以是( C )

A．2　　B．2 或 1

C．1 或 0　　D．2 或 1 或 0

【解析】 这里“成功率是失败率的 2 倍”是干扰条件，对 1 次试验的成功次数没有影响，故 $X$ 的可能取值有两种，即 0,1.

4．盒中装有 5 节同品牌的五号电池，其中混有 2 节废电池．现在无放回地每次取一节电池检验，直到取到好电池为止，求抽取次数 $X$ 的分布列及均值．

【解答】 $X$ 可取的值为 1,2,3，则 $P(X=1)=\frac{3}{5}$，$P(X=2)=\frac{2}{5}\times\frac{3}{4}=\frac{3}{10}$，$P(X=3)=\frac{2}{5}\times\frac{1}{4}\times1=\frac{1}{10}$.

故抽取次数 $X$ 的分布列为

| $X$ | 1 | 2 | 3 |
|---|---|---|---|
| $P$ | $\frac{3}{5}$ | $\frac{3}{10}$ | $\frac{1}{10}$ |

所以 $E(X)=1\times\frac{3}{5}+2\times\frac{3}{10}+3\times\frac{1}{10}=\frac{3}{2}$.

第 2 课时　离散型随机变量的方差

预习评价▶新知初探

**学习目标**

1．理解取有限个值的离散型随机变量的方差及标准差的概念．

2．能计算简单离散型随机变量的方差，并能解决一些实际问题．

3．掌握方差的性质以及两点分布方差的求法，会利用公式求它们的方差．

**问题导引**

1．已知 $X$ 是一个随机变量，设 $a$，$b$ 都是实数，且 $a\neq 0$，则 $Y=aX+b$ 也是一个随机变量，且 $E(Y)=aE(X)+b$，那么这两个随机变量的方差之间有什么联系呢？

2．随机变量的均值、方差与样本均值、方差的关系是怎样的？

**现学现用**

1．概念辨析(正确的打“√”，错误的打“×”)

(1) 离散型随机变量的方差越大，随机变量越稳定．(　×　)

(2) 若 $a$ 是常数，则 $D(a)=0$.(　√　)

(3) 离散型随机变量的方差反映了随机变量偏离于期望的平均程度．(　√　)

2．已知随机变量 $X$，$D(X)=\frac{1}{9}$，则 $X$ 的标准差为__$\frac{1}{3}$__．

3．有两台自动包装机甲与乙，包装质量分别为随机变量 $X_1$，$X_2$，已知 $E(X_1)=E(X_2)$，$D(X_1)>D(X_2)$，则自动包装机__乙__的质量较好．

**对点练▶先掌握基础**

类型 1　离散型随机变量的方差、标准差

**例1**　编号为 1,2,3 的三位学生随意入座编号为 1,2,3 的三个座位，每位学生坐一个座位，设与座位编号相同的学生的人数是 $\xi$，求 $E(\xi)$和 $D(\xi)$．

【解答】 $\xi$ 的所有可能取值为 0,1,3，$\xi=0$ 表示三位同学全坐错了，有 2 种情况，即编号为 1,2,3 的座位上分别坐了编号为 2,3,1 或 3,1,2 的学生，则 $P(\xi=0)=\frac{2}{A_3^3}=\frac{1}{3}$；$\xi=1$ 表示三位同学只有 1 位同学坐对了，则 $P(\xi=1)=\frac{C_3^1}{A_3^3}=\frac{1}{2}$；$\xi=3$ 表示三位学生全坐对，即对号入座，则 $P(\xi=3)=\frac{1}{A_3^3}=\frac{1}{6}$.

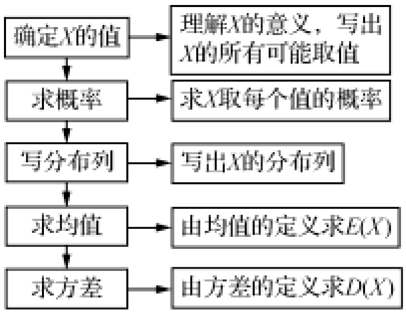
所以，$\xi$ 的分布列为

| $\xi$ | 0 | 1 | 3 |
|---|---|---|---|
| $P$ | $\frac{1}{3}$ | $\frac{1}{2}$ | $\frac{1}{6}$ |

$E(\xi)=0\times\frac{1}{3}+1\times\frac{1}{2}+3\times\frac{1}{6}=1$；

$D(\xi)=\frac{1}{3}\times(0-1)^2+\frac{1}{2}\times(1-1)^2+\frac{1}{6}\times(3-1)^2=1.$

【规律总结】 求离散型随机变量 $X$ 的方差的一般步骤：




类型 2 离散型随机变量方差的性质

**例2** 已知 $\eta$ 的分布列为

| $\eta$ | 0 | 10 | 20 | 50 | 60 |
|---|---|---|---|---|---|
| $P$ | $\frac{1}{3}$ | $\frac{2}{5}$ | $\frac{1}{15}$ | $\frac{2}{15}$ | $\frac{1}{15}$ |

(1) 求 $\eta$ 的方差及标准差；

(2) 设 $Y=2\eta-E(\eta)$，求 $D(Y)$.

【解答】 (1) 因为 $E(\eta)=0\times\frac{1}{3}+10\times\frac{2}{5}+20\times\frac{1}{15}+50\times\frac{2}{15}+60\times\frac{1}{15}=16$，$D(\eta)=(0-16)^2\times\frac{1}{3}+(10-16)^2\times\frac{2}{5}+(20-16)^2\times\frac{1}{15}+(50-16)^2\times\frac{2}{15}+(60-16)^2\times\frac{1}{15}=384$，所以$\sqrt{D(\eta)}=8\sqrt{6}$.

(2) 因为 $Y=2\eta-E(\eta)$，所以 $D(Y)=D[2\eta-E(\eta)]=2^2D(\eta)=4\times384=1\ 536$.

【规律总结】 对于变量间存在关系的方差，在求解过程中应注意方差性质的应用，如 $D(a\xi+b)=a^2D(\xi)$，这样处理既避免了求随机变量 $\eta=a\xi+b$ 的分布列，又避免了繁杂的计算，简化了计算过程.

类型 3　离散型随机变量方差的应用

例3　甲、乙两名射手在一次射击中得分为两个相互独立的随机变量 $\xi$，$\eta$，已知甲、乙两名射手在每次射击中射中的环数大于 6 环，且甲射中 10,9,8,7 环的概率分别为 0.5,3$a$，$a$,0.1，乙射中 10,9,8 环的概率分别为 0.3,0.3,0.2.

(1) 求 $\xi$，$\eta$ 的分布列；

(2) 求 $\xi$，$\eta$ 的均值与方差，并以此比较甲、乙的射击技术.

【解答】 (1) 由题意得：$0.5+3a+a+0.1=1$，解得 $a=0.1$.因为乙射中 10,9,8 环的概率分别为 0.3,0.3,0.2，

所以乙射中 7 环的概率为 $1-(0.3+0.3+0.2)=0.2$.

所以 $\xi$，$\eta$ 的分布列分别为

| $\xi$ | 10 | 9 | 8 | 7 |
|---|---|---|---|---|
| $P$ | 0.5 | 0.3 | 0.1 | 0.1 |

| $\eta$ | 10 | 9 | 8 | 7 |
|---|---|---|---|---|
| $P$ | 0.3 | 0.3 | 0.2 | 0.2 |

(2) 由(1)得 $E(\xi)=10\times0.5+9\times0.3+8\times0.1+7\times0.1=9.2$；$E(\eta)=10\times0.3+9\times0.3+8\times0.2+7\times0.2=8.7$；

$D(\xi)=(10-9.2)^2\times0.5+(9-9.2)^2\times0.3+(8-9.2)^2\times0.1+(7-9.2)^2\times0.1=0.96$；

$D(\eta)=(10-8.7)^2\times0.3+(9-8.7)^2\times0.3+(8-8.7)^2\times0.2+(7-8.7)^2\times0.2=1.21$.

由于 $E(\xi)>E(\eta)$，$D(\xi)<D(\eta)$，说明甲射击的环数的均值比乙高，且成绩比较稳定，所以甲比乙的射击技术好.

变式　有甲、乙两名学生，经统计，他们在解答同一份数学试卷时，各自的成绩在 80 分、90 分、100 分的概率分布大致如下表所示：

甲

| 分数 $X$ | 80 | 90 | 100 |
| --- | --- | --- | --- |
| 概率 $P$ | 0.2 | 0.6 | 0.2 |

乙

| 分数 $Y$ | 80 | 90 | 100 |
| --- | --- | --- | --- |
| 概率 $P$ | 0.4 | 0.2 | 0.4 |

试分析两名学生的成绩水平．

【解答】 因为 $E(X)=80\times0.2+90\times0.6+100\times0.2=90$，$D(X)=(80-90)^2\times0.2+(90-90)^2\times0.6+(100-90)^2\times0.2=40$，$E(Y)=80\times0.4+90\times0.2+100\times0.4=90$，$D(Y)=(80-90)^2\times0.4+(90-90)^2\times0.2+(100-90)^2\times0.4=80$，即 $E(X)=E(Y)$，$D(X)<D(Y)$，所以甲生与乙生的成绩均值一样，甲的方差较小，因此甲生的学习成绩较稳定．

【规律总结】 利用均值和方差的意义分析解决实际问题的步骤：

(1) 比较均值：离散型随机变量的均值反映了离散型随机变量取值的平均水平，因此，在实际决策问题中，需先计算均值，看一下谁的平均水平高．

(2) 比较方差：在均值相等的情况下计算方差，方差反映了离散型随机变量取值的稳定与波动、集中与离散的程度，通过计算方差分析谁的水平发挥相对稳定．

(3) 下结论：依据方差的几何意义做出结论．

课堂评价▶基础达标

1．若随机变量 $\xi$ 满足 $P(\xi=1)=0.3$，$P(\xi=2)=0.7$，则 $E(\xi)$和 $D(\xi)$的值分别为( D )

A．0.6 和 0.7　　B．1.7 和 0.09

C．0.3 和 0.7　　D．1.7 和 0.21

【解析】 $E(\xi)=1\times0.3+2\times0.7=1.7$，$D(\xi)=(1-1.7)^2\times0.3+(2-1.7)^2\times0.7=0.21$．

2．若随机变量 $\xi$ 的分布列如下表，且 $E(\xi)=1.1$，则 $D(\xi)$等于(　C　)

| $\xi$ | 0 | 1 | $x$ |
|---|---|---|---|
| $P$ | $\frac{1}{5}$ | $p$ | $\frac{3}{10}$ |

A．0.36　　　　B．0.52

C．0.49　　　　D．0.68

【解析】由随机变量分布列的性质得 $p=\frac{1}{2}$，则 $E(\xi)=0\times\frac{1}{5}+1\times\frac{1}{2}+\frac{3}{10}x=1.1$，解得 $x=2$，所以 $D(\xi)=(0-1.1)^2\times\frac{1}{5}+(1-1.1)^2\times\frac{1}{2}+(2-1.1)^2\times\frac{3}{10}=0.49$.

3．设 $0<a<1$，随机变量 $X$ 的分布列如下表，则当 $a$ 在(0,1)内增大时 $D(X)$(　D　)

| $X$ | 0 | $a$ | 1 |
|---|---|---|---|
| $p$ | $\frac{1}{3}$ | $\frac{1}{3}$ | $\frac{1}{3}$ |

A．增大　　　　B．减小

C．先增大后减小　　　　D．先减小后增大

【解析】 由分布列得 $E(X)=\frac{1+a}{3}$，则 $D(X)=\left(\frac{1+a}{3}-0\right)^2\times\frac{1}{3}+\left(\frac{1+a}{3}-a\right)^2\times\frac{1}{3}+\left(\frac{1+a}{3}-1\right)^2\times\frac{1}{3}=\frac{2}{9}\left(a-\frac{1}{2}\right)^2+\frac{1}{6}$，则当 $a$ 在(0,1)内增大时，$D(X)$先减小后增大.

4．检测甲、乙水稻各 10 株的分蘖数据，计算出 $E(X_{甲})=E(X_{乙})$，$D(X_{甲})=11$，$D(X_{乙})=3.4$.由此可以估计(　B　)

A．甲种水稻比乙种水稻分蘖整齐

B．乙种水稻比甲种水稻分蘖整齐

C．甲、乙两种水稻分蘖整齐程度相同

D．甲、乙两种水稻分蘖整齐程度不能比较

【解析】 因为 $E(X_{甲})=E(X_{乙})$，$D(X_{甲})>D(X_{乙})$，所以乙种水稻比甲种水稻分蘖整齐.

5．已知随机变量 $X$ 的分布列如下表，求 $D(X)$和 $D(2X-1)$.

| $X$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| $P$ | 0.2 | 0.2 | 0.3 | 0.2 | 0.1 |

【解答】 因为 $E(X)=0\times0.2+1\times0.2+2\times0.3+3\times0.2+4\times0.1=1.8$，所以 $D(X)=(0-1.8)^2\times0.2+(1-1.8)^2\times0.2+(2-1.8)^2\times0.3+(3-1.8)^2\times0.2+(4-1.8)^2\times0.1=1.56$，所以 $D(2X-1)=4D(X)=4\times1.56=6.24$.

## 习题课　均值与方差

### 题型 1　离散型随机变量的均值

**例1**　某婴幼儿游泳馆为了吸引顾客，推出优惠活动，即对首次消费的顾客按 80 元收费，并注册成为会员，对会员消费的不同次数给予相应的优惠，标准如下：

| 消费次数 | 第 1 次 | 第 2 次 | 第 3 次 | 不少于 4 次 |
|---|---|---|---|---|
| 收费比例 | 1 | 0.95 | 0.90 | 0.85 |

该游泳馆从注册的会员中，随机抽取了 100 位会员统计他们的消费次数，得到数据如下：

| 消费次数 | 1 次 | 2 次 | 3 次 | 不少于 4 次 |
|---|---|---|---|---|
| 频数 | 60 | 25 | 10 | 5 |

假设每位顾客游泳 1 次，游泳馆的成本为 30 元，每个会员最多消费 4 次，以事件发生的频率作为相应事件发生的概率，从该游泳馆的会员中随机抽取 2 位，记游泳馆从这 2 位会员的消费中获得的平均利润之差的绝对值为 $X$，求 $X$ 的分布列和均值 $E(X)$.

【解答】 若会员消费 1 次，$P_1=\frac{60}{100}=\frac{3}{5}$，

则平均利润为 50 元，其概率为$\frac{3}{5}$；

若会员消费 2 次，$P_2=\frac{25}{100}=\frac{1}{4}$，

则平均利润为$\frac{50+46}{2}=48$(元)，其概率为$\frac{1}{4}$；

若会员消费 3 次，$P_3=\frac{10}{100}=\frac{1}{10}$，

则平均利润为$\frac{50+46+42}{3}=46$(元)，其概率为$\frac{1}{10}$；

若会员消费 4 次，$P_4=\frac{5}{100}=\frac{1}{20}$，

则平均利润为$\frac{50+46+42+38}{4}=44$(元)，其概率为$\frac{1}{20}$.

由题意知，$X$的所有可能取值为 0,2,4,6，

且 $P(X=0)=\frac{3}{5}\times\frac{3}{5}+\frac{1}{4}\times\frac{1}{4}+\frac{1}{10}\times\frac{1}{10}+\frac{1}{20}\times\frac{1}{20}=\frac{87}{200}$，

$P(X=2)=2\times\left(\frac{3}{5}\times\frac{1}{4}+\frac{1}{4}\times\frac{1}{10}+\frac{1}{10}\times\frac{1}{20}\right)=\frac{9}{25}$，

$P(X=4)=2\times\left(\frac{3}{5}\times\frac{1}{10}+\frac{1}{4}\times\frac{1}{20}\right)=\frac{29}{200}$，

$P(X=6)=2\times\frac{3}{5}\times\frac{1}{20}=\frac{3}{50}$.

所以$X$的分布列为

| $X$ | 0 | 2 | 4 | 6 |
|---|---|---|---|---|
| $P$ | $\frac{87}{200}$ | $\frac{9}{25}$ | $\frac{29}{200}$ | $\frac{3}{50}$ |

所以 $E(X)=0\times\frac{87}{200}+2\times\frac{9}{25}+4\times\frac{29}{200}+6\times\frac{3}{50}=\frac{83}{50}$.

**变式**　某公司年会有幸运抽奖环节，一个箱子里有相同的十个乒乓球，球上分别标有 0,1，2，…，9 这十个自然数，每位员工有放回地依次取出三个球，规定：每次取出的球所标数字不小于后面取出的球所标数字即中奖．中奖奖项：三个数字全部相同中一等奖，奖励 10 000 元现金；三个数字中有两个数字相同中二等奖，奖励 5 000 元现金；三个数字各不相同中三等奖，奖励 2 000 元现金；其他不中奖，没有奖金．

(1) 求员工$A$中二等奖的概率；

(2) 设员工$A$中奖奖金为$X$，求$X$的分布列；

(3) 员工$B$是优秀员工，有两次抽奖机会，求员工$B$中奖奖金的期望．

**【解答】** (1) 记事件$M$为“员工$A$中二等奖”，有放回地依次取三个球的取法有 $10^3$ 种．

中二等奖取法有两类：一类是前两次取到同一数字，从 10 个数字中取出 2 个，较大的数是前两次取出的数，较小的数是第 3 次取出的数，有 $C_{10}^2$种；另一类是后两次取到同一数字，同理有 $C_{10}^2=45$ 种，共 90 种，则 $P(M)=\frac{90}{10^3}=0.09$.

(2) $X$的可能取值为 0,2 000,5 000，10 000，且 $P(X=2\ 000)=\frac{C_{10}^{3}}{10^{3}}=0.12$；

$P(X=5\ 000)=0.09$；

$P(X=10\ 000)=\frac{10}{10^{3}}=0.01$；

$P(X=0)=1-P(X=2\ 000)-P(X=5\ 000)-P(X=10\ 000)=0.78$.

故 $X$ 的分布列为

| $X$ | 10 000 | 5 000 | 2 000 | 0 |
|---|---|---|---|---|
| $P$ | 0.01 | 0.09 | 0.12 | 0.78 |

(3) 由(2)可知 $A$ 中奖奖金的期望 $E(X)=10\ 000\times0.01+5\ 000\times0.09+2\ 000\times0.12+0\times0.78=790$ 元.

员工 $B$ 每次中奖奖金的期望和 $A$ 一样，

由题意可知，员工 $B$ 中奖奖金的期望是 1 580 元.

题型 2　离散型随机变量的方差

**例2**　某滑雪场的收费标准是：滑雪时间不超过 1 小时免费，超过 1 小时的部分每小时收费标准为 40 元(不足 1 小时的部分按 1 小时计算)．有甲、乙两人相互独立地来该滑雪场运动，设甲、乙不超过 1 小时离开的概率分别为$\frac{1}{4}$，$\frac{1}{6}$；1 小时以上且不超过 2 小时离开的概率分别为$\frac{1}{2}$，$\frac{2}{3}$；两人滑雪时间都不会超过 3 小时．设甲、乙两人所付的滑雪费用之和为随机变量 $\xi$，求 $\xi$ 的分布列，均值 $E(\xi)$与方差 $D(\xi)$.

**【解答】**　甲、乙两人滑雪 2 小时以上且不超过 3 小时离开的概率分别为 $\left(1-\frac{1}{4}-\frac{1}{2}\right)=\frac{1}{4}$，$\left(1-\frac{1}{6}-\frac{2}{3}\right)=\frac{1}{6}$，两人所付费用相同，相同的费用可能为 0,40,80 元，

则 $\xi$ 的所有可能取值为 0,40,80,120,160，

且 $P(\xi=0)=\frac{1}{4}\times\frac{1}{6}=\frac{1}{24}$，

$P(\xi=40)=\frac{1}{4}\times\frac{2}{3}+\frac{1}{2}\times\frac{1}{6}=\frac{1}{4}$，

$P(\xi=80)=\frac{1}{4}\times\frac{1}{6}+\frac{1}{2}\times\frac{2}{3}+\frac{1}{4}\times\frac{1}{6}=\frac{5}{12}$,

$P(\xi=120)=\frac{1}{2}\times\frac{1}{6}+\frac{1}{4}\times\frac{2}{3}=\frac{1}{4}$,

$P(\xi=160)=\frac{1}{4}\times\frac{1}{6}=\frac{1}{24}$.

所以 $\xi$ 的分布列为

| $\xi$ | 0 | 40 | 80 | 120 | 160 |
| --- | --- | --- | --- | --- | --- |
| $P$ | $\frac{1}{24}$ | $\frac{1}{4}$ | $\frac{5}{12}$ | $\frac{1}{4}$ | $\frac{1}{24}$ |

$E(\xi)=0\times\frac{1}{24}+40\times\frac{1}{4}+80\times\frac{5}{12}+120\times\frac{1}{4}+160\times\frac{1}{24}=80$，

$D(\xi)=(0-80)^2\times\frac{1}{24}+(40-80)^2\times\frac{1}{4}+(80-80)^2\times\frac{5}{12}+(120-80)^2\times\frac{1}{4}+(160-80)^2\times\frac{1}{24}=\frac{4\,000}{3}$.

题型 3　均值与方差在决策中的应用

**例3**　某投资公司准备将 1 000 万元投资到“低碳”项目上，现有两个项目供选择：

项目一：新能源汽车．据市场调研，投资到该项目上，到年底可能获利 30%，也可能亏损 15%，且这两种情况发生的概率分别为$\frac{7}{9}$和$\frac{2}{9}$.

项目二：通信设备．据市场调研，投资到该项目上，到年底可能获利 50%，可能损失 30%，也可能不赔不赚，且这三种情况发生的概率分别为$\frac{3}{5}$，$\frac{1}{3}$和$\frac{1}{15}$.

针对以上两个投资项目，请你为投资公司选择一个合理的项目，并说明理由．

**【解答】**　若投资“项目一”，设获利为 $X_1$ 万元，则 $X_1$ 的分布列为

| $X_1$ | 300 | $-150$ |
| --- | --- | --- |
| $P$ | $\frac{7}{9}$ | $\frac{2}{9}$ |

所以 $E(X_1)=300\times\frac{7}{9}+(-150)\times\frac{2}{9}=200$，

$D(X_1)=(300-200)^2\times\frac{7}{9}+(-150-200)^2\times\frac{2}{9}=35\,000$.

若投资“项目二”，设获利为 $X_2$ 万元，则 $X_2$ 的分布列为

| $X_2$ | 500 | $-300$ | 0 |
|---|---|---|---|
| $P$ | $\frac{3}{5}$ | $\frac{1}{3}$ | $\frac{1}{15}$ |

所以 $E(X_2)=500\times\frac{3}{5}+(-300)\times\frac{1}{3}+0\times\frac{1}{15}=200$，

$D(X_2)=(500-200)^2\times\frac{3}{5}+(-300-200)^2\times\frac{1}{3}+(0-200)^2\times\frac{1}{15}=140\ 000$.

所以 $E(X_1)=E(X_2)$，$D(X_1)<D(X_2)$．

这说明虽然项目一、项目二获利相等，但项目一更稳妥．

综上所述，建议该投资公司选择项目一投资．

**变式** 东方商店欲每两天购进一次某种食品(保质期两天，购进时该食品为刚生产的)，根据市场调查，该食品每份进价 8 元，售价 12 元，如果两天内无法售出，则食品过期作废，且两天内的销售情况互不影响．为了了解市场的需求情况，现统计该产品在本地区 100 天的销售量如下表：(视样本频率为概率)

| 销售量/份 | 15 | 16 | 17 | 18 |
|---|---|---|---|---|
| 天数 | 20 | 30 | 40 | 10 |

(1) 根据该产品 100 天的销售量统计表，记两天中一共销售该食品份数为 $\xi$，求 $\xi$ 的分布列与均值；

(2) 东方商店一次性购进 32 份或 33 份，以两天内该产品所获得的利润均值为决策依据，判断哪种方案得到的利润更大．

【解答】 (1) $\xi$ 的可能取值为 30,31,32,33,34,35,36，其中 $P(\xi=30)=0.2\times0.2=0.04$，

$P(\xi=31)=2\times0.2\times0.3=0.12$，

$P(\xi=32)=0.3\times0.3+2\times0.2\times0.4=0.25$，

$P(\xi=33)=2\times0.2\times0.1+2\times0.3\times0.4=0.28$，

$P(\xi=34)=0.4\times0.4+2\times0.3\times0.1=0.22$，

$P(\xi=35)=2\times0.4\times0.1=0.08$，

$P(\xi=36)=0.1\times 0.1=0.01$，

所以 $\xi$ 的分布列为

| $\xi$ | 30 | 31 | 32 | 33 | 34 | 35 | 36 |
|---|---|---|---|---|---|---|---|
| $P$ | 0.04 | 0.12 | 0.25 | 0.28 | 0.22 | 0.08 | 0.01 |

所以 $E(\xi)=30\times 0.04+31\times 0.12+32\times 0.25+33\times 0.28+34\times 0.22+35\times 0.08+36\times 0.01=32.8$.

(2) 当一次性购进 32 份食品时，设每两天的利润为 $X$，则 $X$ 的可能取值有 104,116,128，且 $P(X=104)=0.04$，$P(X=116)=0.12$，$P(X=128)=1-0.04-0.12=0.84$，

所以 $E(X)=104\times 0.04+116\times 0.12+128\times 0.84=125.6$.

当一次性购进 33 份食品时，设每两天的利润为 $Y$，则 $Y$ 的可能取值有 96,108,120,132，且 $P(Y=96)=0.04$，$P(Y=108)=0.12$，$P(Y=120)=0.25$，$P(Y=132)=1-0.04-0.12-0.25=0.59$，

所以 $E(Y)=96\times 0.04+108\times 0.12+120\times 0.25+132\times 0.59=124.68$.

因为 $E(X)>E(Y)$，

所以东方商店一次性购进 32 份食品时得到的利润更大.

【规律总结】 离散型随机变量的均值与方差的常见类型及解题策略：(1) 求离散型随机变量的均值与方差，可依题设条件求出离散型随机变量的分布列，然后利用均值、方差公式直接求解.

(2) 由已知均值或方差求参数值，可依据条件利用均值、方差公式得出含有参数的方程(组)，解方程(组)即可求出参数值.

(3) 由已知条件，作出对两种方案的判断．可依据均值、方差的意义，对实际问题作出判断.

## 7.4 二项分布与超几何分布

### 第 1 课时 二项分布

预习评价▸新知初探

**学习目标**

1. 理解 $n$ 次伯努利试验的模型.

2. 理解二项分布.

3. 能利用伯努利试验的模型及二项分布解决一些简单的实际问题.

## 问题导引

1. 独立重复试验满足什么条件?

2. 二项分布是如何定义的?

## 现学现用

1. 概念辨析(正确的打“√”，错误的打“×”)

(1) 某同学投篮的命中率为 0.6，他 10 次投篮中命中的次数 $X$ 是一个随机变量，且 $X\sim B(10,0.6)$. ( √ )

(2) 某福彩的中奖概率为 $p$，某人一次买了 8 张，中奖张数 $X$ 是一个随机变量，且 $X\sim B(8，p)$. ( √ )

(3) 二项分布是一个概率分布列，是一个用公式 $P(X=k)=C_n^k p^k(1-p)^{n-k}$，$k=0,1,2,\cdots,n$ 表示的概率分布列，它表示了 $n$ 次独立重复试验中事件 $A$ 发生次数的概率分布. ( √ )

2. 打靶时，某人每打 10 发可中靶 8 次，则他打 100 发子弹有 4 发中靶的概率为( A )

A. $C_{100}^4 0.8^4\times0.2^{96}$　　B. $0.8^4$

C. $0.8^4\times0.2^{96}$　　D. $0.2^4\times0.2^{96}$

**【解析】** 由题意可知中靶的概率为 0.8，故打 100 发子弹有 4 发中靶的概率为 $C_{100}^4 0.8^4\times0.2^{96}$.

3. 若随机变量 $X\sim B\left(6，\frac{1}{3}\right)$，则 $P(X=2)=\frac{80}{243}$.

**【解析】** $P(X=2)=C_6^2\left(1-\frac{1}{3}\right)^4\times\left(\frac{1}{3}\right)^2=\frac{80}{243}$.

## 对点练 ▶ 先掌握基础

### 类型 1　$n$ 次伯努利试验的概率问题

**例1** 甲、乙两人各进行 3 次射击，甲每次击中目标的概率为$\frac{1}{2}$，乙每次击中目标的概率为$\frac{2}{3}$.

(1) 求甲恰好击中目标 2 次的概率；

(2) 求乙至少击中目标 2 次的概率；

(3) 求乙恰好比甲多击中目标 2 次的概率．

【解答】 设 $A$＝“甲击中目标”，则 $P(A)=\frac{1}{2}$；设 $B$＝“乙击中目标”，则 $P(B)=\frac{2}{3}$.用 $X$ 表示事件 $A$ 发生的次数，$Y$ 表示事件 $B$ 发生的次数，则 $X\sim B\left(3,\frac{1}{2}\right)$，$Y\sim B\left(3,\frac{2}{3}\right)$.

(1) 甲恰好击中目标 2 次等价于 $X=2$，则 $P(X=2)=C_3^2\left(\frac{1}{2}\right)^3=\frac{3}{8}$.

(2) 乙至少击中 2 次等价于 $Y=2$ 或 $Y=3$，于是 $P(Y=2$ 或 $Y=3)=P(Y=2)+P(Y=3)=C_3^2\left(\frac{2}{3}\right)^2\left(\frac{1}{3}\right)+C_3^3\left(\frac{2}{3}\right)^3=\frac{20}{27}$.

(3) 设乙恰好比甲多击中目标 2 次为事件 $C$，乙恰好击中目标 2 次且甲恰好击中目标 0 次为事件 $C_1$，乙恰好击中目标 3 次且甲恰好击中目标 1 次为事件 $C_2$，则 $C=C_1+C_2$，$C_1$，$C_2$ 为互斥事件．

$$P(C)=P(C_1)+P(C_2)=C_3^2\left(\frac{2}{3}\right)^2\frac{1}{3}\times C_3^0\left(\frac{1}{2}\right)^3+C_3^3\left(\frac{2}{3}\right)^3\times C_3^1\left(\frac{1}{2}\right)^3=\frac{1}{18}+\frac{1}{9}=\frac{1}{6}.$$

所以乙恰好比甲多击中目标 2 次的概率为$\frac{1}{6}$.

类型 2　二项分布的分布列

例2　一名学生每天骑自行车上学，从家到学校的途中有 5 个交通岗，假设他在各交通岗遇到红灯的事件是相互独立的，且概率都是$\frac{1}{3}$.

(1) 求这名学生在途中遇到红灯的次数 $\xi$ 的分布列；

(2) 求这名学生在首次遇到红灯或到达目的地停车前经过的路口数 $\eta$ 的分布列．

【解答】 (1) 由题意 $\xi\sim B\left(5,\frac{1}{3}\right)$，则 $P(\xi=k)=C_5^k\left(\frac{1}{3}\right)^k\left(\frac{2}{3}\right)^{5-k}$，$k=0,1,2,3,4,5$.

$P(\xi=0)=C_5^0\left(\frac{1}{3}\right)^0\left(\frac{2}{3}\right)^5=\frac{32}{243}$，

$P(\xi=1)=C_5^1\left(\frac{1}{3}\right)\left(\frac{2}{3}\right)^4=\frac{80}{243}$,

$P(\xi=2)=C_5^2\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^3=\frac{80}{243}$,

$P(\xi=3)=C_5^3\left(\frac{1}{3}\right)^3\left(\frac{2}{3}\right)^2=\frac{40}{243}$,

$P(\xi=4)=C_5^4\left(\frac{1}{3}\right)^4\left(\frac{2}{3}\right)=\frac{10}{243}$,

$P(\xi=5)=C_5^5\left(\frac{1}{3}\right)^5=\frac{1}{243}$,

所以 $\xi$ 的分布列为

| $\xi$ | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| $P$ | $\frac{32}{243}$ | $\frac{80}{243}$ | $\frac{80}{243}$ | $\frac{40}{243}$ | $\frac{10}{243}$ | $\frac{1}{243}$ |

(2) $\eta$ 的分布列为 $P(\eta=k)=P$(前 $k$ 个是绿灯，第 $k+1$ 个是红灯)$=\left(\frac{2}{3}\right)^k\times\frac{1}{3}$，$k=0,1,2,3,4$，

$P(\eta=5)=P$(5 个均为绿灯)$=\left(\frac{2}{3}\right)^5$.

故 $\eta$ 的分布列为

| $\eta$ | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| $P$ | $\frac{1}{3}$ | $\frac{2}{9}$ | $\frac{4}{27}$ | $\frac{8}{81}$ | $\frac{16}{243}$ | $\frac{32}{243}$ |

【规律总结】 (1) 判断一个随机变量是否服从二项分布，关键有两点：一是对立性，即一次试验中，事件发生与否两者必有其一；二是重复性，即试验是独立重复地进行了 $n$ 次.

(2) 对于公式 $P(X=k)=C_n^kp^k(1-p)^{n-k}(k=0,1,2,\cdots,n)$必须在满足“独立重复试验”时才能运用，否则不能应用该公式.

类型 3　二项分布的应用

例3　某车间有 10 台同类型的机床，每台机床配备的电动机功率为 10 kW，已知每台机床工作时，平均每小时实际开动 12 min，且开动与否是相互独立的．现

因当地电力供应紧张，供电部门只提供 50 kW 的电力.

(1) 求这 10 台机床能够正常工作的概率；

(2) 在一个工作班的 8 h 内，不能正常工作的时间大约是多少？

【解答】每台机床正常工作的概率为$\frac{12}{60}=\frac{1}{5}$,而且每台机床分“工作”和“不工作”两种情况,

所以工作机床台数 $\xi\sim B\left(10,\frac{1}{5}\right)$,

$P(\xi=k)=C_{10}^{k}\left(\frac{1}{5}\right)^{k}\left(\frac{4}{5}\right)^{10-k}(k=0,1,2,3,\cdots,10)$.

(1) 因为 50 kW 电力可以同时供给 5 台机床开动，所以 10 台机床同时开动的台数不超过 5 台时都可以正常工作．这一事件的概率为 $P(\xi\leqslant5)$，且 $P(\xi\leqslant5)=C_{10}^{0}\left(\frac{4}{5}\right)^{10}+C_{10}^{1}\times\frac{1}{5}\times\left(\frac{4}{5}\right)^{9}+C_{10}^{2}\left(\frac{1}{5}\right)^{2}\times\left(\frac{4}{5}\right)^{8}+C_{10}^{3}\left(\frac{1}{5}\right)^{3}\left(\frac{4}{5}\right)^{7}+C_{10}^{4}\left(\frac{1}{5}\right)^{4}\times\left(\frac{4}{5}\right)^{6}+C_{10}^{5}\left(\frac{1}{5}\right)^{5}\times\left(\frac{4}{5}\right)^{5}\approx0.994$.

(2) 在电力供应为 50 kW 的条件下，机床不能正常工作的概率仅约为 0.006，从而在一个工作班的 8 h 内，不能正常工作的时间大约有 $8\times60\times0.006=2.88$ (min).

**变式** 实力相等的甲、乙两队参加乒乓球团体比赛，规定 5 局 3 胜制(即 5 局内谁先赢 3 局就算胜出并停止比赛).

(1) 试分别求甲打完 3 局、4 局、5 局才能取胜的概率；

(2) 求按比赛规则甲获胜的概率.

【解答】 甲、乙两队实力相等，所以每局比赛甲获胜的概率为$\frac{1}{2}$，乙获胜的概率为$\frac{1}{2}$.

(1) 记事件 $A=$“甲打完 3 局才能取胜”，记事件 $B=$“甲打完 4 局才能取胜”，记事件 $C=$“甲打完 5 局才能取胜”.

①甲打完 3 局才能取胜，相当于进行 3 次独立重复试验，且每局比赛甲均取

胜，所以 $P(A)=C_3^3\left(\frac{1}{2}\right)^3=\frac{1}{8}$.

②甲打完 4 局才能取胜，相当于进行 4 次独立重复试验，且甲第 4 局比赛取胜，前 3 局为 2 胜 1 负，所以 $P(B)=C_3^2\times\left(\frac{1}{2}\right)^2\times\frac{1}{2}\times\frac{1}{2}=\frac{3}{16}$.

③甲打完 5 局才能取胜，相当于进行 5 次独立重复试验，且甲第 5 局比赛取胜，前 4 局恰好 2 胜 2 负，所以 $P(C)=C_4^2\times\left(\frac{1}{2}\right)^2\times\left(\frac{1}{2}\right)^2\times\frac{1}{2}=\frac{3}{16}$.

(2) 记事件 $D$ = “按比赛规则甲获胜”，则 $D=A+B+C$，

又 $A$，$B$，$C$ 彼此互斥，故 $P(D)=P(A+B+C)=P(A)+P(B)+P(C)=\frac{1}{8}+\frac{3}{16}+\frac{3}{16}=\frac{1}{2}$，即按比赛规则甲获胜的概率为$\frac{1}{2}$.

**【规律总结】** 二项分布实际应用问题的解题思路：

(1) 根据题意设出随机变量，

(2) 分析出随机变量服从二项分布，

(3) 找到参数 $n$(试验的次数)和 $p$(事件发生的概率)，

(4) 写出二项分布的分布列．

课堂评价▸基础达标

1．设随机变量 $\xi\sim B\left(6，\frac{1}{2}\right)$，则 $P(\xi\leqslant 3)$等于(　C　)

A．$\frac{11}{32}$　　B．$\frac{7}{32}$

C．$\frac{21}{32}$　　D．$\frac{7}{64}$

**【解析】** $P(\xi\leqslant 3)=P(\xi=0)+P(\xi=1)+P(\xi=2)+P(\xi=3)=C_6^0\times\left(\frac{1}{2}\right)^6+C_6^1\times\left(\frac{1}{2}\right)^6+C_6^2\times\left(\frac{1}{2}\right)^6+C_6^3\times\left(\frac{1}{2}\right)^6=\frac{21}{32}$.

2．若 $X\sim B\left(4，\frac{1}{2}\right)$，则 $E(2X+1)$等于(　D　)

A. 2　　　　　　　　　　B. 3

C. 4　　　　　　　　　　D. 5

【解析】 因为 $X\sim B\left(4,\ \frac{1}{2}\right)$，所以 $E(X)=4\times\frac{1}{2}=2$，所以 $E(2X+1)=2E(X)+1=5$.

3. 设 $\xi\sim B(18,p)$，$E(\xi)=9$，则 $p$ 的值为( A )

A. $\frac{1}{2}$　　　　　　　　　　B. $\frac{1}{3}$

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

【解析】 因为 $\xi\sim B(18,p)$，$E(\xi)=np=9$，所以 $18p=9$，所以 $p=\frac{1}{2}$.

4. 在某次试验中，事件 $A$ 出现的概率为 $p$，则在 $n$ 次独立重复试验中 $\overline{A}$ 出现 $k$ 次的概率为( D )

A. $1-p^k$　　　　　　　　　　B. $(1-p)^kp^{n-k}$

C. $1-(1-p)^k$　　　　　　　　　　D. $C_n^k(1-p)^kp^{n-k}$

【解析】 $\overline{A}$ 出现 1 次的概率为 $1-p$，由二项分布的概率公式可得 $\overline{A}$ 出现 $k$ 次的概率为 $C_n^k(1-p)^kp^{n-k}$.

5. 某气象站天气预报的准确率为 80%，则“5 次预报中恰有 2 次准确”的概率为 0.05；“5 次预报中至少有 2 次准确”的概率为 0.99.(结果保留到小数点后第 2 位)

【解析】 记“预报一次准确”为事件 $A$，则 $P(A)=0.8$.用 $X$ 表示事件 $A$ 发生的次数，则 $X\sim B(5,0.8)$．“恰有 2 次准确”等价于 $X=2$，于是 $P(X=2)=C_5^2\times0.8^2\times0.2^3=0.051\ 2\approx0.05$，因此 5 次预报中恰有 2 次准确的概率约为 0.05.“5 次预报中至少有 2 次准确”的对立事件为“5 次预报全部不准确或只有 1 次准确”，而“5 次预报全部不准确或只有 1 次准确”等价于 $X=0$ 或 $X=1$.因为 $P(X=0$ 或 $X=1)=P(X=0)+P(X=1)=C_5^0\times0.2^5+C_5^1\times0.8\times0.2^4=0.006\ 72$，所以所求

概率为 1 - 0.006 72≈0.99.即“5 次预报中至少有 2 次准确”的概率约为 0.99.

## 第 2 课时　超几何分布

预习评价▶新知初探

**学习目标**

1. 理解超几何分布及其推导过程.

2. 理解超几何分布与二项分布的区别与联系.

3. 能运用超几何分布解决一些实际问题.

**问题导引**

1. 超几何分布是如何定义的?

2. 若 $X \sim H(N, n, M)$，且 $n+M-N \leqslant 0$，则 $X$ 能取所有不大于 $s$ 的自然数，此时 $X$ 的分布列是怎样的?

**现学现用**

1. 概念辨析(正确的打“√”，错误的打“×”)

(1) 在产品检验中，超几何分布描述的是放回抽样. ( × )

(2) 从 4 名男演员和 3 名女演员中选出 4 名，其中女演员的人数 $X$ 服从超几何分布. ( √ )

(3) 在超几何分布中，只要知道 $N$，$M$ 和 $n$，就可以根据公式，求出 $X$ 取不同值 $m$ 时的概率 $P(X=m)$. ( √ )

2. (多选)有 6 个大小相同的黑球，编号为 1,2,3,4,5,6，还有 4 个同样大小的白球，编号为 7,8,9,10，现从中任取 4 个球，则下面服从超几何分布的是( CD )

A. $X$ 表示取出的最大号码

B. $Y$ 表示取出的最小号码

C. 取出一个黑球记 2 分，取出一个白球记 1 分，$\xi$ 表示取出的 4 个球的总得分

D. $\eta$ 表示取出的黑球个数

3. 带活动门的小盒子里有来自同一巢的 20 只工蜂和 10 只雄蜂，现随机地放出 5 只做试验，$X$ 表示放出的蜂中工蜂的只数，则 $X=2$ 时的概率是( B )

A. $\dfrac{C_{20}^{1}C_{10}^{4}}{C_{30}^{5}}$　　B. $\dfrac{C_{20}^{2}C_{10}^{3}}{C_{30}^{5}}$

C. $\dfrac{C_{20}^{3}C_{10}^{2}}{C_{30}^{5}}$　　　　D. $\dfrac{C_{20}^{4}C_{10}^{1}}{C_{30}^{5}}$

【解析】 依题意，$X$服从超几何分布，所以$P(X=2)=\dfrac{C_{20}^{2}C_{10}^{3}}{C_{30}^{5}}$.

4. 从3台甲型彩电和2台乙型彩电中任取2台，设$X$表示所取的2台彩电中甲型彩电的台数，则$P(X=1)=\underline{\dfrac{3}{5}}$.

【解析】 $X=1$表示的结果是抽取的2台彩电有甲型和乙型彩电各一台，故所求概率$P(X=1)=\dfrac{C_{3}^{1}C_{2}^{1}}{C_{5}^{2}}=\dfrac{3}{5}$.

## 对点练▶先掌握基础

类型1　利用超几何分布求概率

例1　袋中有8个球，其中5个黑球，3个红球，从袋中任取3个球，求取出的红球数$X$的分布列，并求至少有一个红球的概率.

【解答】 由已知可得$X$的取值为0,1,2,3，$X=0$表示取出的3个球全是黑球，$P(X=0)=\dfrac{C_{5}^{3}}{C_{8}^{3}}=\dfrac{10}{56}=\dfrac{5}{28}$，同理$P(X=1)=\dfrac{C_{3}^{1}C_{5}^{2}}{C_{8}^{3}}=\dfrac{30}{56}=\dfrac{15}{28}$，$P(X=2)=\dfrac{C_{3}^{2}C_{5}^{1}}{C_{8}^{3}}=\dfrac{15}{56}$，$P(X=3)=\dfrac{C_{3}^{3}}{C_{8}^{3}}=\dfrac{1}{56}$.所以$X$的分布列为

| $X$ | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| $P$ | $\dfrac{5}{28}$ | $\dfrac{15}{28}$ | $\dfrac{15}{56}$ | $\dfrac{1}{56}$ |

至少有一个红球的概率为$P(X\geqslant 1)=1-\dfrac{5}{28}=\dfrac{23}{28}$.

【规律总结】 超几何分布的求解步骤：

(1) 辨模型：结合实际情景分析所求概率分布问题是否具有明显的两部分组成，如“男生、女生”，“正品、次品”，“优、劣”等，或可转化为明显的两部分．具有该特征的概率模型为超几何分布模型．

(2) 算概率：可以直接借助公式$P(X=k)=\dfrac{C_{M}^{k}C_{N-M}^{n-k}}{C_{N}^{n}}$求解，也可以利用排列组

合及概率的知识求解，需注意借助公式求解时应理解参数 $M$，$N$，$n$，$k$ 的含义．

(3) 列分布表：把求得的概率值通过表格表示出来．

类型 2　两点分布和超几何分布

例2　在一次购物抽奖活动中，假设 10 张奖券中有一等奖奖券 1 张，可获价值 50 元的奖品，有二等奖奖券 3 张，每张可获价值 10 元的奖品，其余 6 张没有奖品．

(1) 顾客甲从 10 张奖券中任意抽取 1 张，求中奖次数 $X$ 的分布列．

(2) 顾客乙从 10 张奖券中任意抽取 2 张．

①求顾客乙中奖的概率；

②设顾客乙获得的奖品总价值为 $Y$ 元，求 $Y$ 的分布列．

【解答】 (1) 抽奖 1 次，只有中奖和不中奖两种情况，故 $X$ 的取值只有 0 和 1 两种情况，

$P(X=1)=\frac{C_4^1}{C_{10}^1}=\frac{4}{10}=\frac{2}{5}$，

则 $P(X=0)=1-P(X=1)=1-\frac{2}{5}=\frac{3}{5}$.

因此 $X$ 的分布列为

| $X$ | 0 | 1 |
| --- | --- | --- |
| $P$ | $\frac{3}{5}$ | $\frac{2}{5}$ |

(2) ①顾客乙中奖可分为：所抽取的 2 张奖券中有 1 张中奖或 2 张都中奖，故所求概率 $P=\frac{C_4^1C_6^1+C_4^2C_6^0}{C_{10}^2}=\frac{30}{45}=\frac{2}{3}$.

②$Y$ 的所有可能取值为 0,10,20,50,60，且

$P(Y=0)=\frac{C_4^0C_6^2}{C_{10}^2}=\frac{15}{45}=\frac{1}{3}$，

$P(Y=10)=\frac{C_3^1C_6^1}{C_{10}^2}=\frac{18}{45}=\frac{2}{5}$，

$P(Y=20)=\frac{C_3^2C_6^0}{C_{10}^2}=\frac{3}{45}=\frac{1}{15}$，

$P(Y=50)=\frac{C_1^1C_6^1}{C_{10}^2}=\frac{6}{45}=\frac{2}{15}$,

$P(Y=60)=\frac{C_1^1C_3^1}{C_{10}^2}=\frac{3}{45}=\frac{1}{15}$.

因此随机变量 $Y$ 的分布列为

| $Y$ | 0 | 10 | 20 | 50 | 60 |
|---|---|---|---|---|---|
| $P$ | $\frac{1}{3}$ | $\frac{2}{5}$ | $\frac{1}{15}$ | $\frac{2}{15}$ | $\frac{1}{15}$ |

【规律总结】 解决超几何分布问题的两个关键点:

(1) 超几何分布是概率分布的一种形式，一定要注意公式中字母的范围及其意义，解决问题时可以直接利用公式求解，但不能机械地记忆.

(2) 超几何分布中，只要知道 $M$，$N$，$n$，就可以利用公式求出 $X$ 取不同值 $k$ 的概率 $P(X=k)$，从而求出 $X$ 的分布列.

类型 3 超几何分布与二项分布

例3 袋中有 8 个白球、2 个黑球，从中随机地连续抽取 3 次，每次取 1 个球，用 $X$ 表示取出黑球的个数.

(1) 无放回抽样时，求取到黑球的个数 $X$ 的分布列；

(2) 有放回抽样时，求取到黑球的个数 $X$ 的分布列.

【解答】 (1) 无放回抽样时，各次试验结果不独立，其中 $P(X=k)=\frac{C_2^kC_8^{3-k}}{C_{10}^3}$，$k=0,1,2$，所以 $X$ 的分布列为

| $X$ | 0 | 1 | 2 |
|---|---|---|---|
| $P$ | $\frac{7}{15}$ | $\frac{7}{15}$ | $\frac{1}{15}$ |

(2) 有放回抽样时，取到黑球数 $X$ 可能的取值为 0,1,2,3，又每次取到黑球的概率均为$\frac{1}{5}$，3 次取球可看成 3 次独立重复试验，则 $X\sim B\left(3,\frac{1}{5}\right)$，其中 $P(X=k)=C_3^k\times\left(\frac{1}{5}\right)^k\times\left(\frac{4}{5}\right)^{3-k}$，$k=0,1,2$，故 $X$ 的分布列为

| $X$ | 0 | 1 | 2 | 3 |
|---|---|---|---|---|

| | | | | |
|---|---|---|---|---|
| $P$ | $\frac{64}{125}$ | $\frac{48}{125}$ | $\frac{12}{125}$ | $\frac{1}{125}$ |

**变式** 老师要从10篇课文中随机抽3篇让学生背诵，某同学只能背诵其中的6篇.

(1) 抽到他能背诵的课文的数量的分布列；

(2) 若老师从10篇课文中有放回抽取3篇，求抽到他能背诵的课文的数量的分布列.

【解答】(1) 设抽到他能背诵的课文的数量为$X$，则$X$服从超几何分布，$P(X=k)=\frac{C_6^kC_4^{3-k}}{C_{10}^3}(k=0,1,2,3)$.

故$X$的分布列为

| $X$ | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| $P$ | $\frac{1}{30}$ | $\frac{3}{10}$ | $\frac{1}{2}$ | $\frac{1}{6}$ |

(2) 老师抽取一篇抽到该生能背诵的概率为$p=\frac{6}{10}=\frac{3}{5}$，设抽到他能背诵的课文的数量为$X$，则$X\sim B\left(3，\frac{3}{5}\right)$，$P(X=k)=C_3^k\times\left(\frac{3}{5}\right)^k\times\left(\frac{2}{5}\right)^{3-k}$，$k=0,1,2,3$，

故$X$的分布列为

| $X$ | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| $P$ | $\frac{8}{125}$ | $\frac{36}{125}$ | $\frac{54}{125}$ | $\frac{27}{125}$ |

课堂评价▶基础达标

1. 盒中有 10 个螺丝钉，其中 3 个是坏的，现从盒中随机地抽取 4 个，则概率是$\frac{3}{10}$的事件为(　C　)

A. 恰有 1 个是坏的

B. 4 个全是好的

C. 恰有 2 个是好的

D. 至多有 2 个是坏的

**【解析】** 对于 A，概率为$\frac{C_3^1C_7^3}{C_{10}^4}=\frac{1}{2}$.对于 B，概率为$\frac{C_7^4}{C_{10}^4}=\frac{1}{6}$.对于 C，概率为$\frac{C_3^2C_7^2}{C_{10}^4}=\frac{3}{10}$.对于 D，包括没有坏的，有 1 个坏的和 2 个坏的三种情况．根据 A 选项，恰好有一个坏的概率已经是$\frac{1}{2}>\frac{3}{10}$，故 D 选项不正确．

2. 已知在 10 件产品中可能存在次品，从中抽取 2 件检查，其次品数为 $\xi$，已知 $P(\xi=1)=\frac{16}{45}$，且该产品的次品率不超过 40%，则这 10 件产品的次品率为(　B　)

A. 10%　　　　B. 20%

C. 30%　　　　D. 40%

**【解析】** 设 10 件产品中有 $x$ 件次品，则 $P(\xi=1)=\frac{C_x^1C_{10-x}^1}{C_{10}^2}=\frac{x(10-x)}{45}=\frac{16}{45}$，所以 $x=2$ 或 $x=8$.因为该产品的次品率不超过 40%，所以 $x=2$，所以这 10 件产品的次品率为$\frac{2}{10}=20\%$.

3. 在 15 个村庄中，有 7 个村庄交通不方便，若用随机变量 $X$ 表示任选 10 个村庄中交通不方便的村庄的个数，则 $X$ 服从超几何分布，其参数为(　A　)

A. $N=15$，$M=7$，$n=10$

B. $N=15$，$M=10$，$n=7$

C. $N=22$，$M=10$，$n=7$

D. $N=22$，$M=7$，$n=10$

4. 袋中有 10 个球，其中 7 个红球，3 个白球，任意取出 3 个，这 3 个都是红球的概率是(　B　)

A. $\frac{1}{120}$　　B. $\frac{7}{24}$

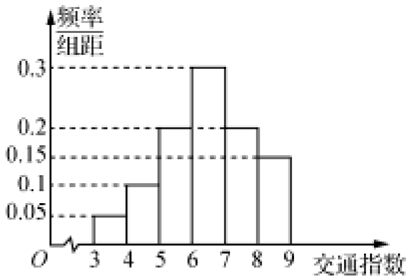
C. $\frac{7}{10}$　　D. $\frac{3}{7}$

【解析】 设“3 个球都是红球”为事件 $A$，则 $P(A)=\frac{C_7^3C_3^0}{C_{10}^3}=\frac{7}{24}$.

5. 交通指数是交通拥堵指数的简称，是综合反映道路网畅通或拥堵的概念. 记交通指数为 $T$，其范围为[0,10]，分别有五个级别：$T\in[0,2)$畅通；$T\in[2,4)$基本畅通；$T\in[4,6)$轻度拥堵；$T\in[6,8)$中度拥堵；$T\in[8,10]$严重拥堵. 晚高峰时段，从某市交通指挥中心选取了市区 20 个交通路段，依据其交通指数数据绘制的频率分布直方图如图所示.

(1) 这 20 个路段轻度拥堵、中度拥堵的路段各有多少个？

(2) 从这 20 个路段中随机抽出 3 个路段，用 $X$ 表示抽取的中度拥堵的路段的个数，求 $X$ 的分布列.




【解答】 (1) 由直方图得，轻度拥堵的路段个数是$(0.1+0.2)\times1\times20=6$，

中度拥堵的路段个数是$(0.3+0.2)\times1\times20=10$.

(2) $X$ 的可能取值为 0,1,2,3，

则 $P(X=0)=\frac{C_{10}^0C_{10}^3}{C_{20}^3}=\frac{2}{19}$，$P(X=1)=\frac{C_{10}^1C_{10}^2}{C_{20}^3}=\frac{15}{38}$，$P(X=2)=\frac{C_{10}^2C_{10}^1}{C_{20}^3}=\frac{15}{38}$，$P(X=3)=\frac{C_{10}^3C_{10}^0}{C_{20}^3}=\frac{2}{19}$.所以 $X$ 的分布列为

| $X$ | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| $P$ | $\frac{2}{19}$ | $\frac{15}{38}$ | $\frac{15}{38}$ | $\frac{2}{19}$ |

## 7.5　正态分布

预习评价▶新知初探

### 学习目标

1. 了解正态曲线和正态分布的意义.

2. 明确正态分布中参数$\mu$，$\sigma$的意义及其对正态曲线的影响.

3. 了解$3\sigma$原则，会用正态分布解决实际问题.

## 问题导引

1. 已知$X$服从参数为100,0.5的二项分布，即$X\sim B(100,0.5)$，你能计算出$P(X=50)$的值吗？

2. 举出现实生活中，服从或近似服从正态分布的随机变量.

## 现学现用

1. 概念辨析(正确的打“√”，错误的打“×”)

(1) 正态变量函数表述式中参数$\mu$，$\sigma$的意义分别是样本的均值与方差. ( × )

(2) 正态曲线是单峰的，其与$x$轴围成的面积是随参数$\mu$，$\sigma$的变化而变化的. ( × )

(3) 当$\mu$一定时，曲线的形状由$\sigma$确定，$\sigma$越小，曲线越“瘦高”，$\sigma$越大，曲线越“矮胖”. ( √ )

2. (多选)把一条正态曲线$a$沿着横轴方向向右移动2个单位，得到一条新的曲线$b$，下列说法中正确的是( ABD )

A. 曲线$b$仍然是正态曲线

B. 曲线$a$和曲线$b$的最高点的纵坐标相等

C. 以曲线$b$为正态分布的总体的方差比以曲线$a$为正态分布的总体的方差大2

D. 以曲线$b$为正态分布的总体的均值比以曲线$a$为正态分布的总体的均值大2

【解析】 正态曲线向右平移2个单位，$\sigma$不发生变化，故C错误，其他正确.

3. 若随机变量$X\sim N(2,\sigma^2)$，则$P(X<2)$等于( D )

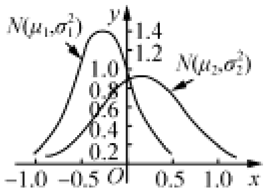
A. $\frac{1}{5}$　　B. $\frac{1}{4}$

C. $\frac{1}{3}$　　D. $\frac{1}{2}$

**对点练 ▶ 先掌握基础**

类型 1　正态曲线的图象及其应用

**例1**　(1) 设两个正态分布 $N(\mu_1, \sigma_1^2)(\sigma_1>0)$ 和 $N(\mu_2, \sigma_2^2)(\sigma_2>0)$ 的密度函数图象如图(1)所示，则(　A　)




(例 1(1))

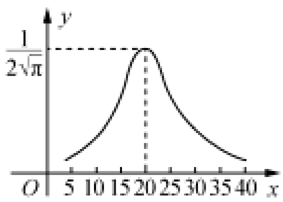
A. $\mu_1<\mu_2$，$\sigma_1<\sigma_2$

B. $\mu_1<\mu_2$，$\sigma_1>\sigma_2$

C. $\mu_1>\mu_2$，$\sigma_1<\sigma_2$

D. $\mu_1>\mu_2$，$\sigma_1>\sigma_2$

(2) 如图(2)所示是一个正态曲线，试根据该图象写出其正态分布的概率密度函数的解析式，求出总体随机变量的均值和方差.




(例 1(2))

**【解答】** 从正态曲线可知，该正态曲线关于直线 $x=20$ 对称，最大值为 $\frac{1}{2\sqrt{\pi}}$，所以 $\mu=20$，$\frac{1}{\sqrt{2\pi}\sigma}=\frac{1}{2\sqrt{\pi}}$，所以 $\sigma=\sqrt{2}$.于是 $f(x)=\frac{1}{2\sqrt{\pi}}e^{\frac{(x-20)^2}{4}}$，$x\in(-\infty, +\infty)$，总体随机变量的均值是 $\mu=20$，方差是 $\sigma^2=(\sqrt{2})^2=2$.

类型 2　服从正态分布变量的概率问题

**例2**　(1) 已知随机变量 $\xi$ 服从正态分布 $N(2, \sigma^2)$，且 $P(\xi<4)=0.8$，则 $P(0<\xi<2)$ 等于(　C　)

A. 0.6　　　　B. 0.4

C. 0.3　　　　D. 0.2

**【解析】** 因为随机变量 $X$ 服从正态分布 $N(2, \sigma^2)$，所以 $\mu=2$，对称轴是 $x=2$.因为 $P(\xi<4)=0.8$，所以 $P(\xi\geqslant 4)=P(\xi<0)=0.2$，所以 $P(0<\xi<4)=0.6$，所以

$P(0<\xi<2)=0.3$.

(2) 在某项测量中，测量结果服从正态分布 $N(1,4)$，则正态总体 $X$ 在$(-1,1)$内取值的概率为 0.341_3.

【解析】 由题意得 $\mu=1,\sigma=2$，所以 $P(-1<X<3)=P(1-2<X<1+2)=0.6826$.因为正态曲线关于 $x=1$ 对称，所以 $P(-1<X<1)=P(1<X<3)=\frac{1}{2}P(-1<X<3)=0.341\ 3$.

【规律总结】 利用正态分布求概率的两个方法：

(1) 对称法：由于正态曲线是关于直线 $x=\mu$ 对称的，且概率的和为 1，故关于直线 $x=\mu$ 对称的区间上概率相等．如：①$P(X<a)=1-P(X\geqslant a)$；②$P(X<\mu-a)=P(X>\mu+a)$．

(2) “$3\sigma$”法：利用 $X$ 落在区间$[\mu-\sigma,\ \mu+\sigma]$，$[\mu-2\sigma,\ \mu+2\sigma]$，$[\mu-3\sigma,\ \mu+3\sigma]$内的概率分别是 68.3%,95.4%,99.7%求解．

类型 3 正态分布的实际应用

例3 在某次数学考试中，考生的成绩 $X$ 服从正态分布，即 $X\sim N(90,100)$.

(1) 求考试成绩 $X$ 位于区间[70,110]内的概率．

(2) 若这次考试共有 2 000 名考生，试估计考试成绩在[80,100]之间的考生人数．

【解答】 由 $X\sim N(90,100)$，

则 $\mu=90$，$\sigma=\sqrt{100}=10$.

(1) $P(70\leqslant X\leqslant 110)=P(90-2\times 10\leqslant X\leqslant 90+2\times 10)=0.954\ 4$，即成绩 $X$ 位于区间[70,110]内的概率为 0.954 4.

(2) $P(80\leqslant X\leqslant 100)=P(90-10\leqslant X\leqslant 90+10)=0.682\ 6$，所以 $2\ 000\times 0.682\ 6\approx 1\ 365$(人)．

即考试成绩在[80,100]之间的考生大约有 1 365 人．

**变式** 某城市从南郊某地乘公共汽车前往北区火车站有两条路线可走，第一条路线穿过市区，路线较短，但交通拥挤，所需时间(单位：分)服从正态分布 $N(50,10^2)$；第二条路线沿环城公路走，路程较长，但交通阻塞少，所需时间服从正态分布 $N(60,4^2)$.

(1) 若只有 70 分钟可用，应走哪条路线？

(2) 若只有 65 分钟可用，又应走哪条路线？

【解答】 由已知 $X\sim N(50,10^2)$，

$Y\sim N(60,4^2)$，

由正态分布的 $2\sigma$ 区间性质 $P(\mu-2\sigma\leqslant\xi\leqslant\mu+2\sigma)=0.954\ 4$，

对 $X$：$\mu=50$，$\sigma=10$,$2\sigma$ 区间为$(30,70)$；

对 $Y$：$\mu=60$，$\sigma=4$,$2\sigma$ 区间为$(52,68)$，

要尽量保证用时在 $X\subseteq(30,70)$，$Y\subseteq(52,68)$才能保证有 95%以上的概率准时到达.

(1) 时间只有 70 分钟可用，应该走第二条路线.

(2) 时间只有 65 分钟可用，两种方案都能保证有 95%以上的概率准时到达，但是走市区平均用时比路线二少了 10 分钟，应该走第一条路线.

【规律总结】 解答正态分布的实际应用题，其关键是如何转化，同时应熟练掌握正态分布在$[\mu-\sigma，\mu+\sigma]$，$[\mu-2\sigma，\mu+2\sigma]$，$[\mu-3\sigma，\mu+3\sigma]$三个区间内的概率．在此过程中用到归纳思想和数形结合思想.

课堂评价▶基础达标

1. 正态曲线关于 $y$ 轴对称，当且仅当它所对应的正态总体的均值为(　C　)

A. 1　　　　B. －1

C. 0　　　　D. 不确定

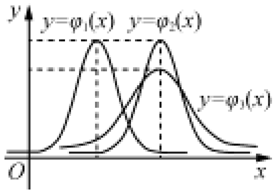
2. 已知某批零件的长度误差(单位：mm)服从正态分布 $N(0,3^2)$，从中随机取一件，其长度误差落在区间[3,6]内的概率为(　B　)

A. 0.045 6　　　　B. 0.135 9

C. 0.271 8　　　　　　　　D. 0.317 4

【解析】 $P(-3\leqslant\xi\leqslant3)=0.682\ 6$，$P(-6\leqslant\xi\leqslant6)=0.954\ 4$，则 $P(3\leqslant\xi\leqslant6)=\frac{1}{2}\times(0.954\ 4-0.682\ 6)=0.135\ 9$.

3. 若正态分布密度函数 $\varphi_i(x)=\frac{1}{\sqrt{2\pi}\sigma_i}e^{-\frac{(x-\mu_i)^2}{2\sigma_i^2}}$ $(x\in\mathbf{R},i=1,2,3)$的图象如图所示，则(　D　)




A. $\mu_1<\mu_2=\mu_3$，$\sigma_1=\sigma_2>\sigma_3$

B. $\mu_1>\mu_2=\mu_3$，$\sigma_1=\sigma_2<\sigma_3$

C. $\mu_1=\mu_2<\mu_3$，$\sigma_1<\sigma_2=\sigma_3$

D. $\mu_1<\mu_2=\mu_3$，$\sigma_1=\sigma_2<\sigma_3$

【解析】 因为正态曲线关于 $x=\mu$ 对称，且 $\mu$ 越大图象越靠近右边，所以可得 $\mu_1<\mu_2=\mu_3$.又因为 $\sigma$ 的值反映的是这组数据的集中情况，其 $\sigma$ 值越小图象越瘦长，$\sigma$ 值越大图象越矮胖，所以可得 $\sigma_1=\sigma_2<\sigma_3$.

4. 若 $\xi\sim N\left(1，\frac{1}{4}\right)$，$\eta=6\xi$，则 $E(\eta)=$__6__.

【解析】 因为 $\xi\sim N\left(1，\frac{1}{4}\right)$，所以 $E(\xi)=1$，所以 $E(\eta)=6E(\xi)=6$.

5. 在某项测量中，测量结果服从正态分布 $N(1，4)$，则 $P(X\geqslant5)=$0.022 8.

【解析】 因为 $P(X\geqslant5)=P(X\leqslant-3)$，所以 $P(X\geqslant5)=\frac{1}{2}[1-P(-3<X\leqslant5)]=\frac{1}{2}[1-P(1-4<X\leqslant1+4)]=\frac{1}{2}[1-P(\mu-2\sigma<X\leqslant\mu+2\sigma)]=\frac{1}{2}\times(1-0.954\ 4)=0.022\ 8$.

## 章复习　能力整合与素养提升

### 要点回顾▶连点成面

1. 条件概率及其性质

对于任何两个事件 $A$ 和 $B$，在已知事件 $A$ 发生的条件下，事件 $B$ 发生的概率

叫做条件概率，用符号 $P(B|A)$来表示，其公式为 $P(B|A)=\frac{P(AB)}{P(A)}(P(A)>0)$.

在古典概型中，若用 $n(A)$表示事件 $A$ 中基本事件的个数，则 $P(B|A)=\frac{n(AB)}{n(A)}$.

2. 离散型随机变量的均值与方差：若离散型随机变量 $\xi$ 的分布列为

| $\xi$ | $x_1$ | $x_2$ | … | $x_i$ | … | $x_n$ |
|---|---|---|---|---|---|---|
| $P$ | $p_1$ | $p_2$ | … | $p_i$ | … | $p_n$ |

(1) 均值：称 $E(\xi)=x_1\cdot p_1+x_2\cdot p_2+\cdots+x_n\cdot p_n$ 为随机变量 $\xi$ 的均值或数学期望，它反映了离散型随机变量取值的平均水平.

(2) 方差：称 $D(\xi)=\sum_{i=1}^{n}[x_i-E(\xi)]^2\cdot p_i$ 为随机变量 $\xi$ 的方差，它刻画了随机变量 $\xi$ 与其均值 $E(\xi)$的平均偏离程度 ，其算术平方根$\sqrt{D(\xi)}$为随机变量 $\xi$ 的标准差.

(3) 均值与方差的性质：①$\underline{E(a\xi+b)=aE(\xi)+b}$；②$\underline{D(a\xi+b)=a^2D(\xi)}$.

3. 两点分布：若随机变量 $X$ 的分布列如下(其中 $0<p<1$)，则称这样的分布列为两点分布列．若 $\xi$ 服从两点分布，则 $\underline{E(\xi)=p}$，$\underline{D(\xi)=p(1-p)}$.

| $X$ | 0 | 1 |
|---|---|---|
| $P$ | $1-p$ | $p$ |

4. 超几何分布：在含有 $M$ 件次品的 $N$ 件产品中任取 $n$ 件，其中恰有 $X$ 件次品，则事件“$X=k$”发生的概率 $P(X=K)=\frac{C_M^kC_{N-M}^{n-k}}{C_N^n}$，$k=0,1,2$，…，$m$，其中 $m=\min\{M，n\}$，且 $n\leq N$，$M\leq N$，$n$，$M$，$N\in\mathbf{N}^*$，称随机变量 $X$ 服从超几何分布.

5. 二项分布：在 $n$ 次独立重复试验中，设事件 $A$ 发生的次数为 $k$，在每次试验中事件 $A$ 发生的概率为 $p$，那么在 $n$ 次独立重复试验中，事件 $A$ 恰好发生 $k$ 次的概率为 $P(X=k)=C_n^kp^k(1-p)^{n-k}(k=0,1,2，\cdots，n)$，此时称随机变量 $X$ 服从二项分布，记作 $X\sim B(n，p)$，并称 $p$ 为成功概率．期望：$\underline{E(X)=np}$；方差：$\underline{D(X)=np(1-p)}$.

6. 正态分布

(1) 正态曲线有以下特点：

①曲线在 $x$ 轴的上方，与 $x$ 轴不相交；

②曲线是单峰的，它关于直线 $x=\mu$ 对称；

③曲线在 $x=\mu$ 处达到峰值 $\frac{1}{\sigma\sqrt{2\pi}}$ (最高点)；

④$x$ 轴与正态曲线所夹面积恒等于 1．

(2) 若 $X\sim N(\mu,\sigma^2)$，则 $E(X)=$__$\mu$__，$D(X)=$__$\sigma^2$__．

考法聚焦▶核心突破

考法 1　条件概率

例1　在 100 件产品中有 95 件合格品，5 件不合格品，现从中不放回地取两次，每次任取一件，则在第一次取到不合格品后，第二次取到不合格品的概率为$\underline{\dfrac{4}{99}}$.

【解析】 方法一(应用条件概率公式求解)：设事件 $A$ 为"第一次取到不合格品"，事件 $B$ 为"第二次取到不合格品"，则所求的概率为 $P(B|A)$．因为 $P(AB)=\dfrac{C_5^2}{C_{100}^2}=\dfrac{1}{495}$，$P(A)=\dfrac{C_5^1}{C_{100}^1}=\dfrac{1}{20}$，所以 $P(B|A)=\dfrac{P(AB)}{P(A)}=\dfrac{\frac{1}{495}}{\frac{1}{20}}=\dfrac{4}{99}$.方法二(缩小样本空间求解)：第一次取到不合格品后，也就是在第二次取之前，还有 99 件产品，其中有 4 件不合格品，因此第二次取到不合格品的概率为$\dfrac{4}{99}$.

【类题固法】

1. 已知盒中装有 3 只螺口灯泡与 7 只卡口灯泡，这些灯泡的外形与功率都相同且灯口向下放着．现需要一只卡口灯泡，电工师傅每次从中任取一只且不放回，则在他第 1 次取到的是螺口灯泡的条件下，第 2 次取到的是卡口灯泡的概率为( D )

A. $\dfrac{3}{10}$　　B. $\dfrac{2}{9}$

C. $\dfrac{7}{8}$　　D. $\dfrac{7}{9}$

【解析】 方法一：设事件 $A$ 为"第 1 次取到的是螺口灯泡"，事件 $B$ 为"第 2 次取到的是卡口灯泡"，则 $P(A)=\dfrac{3}{10}$，$P(AB)=\dfrac{3}{10}\times\dfrac{7}{9}=\dfrac{7}{30}$，$P(B|A)=\dfrac{P(AB)}{P(A)}=\dfrac{\frac{7}{30}}{\frac{3}{10}}=\dfrac{7}{9}$.方法二：第 1 次取到螺口灯泡后还剩余 9 只灯泡，其中有 7 只卡口灯

泡，故第 2 次取到卡口灯泡的概率为$\frac{C_7^1}{C_9^1}=\frac{7}{9}$.

2. 甲、乙等 4 人参加 4×100 米接力赛，在甲不跑第一棒的条件下，乙不跑第二棒的概率是(　D　)

A. $\frac{2}{9}$　　　　B. $\frac{4}{9}$

C. $\frac{2}{3}$　　　　D. $\frac{7}{9}$

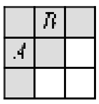
【解析】 甲不跑第一棒共有 $A_3^1 \cdot A_3^3=18$(种)情况，甲不跑第一棒且乙不跑第二棒共有两类：①乙跑第一棒，有 $A_3^3=6$(种)情况；②乙不跑第一棒，有 $A_2^1 \cdot A_2^1 \cdot A_2^2=8$(种)情况，所以在甲不跑第一棒的条件下，乙不跑第二棒的概率为$\frac{6+8}{18}=\frac{7}{9}$.

3. 一盒中放有大小相同的 10 个小球，其中 8 个黑球、2 个红球，现甲、乙两人先后各自从盒子中无放回地任意取 2 个小球，已知甲取到了 2 个黑球，则乙也取到 2 个黑球的概率是__$\frac{15}{28}$__.

【解析】 记事件“甲取到 2 个黑球”为 $A$，“乙取到 2 个黑球”为 $B$，则有 $P(B|A)=\frac{C_6^2}{C_8^2}=\frac{15}{28}$.

4. 一个正方形被平均分成 9 个部分，向大正方形区域随机地投掷一个点(每次都能投中)．设投中最左侧 3 个小正方形区域的事件记为 $A$，投中最上面 3 个小正方形或正中间的 1 个小正方形区域的事件记为 $B$，则 $P(A|B)=$__$\frac{1}{4}$__.

【解析】 如图，$n(\Omega)=9$，$n(A)=3$，$n(B)=4$，




所以 $n(AB)=1$，所以 $P(A|B)=\frac{n(AB)}{n(B)}=\frac{1}{4}$.

考法 2　均值与方差

例2　已知离散型随机变量 $X$ 满足 $P(X=x_1)=\frac{2}{3}$，$P(X=x_2)=\frac{1}{3}$，且 $x_1<x_2$，若

$E(X)=\frac{4}{3}$，$D(X)=\frac{2}{9}$，则 $x_1+x_2$ 的值为__3__.

【解析】 由题得

$$\begin{cases}x_1\times\frac{2}{3}+x_2\times\frac{1}{3}=\frac{4}{3},\\ \frac{2}{3}\left(x_1-\frac{4}{3}\right)^2+\frac{1}{3}\left(x_2-\frac{4}{3}\right)^2=\frac{2}{9},\end{cases}$$

即 $\begin{cases}2x_1+x_2=4,\\ 2\left(x_1-\frac{4}{3}\right)^2+\left(x_2-\frac{4}{3}\right)^2=\frac{2}{3},\end{cases}$

解得 $\begin{cases}x_1=\frac{5}{3},\\ x_2=\frac{2}{3},\end{cases}$ 或 $\begin{cases}x_1=1,\\ x_2=2.\end{cases}$

因为 $x_1<x_2$，所以 $\begin{cases}x_1=1,\\ x_2=2,\end{cases}$

所以 $x_1+x_2=3$.

【类题固法】

1. 现对某项目投资 10 万元，一年后利润是 1.2 万元，1.18 万元，1.17 万元的概率分别为$\frac{1}{6}$，$\frac{1}{2}$，$\frac{1}{3}$，随机变量 $X$ 表示对此项目投资 10 万元一年后的利润，则 $X$ 的均值为( A )

A. 1.18　　B. 3.55

C. 1.23　　D. 2.38

【解析】 由题意知，$X$ 的分布列为

| $X$ | 1.2 | 1.18 | 1.17 |
|---|---|---|---|
| $P$ | $\frac{1}{6}$ | $\frac{1}{2}$ | $\frac{1}{3}$ |

所以 $E(X)=1.2\times\frac{1}{6}+1.18\times\frac{1}{2}+1.17\times\frac{1}{3}=1.18$.

2. 某公司有 5 万元资金用于投资开发项目，如果成功，一年后可获利 12%；

如果失败，一年后将丧失全部资金的50%，下表是过去200例类似项目开发的实施结果，则估计该公司一年后可获收益的均值是4 760元.

| 投资成功 | 投资失败 |
| --- | --- |
| 192 例 | 8 例 |

【解析】 由题意知，一年后获利6 000元的概率为0.96，获利－25 000元的概率为0.04，故一年后收益的均值是$6\ 000\times0.96+(-25\ 000)\times0.04=4\ 760$(元).

3. 某射手射击所得环数$\xi$的分布列如下：

| $\xi$ | 7 | 8 | 9 | 10 |
| --- | --- | --- | --- | --- |
| $P$ | $x$ | 0.1 | 0.3 | $y$ |

已知$\xi$的期望$E(\xi)=8.9$，则$y$的值为0.4.

【解析】 由已知得

$$\begin{cases}x+0.1+0.3+y=1,\\7x+8\times0.1+9\times0.3+10y=8.9,\end{cases}$$

解得$y=0.4$.

考法3 二项分布

例3 某气象站天气预报的准确率为80%.

(1) 求5次预报中恰有2次准确的概率；

(2) 求5次预报中至少有2次准确的概率；

(3) 求5次预报中恰有2次准确，且其中第3次预报准确的概率.(结果保留到小数点后第2位)

【解答】 令$X$表示5次天气预报中准确的次数，则$X\sim B(5,0.8)$.

(1) “5次预报中恰有2次准确”的概率为$P(X=2)=C_5^2\times0.8^2\times(1-0.8)^3=10\times0.64\times0.008\approx0.05$.

(2) “5次预报中至少有2次准确”的概率为$P(X\geqslant2)=1-P(X=0)-P(X=1)=1-C_5^0\times0.8^0\times(1-0.8)^5-C_5^1\times0.8\times(1-0.8)^4=1-0.000\ 32-0.006\ 4\approx0.99$.

(3) “5次预报中恰有2次准确，且其中第3次预报准确”的概率为$C_4^1$

$\times0.8\times(1-0.8)^3\times0.8\approx0.02$.

【类题固法】

1. 某群体中的每位成员使用移动支付的概率都为 $p$，各成员的支付方式相互独立.设 $X$ 为该群体的 10 位成员中使用移动支付的人数，$D(X)=2.1$，$P(X=4)<P(X=6)$，则 $p$ 等于 ( A )

A. 0.7　　B. 0.6

C. 0.4　　D. 0.3

【解析】 由已知 $X\sim B(10,p)$，由 $P(X=4)<P(X=6)$，得 $C_{10}^{4}p^4(1-p)^6<C_{10}^{6}p^6(1-p)^4$，化简得 $1-2p<0$，解得 $p>\frac{1}{2}$.因为 $D(X)=2.1$，所以 $10p(1-p)=2.1$，解得 $p=0.7$ 或 $p=0.3$(舍去).

2. 位于坐标原点的一个质点 $P$ 按下述规则移动：质点每次移动一个单位，移动的方向为向上或向右，且向上、向右移动的概率都是$\frac{1}{2}$，则质点 $P$ 移动五次后位于点(2,3)的概率是$\underline{\frac{5}{16}}$.

【解析】 由于质点每次移动一个单位，移动的方向为向上或向右，移动五次后位于点(2,3)，所以质点 $P$ 必须向右移动两次，向上移动三次，故其概率为 $C_5^3\left(\frac{1}{2}\right)^3\cdot\left(\frac{1}{2}\right)^2=C_5^3\left(\frac{1}{2}\right)^5=\frac{5}{16}$.

3. 为研究家用轿车在高速公路上的车速情况，交通部门随机选取 100 名家用轿车驾驶员进行调查，得到其在高速公路上行驶时的平均车速情况为：在 55 名男性驾驶员中，平均车速超过 100 km/h 的有 40 人，不超过 100 km/h 的有 15 人；在 45 名女性驾驶员中，平均车速超过 100 km/h 的有 20 人，不超过 100 km/h 的有 25 人.

(1) 在被调查的驾驶员中，从平均车速不超过 100 km/h 的人中随机抽取 2 人，求这 2 人恰好有 1 名男性驾驶员和 1 名女性驾驶员的概率；

(2) 以上述样本数据估计总体，从高速公路上行驶的家用轿车中随机抽取 3 辆，记这 3 辆车平均车速超过 100 km/h 且为男性驾驶员的车辆为 $X$，求 $X$ 的分布

列．

【解答】 (1) 平均车速不超过 100 km/h 的驾驶员有 40 人，从中随机抽取 2 人的方法总数为 $C_{40}^2$，记“这 2 人恰好有 1 名男性驾驶员和 1 名女性驾驶员”为事件 $A$，则事件 $A$ 所包含的基本事件数为 $C_{15}^1C_{25}^1$，故所求的概率 $P(A)=\frac{C_{15}^1C_{25}^1}{C_{40}^2}=\frac{15\times25}{20\times39}=\frac{25}{52}$.

(2) 根据样本估计总体的思想，从总体中任取 1 辆车，平均车速超过 100km/h 且为男性驾驶员的概率为 $\frac{40}{100}=\frac{2}{5}$，

故 $X\sim B\left(3,\ \frac{2}{5}\right)$.$X$ 的可能取值为 0,1,2,3，

且 $P(X=0)=C_3^0\left(\frac{2}{5}\right)^0\times\left(\frac{3}{5}\right)^3=\frac{27}{125}$，

$P(X=1)=C_3^1\times\frac{2}{5}\times\left(\frac{3}{5}\right)^2=\frac{54}{125}$，

$P(X=2)=C_3^2\left(\frac{2}{5}\right)^2\times\frac{3}{5}=\frac{36}{125}$，

$P(X=3)=C_3^3\left(\frac{2}{5}\right)^3\times\left(\frac{3}{5}\right)^0=\frac{8}{125}$.

所以 $X$ 的分布列为

| $X$ | 0 | 1 | 2 | 3 |
| --- | --- | --- | --- | --- |
| $P$ | $\frac{27}{125}$ | $\frac{54}{125}$ | $\frac{36}{125}$ | $\frac{8}{125}$ |

考法 4　超几何分布

例4　为推动乒乓球运动的发展，某乒乓球比赛允许不同协会的运动员组队参加．现有来自甲协会的运动员 3 名，其中种子选手 2 名；乙协会的运动员 5 名，其中种子选手 3 名．从这 8 名运动员中随机选择 4 人参加比赛．

(1) 设事件 $A$ 为“选出的 4 人中恰有 2 名种子选手，且这 2 名种子选手来自同一个协会”，求事件 $A$ 发生的概率；

(2) 设 $X$ 为选出的 4 人中种子选手的人数，求随机变量 $X$ 的分布列和均值．

【解答】 (1) 由已知有 $P(A)=\frac{C_2^2C_3^2+C_3^2C_3^2}{C_8^4}=\frac{6}{35}$，

所以事件 $A$ 发生的概率为$\frac{6}{35}$.

(2) 随机变量 $X$ 的所有可能取值为 1,2,3,4，

且 $P(X=k)=\frac{C_5^kC_3^{4-k}}{C_8^4}(k=1,2,3,4)$.

$P(X=1)=\frac{C_5^1C_3^3}{C_8^4}=\frac{1}{14}$，

$P(X=2)=\frac{C_5^2C_3^2}{C_8^4}=\frac{3}{7}$，

$P(X=3)=\frac{C_5^3C_3^1}{C_8^4}=\frac{3}{7}$，

$P(X=4)=\frac{C_5^4C_3^0}{C_8^4}=\frac{1}{14}$.

所以随机变量 $X$ 的分布列为

| $X$ | 1 | 2 | 3 | 4 |
| --- | --- | --- | --- | --- |
| $P$ | $\frac{1}{14}$ | $\frac{3}{7}$ | $\frac{3}{7}$ | $\frac{1}{14}$ |

所以 $E(X)=1\times\frac{1}{14}+2\times\frac{3}{7}+3\times\frac{3}{7}+4\times\frac{1}{14}=\frac{5}{2}$.

【类题固法】

1. 在新冠肺炎疫情期间，某医院有 10 名医生报名参加“援鄂医疗队”，其中有 3 名女医生. 现从中抽选 5 名医生，用 $X$ 表示抽到男医生的人数，则 $X=3$ 的概率为( D )

A. $\frac{7}{12}$　　B. $\frac{5}{36}$

C. $\frac{1}{12}$　　D. $\frac{5}{12}$

【解析】 “从 10 名医生中抽选 5 名医生”的事件总数为 $C_{10}^5$，“从 10 名医生中抽选 5 名医生，其中有 3 名男医生”包含的基本事件数为 $C_7^3C_3^2$，所以 $P(X=$

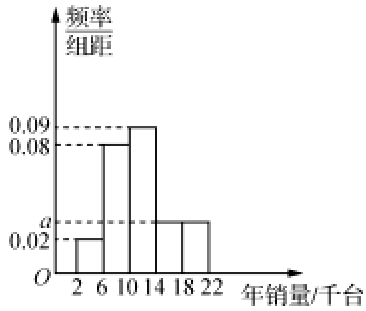
$3)=\frac{C_7^3C_3^2}{C_{10}^5}=\frac{5}{12}$.

2. 某医疗器械公司在全国共有 100 个销售点，总公司每年会根据每个销售点的年销量进行评价分析，规定每个销售点的年销售任务为一万四千台器械．根据这 10 个销售点的年销量绘制出如图所示的频率分布直方图．

(1) 求完成年销售任务的销售点数量；

(2) 若用分层抽样的方法从这 100 个销售点中抽取容量为 25 的样本，求[2,6)，[6,10)，[10,14)，[14,18)，[18,22)，(单位：千台)每组中分别应抽取的销售点数量；

(3) 在(2)的条件下，从前两组[2,6)，[6,10)中的销售点随机选取 3 个，记这 3 个销售点在[6,10)中的个数为 $X$，求 $X$ 的分布列和期望．




(第 2 题)

【解答】 (1) 由 $(0.02+0.08+0.09+2a)\times4=1$，得 $a=0.03$，则完成年销售任务的销售点个数为 $0.03\times2\times4\times100=24$.

(2) 各组应抽取的销售点数量比例为 $2:8:9:3:3$，

则各组应抽取的销售点数量分别为 2,8,9,3,3.

(3) 在第(2)问的容量为 25 的样本中，[2,6)，[6,10)中的销售点数量分别为 2,8，则 $X$ 所有可能的取值为 1,2,3，且

$P(X=1)=\frac{C_2^2C_8^1}{C_{10}^3}=\frac{1}{15}$，$P(X=2)=\frac{C_2^1C_8^2}{C_{10}^3}=\frac{7}{15}$，$P(X=3)=\frac{C_8^3}{C_{10}^3}=\frac{7}{15}$，

所以 $X$ 的分布列为

| $X$ | 1 | 2 | 3 |
| --- | --- | --- | --- |
| $P$ | $\frac{1}{15}$ | $\frac{7}{15}$ | $\frac{7}{15}$ |

所以 $E(X)=1\times\frac{1}{15}+2\times\frac{7}{15}+3\times\frac{7}{15}=\frac{12}{5}$.

考法 5　正态分布

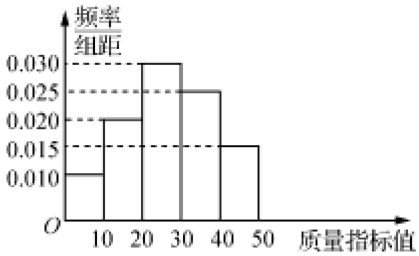
例5　2020 年春节前夕，$A$ 市某质检部门随机抽取了 100 包某种品牌的速冻水饺，检测其某项质量指标值，所得频率分布直方图如图.

(1) 求所抽取的 100 包速冻水饺该项质量指标值的样本平均数 $\overline{x}$ (同一组中的数据用该组区间的中点值作代表);

(2) ①由直方图可以认为，速冻水饺的该项质量指标值 $Z$ 服从正态分布 $N(\mu,\sigma^2)$，利用该正态分布，求 $Z$ 落在(14.55,38.45]内的概率;

②将频率视为概率，若某人从超市购买了 4 包这种品牌的速冻水饺，记这 4 包速冻水饺中这种质量指标值位于(10,30)内的包数为 $X$，求 $X$ 的分布列和均值.

附：这 100 包速冻水饺的质量指标值的标准差为 $\sigma=\sqrt{142.75}\approx 11.95$；若 $\xi\sim N(\mu,\sigma^2)$，则 $P(\mu-\sigma<\xi\leqslant\mu+\sigma)=0.682\,6$，$P(\mu-2\sigma<\xi\leqslant\mu+2\sigma)=0.954\,4$.




【解答】 (1) 所抽取的 100 包速冻水饺该项质量指标值的平均数 $\overline{x}=5\times 0.1+15\times 0.2+25\times 0.3+35\times 0.25+45\times 0.15=26.5$.

(2) ①因为 $Z$ 服从正态分布 $N(\mu,\sigma^2)$，且 $\mu=26.5$，$\sigma\approx 11.95$，

所以 $P(14.55<Z\leqslant 38.45)=P(26.5-11.95<Z\leqslant 26.5+11.95)=0.682\,6$，

所以 $Z$ 落在(14.55,38.45]内的概率是 0.682 6.

②根据题意得 $X\sim B\left(4,\frac{1}{2}\right)$，

$P(X=0)=C_4^0\left(\frac{1}{2}\right)^4=\frac{1}{16}$，$P(X=1)=C_4^1\left(\frac{1}{2}\right)^4=\frac{1}{4}$，$P(X=2)=C_4^2\left(\frac{1}{2}\right)^4=\frac{3}{8}$，

$P(X=3)=C_4^3\left(\frac{1}{2}\right)^4=\frac{1}{4}$，

$P(X=4)=C_4^4\left(\frac{1}{2}\right)^4=\frac{1}{16}$.

所以 $X$ 的分布列为

| $X$ | 0 | 1 | 2 | 3 | 4 |
| --- | --- | --- | --- | --- | --- |
| $P$ | $\frac{1}{16}$ | $\frac{1}{4}$ | $\frac{3}{8}$ | $\frac{1}{4}$ | $\frac{1}{16}$ |

所以 $E(X)=4\times\frac{1}{2}=2$.

**【类题固法】**

1. 设每天从甲地去乙地的旅客人数为随机变量 $X$，且 $X\sim N(800,50^2)$，则一天中从甲地去乙地的旅客人数不超过 900 的概率为(　A　)

(参考数据：若 $X\sim N(\mu,\sigma^2)$，则 $P(\mu-\sigma<X\leqslant\mu+\sigma)=0.682\ 6$，$P(\mu-2\sigma<X\leqslant\mu+2\sigma)=0.954\ 4$，$P(\mu-3\sigma<X\leqslant\mu+3\sigma)=0.997\ 4$)

A. 0.977 2　　　B. 0.682 6

C. 0.997 4　　　D. 0.954 4

**【解析】** 因为 $X\sim N(800,50^2)$，所以 $P(700<X\leqslant 900)=0.954\ 4$，所以 $P(X>900)=\frac{1-0.954\ 4}{2}=0.022\ 8$，所以 $P(X\leqslant 900)=1-0.022\ 8=0.977\ 2$.

2. 某班有 50 名学生，一次考试的数学成绩 $\xi$ 服从正态分布 $N(100,10^2)$，已知 $P(90\leqslant\xi\leqslant 100)=0.3$，估计该班学生数学成绩在 110 分以上的人数为__10__.

**【解析】** 由题意知 $P(\xi>110)=\frac{1-2P(90\leqslant\xi\leqslant 100)}{2}=0.2$，所以该班学生数学成绩在 110 分以上的人数为 $0.2\times 50=10$.

3. 若随机变量 $X\sim N(\mu,\sigma^2)$，且 $P(X>5)=P(X<-1)=0.2$，则 $P(2<X<5)=$__0.3__.

**【解析】** 因为 $P(X>5)=P(X<-1)$，所以 $\mu=\frac{5-1}{2}=2$，所以 $P(2<X<5)=\frac{1}{2}P(-1<X<5)=\frac{1}{2}\times(1-0.2-0.2)=0.3$.

4. 已知两个随机变量 $X$，$Y$ 满足 $X+2Y=4$，且 $X\sim N(1,2^2)$，则 $E(Y)=$__$\frac{3}{2}$__，

$D(Y)=\underline{\quad 1\quad}$.

【解析】 由 $X\sim N(1,2^2)$，得 $E(X)=1$，$D(X)=4$.又 $X+2Y=4$，所以 $Y=2-\frac{X}{2}$，

所以 $E(Y)=2-\frac{1}{2}E(X)=\frac{3}{2}$，$D(Y)=\frac{1}{4}D(X)=1$.

# 第八章

# 成对数据的统计分析

重点与难点

1. 相关性

2. 相关系数

3. 一元线性回归模型

4. 残差分析

5. 非线性回归模型

6. 列联表

7. 独立性检验

## 8.1　成对数据的统计相关性

### 第 1 课时　变量的相关关系

预习评价▸新知初探

**学习目标**

1. 了解变量间的相关关系，会画散点图.

2. 会用散点图判断两个变量之间是否具有相关关系.

**问题导引**

1.　变量间的关系一定是正相关或负相关么？

2.　如何判断变量之间的相关关系？

**现学现用**

1. 概念辨析(正确的打“√”，错误的打“×”)

(1) 粮食产量与施肥量之间的关系是函数关系.(　×　)

(2) 一定范围内，学生的成绩与学习时间成正相关关系.(　√　)

(3) 人的体重与视力成负相关关系．(　×　)

2.　下列两变量具有相关关系的是(　B　)

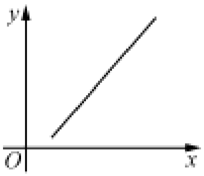
A. 正方体的体积与边长

B. 人的身高与体重

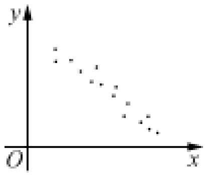
C. 匀速行驶车辆的行驶距离与时间

D. 球的半径与体积

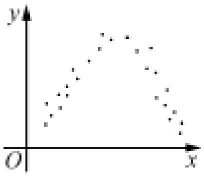
3. 下列四个图象中，两个变量具有正相关关系的是(　D　)



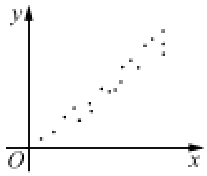

A




B




C




D

## 对点练 ▶ 先掌握基础

类型 1　变量之间的相关关系

例1　下列语句所表示的事件中的因素不具有相关关系的是(　D　)

A. 瑞雪兆丰年

B. 读书破万卷，下笔如有神

C. 吸烟有害健康

D. 喜鹊叫喜，乌鸦叫丧

**【解析】**“瑞雪兆丰年”和“读书破万卷，下笔如有神”是根据多年经验总结归纳出来的，吸烟有害健康具有科学根据，所以它们都是相关关系，所以 A，B，C 三项具有相关关系；结合生活经验知喜鹊和乌鸦发出叫声是它们自身的生理反应，与人无任何关系，故 D 项不具有相关关系．

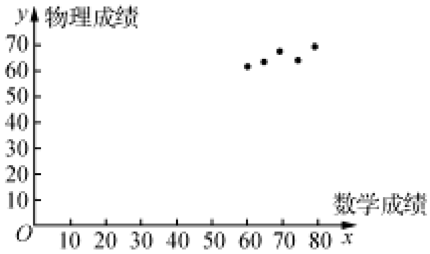
类型 2　变量之间的相关关系判断

例2　5 名学生的数学和物理成绩如下表：

| 学生<br>成绩<br>学科 | $A$ | $B$ | $C$ | $D$ | $E$ |
|---|---|---|---|---|---|
| 数学 | 80 | 75 | 70 | 65 | 60 |
| 物理 | 70 | 66 | 68 | 64 | 62 |

画出散点图，并判断它们是否具有相关关系.

【解答】 把数学成绩作为横坐标，把相应的物理成绩作为纵坐标，在直角坐标系中描点($x_i$，$y_1$)($i=1$,2，…，5)，作出散点图如图.



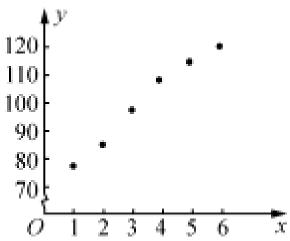

从图中可以直观地看出数学成绩和物理成绩具有相关关系，且当数学成绩增大时，物理成绩也在由小变大，即它们正相关.

**变式** 某个男孩的年龄与身高的统计数据如下表所示.

| 年龄 $x$/岁 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 身高 $y$/cm | 78 | 87 | 98 | 108 | 115 | 120 |

画出散点图并判断 $y$ 与 $x$ 是否具有线性相关关系.

【解答】 散点图如图所示．由图知，所有数据点接近一条直线排列，因此认为 $y$ 与 $x$ 有线性相关关系.




【规律总结】 判断两个变量 $x$ 和 $y$ 之间是否具有线性相关关系，常用的简便方法就是绘制散点图，如果发现点的分布从整体上看大致在一条直线附近，那么

这两个变量就是线性相关的，注意不要受个别点的位置的影响．

课堂评价▶基础达标

1．已知 $x$，$y$ 是两个变量，下列四个关系中，$x$，$y$ 呈负相关的是(　D　)

A. $y=x^2-1$　　　　　　　　B. $y=-x^2+1$

C. $y=x-1$　　　　　　　　　D. $y=-x+1$

2．　对变量 $x$，$y$ 有观测数据$(x_i，y_i)(_{i=1},2,3，\cdots，10)$，得散点图 1；对变量 $u$，$v$ 有观测数据$(u_i，v_i)(_{i=1},2,3，\cdots，10)$，得散点图 2，由这两个散点图可以断定(　C　)

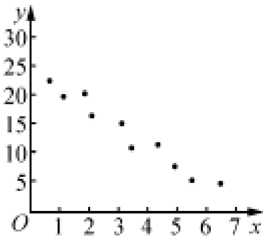



图 1

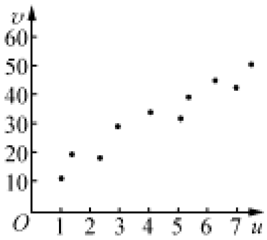



图 2

A. $x$ 与 $y$ 正相关，$u$ 与 $v$ 正相关

B. $x$ 与 $y$ 正相关，$u$ 与 $v$ 负相关

C. $x$ 与 $y$ 负相关，$u$ 与 $v$ 正相关

D. $x$ 与 $y$ 负相关，$u$ 与 $v$ 负相关

3．某商家今年上半年各月的人均销售额(单位：千元)与利润率统计如下表：

| 月份 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 人均销售额 | 6 | 5 | 8 | 3 | 4 | 7 |
| 利润率(%) | 12.6 | 10.4 | 18.5 | 3.0 | 8.1 | 16.3 |

根据表中数据，下列说法正确的是(　C　)

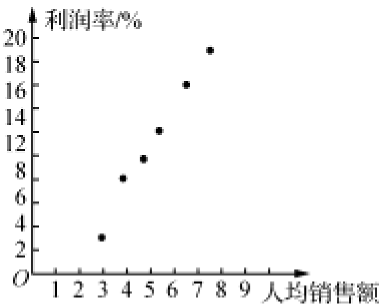
A.  利润率与人均销售额成正比例函数关系

B.  利润率与人均销售额成反比例函数关系

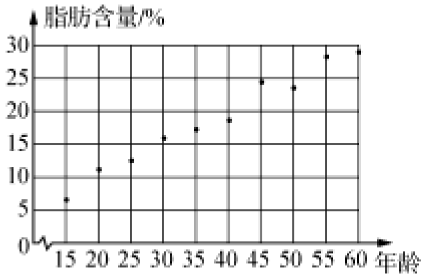
C.  利润率与人均销售额成正相关关系

D.  利润率与人均销售额成负相关关系

**【解析】** 根据题意，画出利润率与人均销售额的散点图，如图所示，由散点图知，利润率与人均销售额成正相关关系．




4. 在一次对人体脂肪含量和年龄关系的研究中，研究人员获得了一组样本数据，并制作成如图所示的人体脂肪含量与年龄关系的散点图．下列结论中正确的是(　B　)




A. 人体脂肪含量与年龄正相关，且脂肪含量的中位数等于 20%

B. 人体脂肪含量与年龄正相关，且脂肪含量的中位数小于 20%

C. 人体脂肪含量与年龄负相关，且脂肪含量的中位数等于 20%

D. 人体脂肪含量与年龄负相关，且脂肪含量的中位数小于 20%

【解析】 从散点图可以看出，年龄增大，脂肪含量也随之增加，故为正相关.中间的两个点即第 5,6 两个点脂肪含量均低于 20%，故脂肪含量的中位数小于 20%.

## 第 2 课时　样本相关系数

预习评价▸新知初探

### 学习目标

1. 了解样本相关系数的含义.
2. 会求样本相关系数.

### 问题导引

1. 样本相关系数的取值范围是什么？
2. 样本相关系数小，则成对样本数据的相关性就弱吗？

### 现学现用

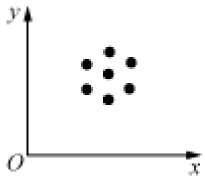
1. 概念辨析(正确的打“√”，错误的打“×”)

(1) 变量间的线性相关系数 $r$ 的取值范围为[0,1]．(　×　)

(2) 变量间的样本相关系数 $r$ 的绝对值越接近 0，则变量间的线性相关程度越低．(　√　)

(3) 若两个随机变量的线性相关性越强，则相关系数 $r$ 的值越接近于 1.(　×　)

2. 若变量 $x$，$y$ 的散点图如图所示，则 $x$，$y$ 之间的样本相关系数 $r$ 最接近的值为(　C　)




A. 1　　　　B. $-0.5$

C. 0　　　　D. 0.5

3. 已知变量 $x$ 和变量 $y$ 的 3 对随机观测数据为(2,2)，(3，$-1$)，(5，$-7$)，则成对样本数据的样本相关系数是__$-1$__.

【解析】 因为 $\overline{x}=\frac{10}{3}$，$\overline{y}=-2$，

所以 $r=\dfrac{\sum\limits_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\sum\limits_{i=1}^{n}(x_i-\overline{x})^2}\sqrt{\sum\limits_{i=1}^{n}(y_i-\overline{y})^2}}=$

$$\frac{\left(2-\frac{10}{3}\right)\times 4+\left(3-\frac{10}{3}\right)\times 1+\left(5-\frac{10}{3}\right)\times(-5)}{\sqrt{\left(2-\frac{10}{3}\right)^2+\left(3-\frac{10}{3}\right)^2+\left(5-\frac{10}{3}\right)^2}\times\sqrt{4^2+1^2+(-5)^2}}$$

$=-1$.

## 对点练 ▶ 先掌握基础

### 类型 1　相关系数的性质

例1　对两个变量 $x$，$y$ 进行线性相关检验，得线性相关系数 $r_1=0.785\ 9$，对两个变量 $u$，$v$ 进行线性相关检验，得线性相关系数 $r_2=-0.956\ 8$，则下列判断正确的是(　C　)

A. 变量 $x$ 与 $y$ 正相关，变量 $u$ 与 $v$ 负相关，变量 $x$ 与 $y$ 的线性相关性较强

B. 变量 $x$ 与 $y$ 负相关，变量 $u$ 与 $v$ 正相关，变量 $x$ 与 $y$ 的线性相关性较强

C. 变量 $x$ 与 $y$ 正相关，变量 $u$ 与 $v$ 负相关，变量 $u$ 与 $v$ 的线性相关性较强

D. 变量 $x$ 与 $y$ 负相关，变量 $u$ 与 $v$ 正相关，变量 $u$ 与 $v$ 的线性相关性较强

【解析】 由线性相关系数 $r_1=0.785\ 9>0$ 知 $x$ 与 $y$ 正相关，由线性相关系数 $r_2=-0.956\ 8<0$ 知 $u$ 与 $v$ 负相关．又$|r_1|<|r_2|$，所以变量 $u$ 与 $v$ 的线性相关性比 $x$ 与 $y$ 的线性相关性强．

**变式** 甲、乙、丙、丁四位同学各自对 $x$，$y$ 两变量的线性相关性作试验，分别求得相关系数 $r$，如下表：

| 相关系数 | 甲 | 乙 | 丙 | 丁 |
|---|---|---|---|---|
| $r$ | $-0.82$ | 0.78 | 0.69 | 0.87 |

则试验结果体现两变量有更强线性相关性的是( D )

A. 甲　　B. 乙

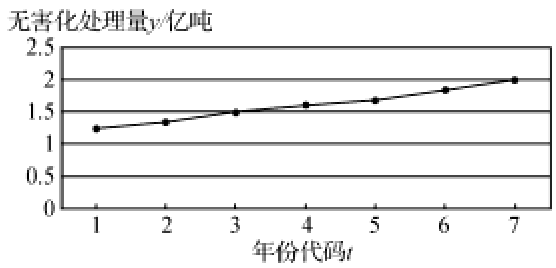
C. 丙　　D. 丁

【规律总结】 样本相关系数 $r$ 的性质：①$|r|\leqslant 1$.②当 $r>0$ 时，表明两个变量正相关；当 $r<0$ 时，表明两个变量负相关．③$|r|$越接近于 1，表明两个变量的线性相关性越强．④$|r|$越接近于 0，表明两个变量的线性相关性越弱．

类型 2　样本相关系数的计算及应用

**例2** 如图是我国 2015 年至 2021 年生活垃圾无害化处理量(单位：亿吨)的折线图．

注：年份代码 1～7 分别对应年份 2015～2021.




请求出相关系数 $r$，并用相关系数的大小说明 $y$ 与 $t$ 相关性的强弱．

参考数据：$\sum\limits_{i=1}^{7}y_i=10.97$，$\sum\limits_{i=1}^{7}t_iy_i=47.36$，

$\sqrt{\sum\limits_{i=1}^{7}(y_i-\overline{y})^2}=0.664$，$\sqrt{7}\approx2.646$.

【解答】 由折线图中数据和参考数据得

$\overline{t}=4$，$\sum\limits_{i=1}^{7}(t_i-\overline{t})^2=28$，$\sqrt{\sum\limits_{i=1}^{7}(y_i-\overline{y})^2}=0.664$，

$\sum\limits_{i=1}^{7}(t_i-\overline{t})(y_i-\overline{y})=\sum\limits_{i=1}^{7}t_iy_i-\overline{t}\sum\limits_{i=1}^{7}y_i=47.36-4\times10.97=3.48$，所以 $r\approx$

$\dfrac{3.48}{0.664\times2\times2.646}\approx0.99$，

所以 $y$ 与 $t$ 的线性相关程度比较高.

**变式** 测得 10 对父子身高(单位：英寸，注：1 英寸＝0.025 4m)如下：

| 父亲身高($x$) | 60 | 62 | 64 | 65 | 66 | 67 | 68 | 70 | 72 | 74 |
|---|---|---|---|---|---|---|---|---|---|---|
| 儿子身高($y$) | 63.6 | 65.2 | 66 | 65.5 | 66.9 | 67.1 | 67.4 | 68.3 | 70.1 | 70 |

请求出相关系数 $r$，并用相关系数的大小说明 $y$ 与 $x$ 相关性的强弱.

参考数据：$\sum\limits_{i=1}^{10}y_i^2=44\ 941.93$，$\sum\limits_{i=1}^{10}x_i^2=44\ 794$，$\sum\limits_{i=1}^{10}x_iy_i=44\ 842.4$.

【解答】 $\overline{x}=66.8$，$\overline{y}=67.01$，$\sum\limits_{i=1}^{10}x_i^2=44\ 794$，$\sum\limits_{i=1}^{10}y_i^2=44\ 941.93$，$\overline{x}\ \overline{y}=4$

$476.27$，$\overline{x}^2=4\ 462.24$，$\overline{y}^2=4\ 490.34$，$\sum\limits_{i=1}^{10}x_iy_i=44\ 842.4$，

所以 $r=\dfrac{\sum\limits_{i=1}^{10}x_iy_i-10\overline{x}\ \ \overline{y}}{\sqrt{\left(\sum\limits_{i=1}^{10}x_i^2-10\overline{x}^2\right)\left(\sum\limits_{i=1}^{10}y_i^2-10\overline{y}^2\right)}}$

$=\dfrac{44\ 842.4-10\times4\ 476.27}{\sqrt{(4\ 4794-44\ 622.4)\times(4\ 4941.93-44\ 903.4)}}\approx0.98$，所以 $y$ 与 $x$ 之间具有较强的线性相关关系.

【规律总结】 用相关系数 $r$ 可以衡量两个变量之间的相关关系的强弱，$r$ 的

绝对值越接近于 1，表示两个变量的线性相关性越强，且 $r$ 的正负即表示两个变量相关性的正负．相关系数的取值范围是[－1，1]．

课堂评价▸基础达标

1. 为了比较甲、乙、丙三组数据的样本相关性的强弱，小郑分别计算了甲、乙、丙三组数据的样本相关系数，其数值分别为 0.939,0.937,0.948，则(　D　)

A. 甲组数据的线性相关性最强，乙组数据的线性相关性最弱

B. 乙组数据的线性相关性最强，丙组数据的线性相关性最弱

C. 丙组数据的线性相关性最强，甲组数据的线性相关性最弱

D. 丙组数据的线性相关性最强，乙组数据的线性相关性最弱

2. 在一组样本数据$(x_1，y_1)$，$(x_2，y_2)$，…，$(x_n，y_n)(n\geqslant 2，x_1，x_2，…，x_n$ 不相等)的散点图中，若所有样本点$(x_i，y_i)(_{i=1},2，…，n)$都在直线 $y=\frac{1}{2}x+1$ 上，则这组样本数据的样本相关系数为(　D　)

A. $-1$　　　　B. 0

C. $\frac{1}{2}$　　　　D. 1

【解析】 由题意知，这组样本数据完全正相关，故相关系数为 1.

3. 对于样本相关系数 $r$，叙述最准确的是(　C　)

A. $|r|\in(0，+\infty)$，$|r|$越大相关程度越大，反之相关程度越小

B. $r\in(-\infty，+\infty)$，$r$ 越大相关程度越大，反之相关程度越小

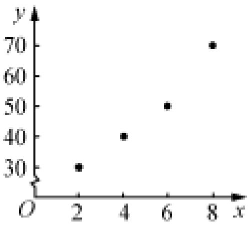
C. $|r|\leqslant 1$，且$|r|$越接近 1 相关程度越大，$|r|$越接近 0，相关程度越小

D. 以上说法都不对

4. 某厂的生产原料耗费 $x$(单位：百万元)与销售额 $y$(单位：百万元)之间有如图的对应关系，判断 $x$ 与 $y$ 之间是否具有样本相关关系；若有，判断相关性的强弱.

| $x$ | 2 | 4 | 6 | 8 |
|---|---|---|---|---|
| $y$ | 30 | 40 | 50 | 70 |

【解答】 画出散点图如图所示，由图可知 $x$，$y$ 有线性相关关系.




$\overline{x}=5$，$\overline{y}=47.5$，$\sum\limits_{i=1}^{4}x_i^2=120$，$\sum\limits_{i=1}^{4}y_i^2=9\,900$，

$\sum\limits_{i=1}^{4}x_iy_i=1\,080$，

$$r=\frac{\sum\limits_{i=1}^{4}x_iy_i-4\overline{x}\ \overline{y}}{\sqrt{\left(\sum\limits_{i=1}^{4}x_i^2-4\overline{x}^2\right)\left(\sum\limits_{i=1}^{4}y_i^2-4\overline{y}^2\right)}}$$

$$=\frac{1\,080-4\times5\times47.5}{\sqrt{(120-4\times5^2)\times(9\,900-4\times47.5^2)}}$$

$\approx0.982\,7$，

故 $x$ 与 $y$ 之间具有很强的正相关关系.

## 8.2 一元线性回归模型及其应用

### 第 1 课时 一元线性回归模型与最小二乘法

预习评价▶新知初探

#### 学习目标

1. 了解一元线性回归模型.
2. 了解最小二乘法.
3. 会求简单的经验回归方程.

#### 问题导引

1. 回归分析的含义是什么？？
2. 经验回归方程中 $\hat{a}$ 与 $\hat{b}$ 的含义是什么？怎样求解？

#### 现学现用

1. 概念辨析(正确的打“√”，错误的打“×”)

(1) 求经验回归方程前可以不进行相关性检验.( × )

(2) 经验回归直线 $\hat{y}=\hat{b}x+\hat{a}$ 一定过样本点中心($\overline{x}$，$\overline{y}$)( √ )

(3) 若某商品的销售量 $y$(件)关于销售价格 $x$(元/件)的经验回归方程为$\hat{y}=-5x+350$，当销售价格为 10 元时，销售量一定为 300 件．( × )

2. 散点图在回归分析中的作用是( D )

A. 查找个体个数

B. 比较个体数据大小关系

C. 探究个体分类

D. 粗略判断变量是否相关

3. $x$ 与 $y$ 之间的经验回归方程表示( D )

A. $x$ 与 $y$ 之间的函数关系

B. $x$ 与 $y$ 之间的不确定性关系

C. $x$ 与 $y$ 之间的真实关系形式

D. $x$ 与 $y$ 之间的真实关系达到最大限度的吻合

4. 已知经验回归方程$\hat{y}=0.75x+0.7$，则 $x=11$ 时，$y$ 的估计值为 8.95.

## 对点练▶先掌握基础

### 类型 1 对经验回归方程和最小二乘法的理解

例1 (多选)由一组样本数据$(x_1,y_1)$，$(x_2,y_2)$，…，$(x_n,y_n)$得到的经验回归方程为$\hat{y}=bx+a$，那么下面说法不正确的是( AB )

A. 通过 $y=bx+a$ 及回归系数 $b$，可以精确反映变量的取值和变化趋势

B. 直线$\hat{y}=bx+a$ 至少经过点$(x_1,y_1)$，$(x_2,y_2)$，…，$(x_n,y_n)$中的一个

C. 直线$\hat{y}=bx+a$ 的斜率为$\dfrac{\sum\limits_{i=1}^{n}x_iy_i-n\overline{x}\,\overline{y}}{\sum\limits_{i=1}^{n}x_i^2-n\overline{x}^2}$

D. 直线$\hat{y}=bx+a$ 和各点$(x_1,y_1)$，$(x_2,y_2)$，…，$(x_n,y_n)$的总偏差$\sum\limits_{i=1}^{n}[y_i-(bx_i+a)]^2$是该坐标平面上所有直线与这些点的偏差中最小的直线

【解析】 对于 A，通过经验回归方程 $y=bx+a$ 及回归系数 $b$，可估计和预测变量的取值和变化趋势，A 错误；对于 B，经验回归直线描述样本点的变化趋势和相关关系，未必经过样本点，B 错误；对于 C，由最小二乘法知$\hat{b}=$

$\dfrac{\sum\limits_{i=1}^{n}x_iy_i-n\overline{x}\ \overline{y}}{\sum\limits_{i=1}^{n}x_i^2-n\overline{x}^2}$，C 正确；对于 D，经验回归直线是所有直线中与样本点离散度最低的，由此可知经验回归直线的总偏差$\sum\limits_{i=1}^{n}[y_i-(bx_i+a)]^2$是该坐标平面上所有直线与这些点的偏差中最小的，D 正确．

【规律总结】 对经验回归方程的理解

(1) 从参数计算公式$\hat{a}=\overline{y}-\hat{b}\overline{x}$中，我们可以看出，回归直线方程$\hat{y}=\hat{b}x+\hat{a}$一定经过点($\overline{x}$，$\overline{y}$)．我们把($\overline{x}$，$\overline{y}$)称为样本点的中心，因此，回归直线必过样本点的中心．

(2) 经验回归方程$\hat{y}=\hat{b}x+\hat{a}$中的截距$\hat{a}$和斜率$\hat{b}$都是通过估计而得来的，存在着误差，这种误差可能导致预测结果的偏差．

(3) 经验回归方程$\hat{y}=\hat{b}x+\hat{a}$中的$\hat{b}$表示$x$增加 1 个单位时，$y$的平均变化量为$\hat{b}$，而$\hat{a}$表示$y$不随$x$的变化而变化的量．

类型 2　求经验回归方程

例2　第十八届中国国际农产品交易会上，重庆市全面推广“遂宁红薯”及“遂宁鲜”农产品区域公用品牌，并组织了 100 家企业、1 000 个产品进行展示展销，扩大优质特色农产品市场的占有率和影响力，提升遂宁特色农产品的社会认知度和美誉度，让来自世界各地的与会者和消费者更深入了解遂宁．某记者对本次农交会进行了跟踪报道和实际调查，对某特产的最满意度$x$(%)和对应销售额$y$(单位：万元)调查得到以下数据：

| 时间 | 第一天 | 第二天 | 第三天 | 第四天 | 第五天 |
|---|---|---|---|---|---|
| 最满意度 $x$/% | 22 | 34 | 25 | 20 | 19 |
| 销售额 $y$/万元 | 78 | 90 | 86 | 76 | 75 |

(1) 求销量额$y$关于最满意度$x$的相关系数$r$.我们约定：销量额$y$关于最满意度$x$的相关系数$r$的绝对值在 0.95 以上(含 0.95)是线性相关性较强；否则，线性相关性较弱．请你对线性相关性强弱作出判断，并给出理由．

(2) 如果没有达到较强线性相关，则采取“末位淘汰”制(即销售额最少的那一天不作为计算数据)，并求在剔除“末位淘汰”的那一天后的销量额 $y$ 关于最满意度 $x$ 的经验回归方程(系数精确到 0.1).

参考数据：$\overline{x}=24$，$\overline{y}=81$，$\sum_{i=1}^{5}x_i^2-5\overline{x}^2=146$，

$\sum_{i=1}^{5}y_i^2-5\overline{y}^2=176$，$\sum_{i=1}^{5}x_iy_i-5\overline{x}\,\overline{y}=151$，$\sqrt{146}\approx12.08$，$\sqrt{176}\approx13.27$.

【解答】 (1) $r=\dfrac{\sum_{i=1}^{5}x_iy_i-5\overline{x}\,\overline{y}}{\sqrt{\sum_{i=1}^{5}x_i^2-5\overline{x}^2}\sqrt{\sum_{i=1}^{5}y_i^2-5\overline{y}^2}}=\dfrac{151}{\sqrt{146\times176}}=\dfrac{151}{12.08\times13.27}$

$\approx0.94$.因为 $r\approx0.94<0.95$，所以线性相关性较弱.

(2) 由(1)可得没有达到较强线性相关，则淘汰销售额为 75 万元的数据．剔除数据后的 $\overline{x}'=25.25$，$\overline{y}'=82.5$，

$\sum_{i=1}^{4}x_iy_i=22\times78+34\times90+25\times86+20\times76=8\ 446$，

$4\overline{x}'\cdot\overline{y}'=4\times25.25\times82.5=8332.5$，

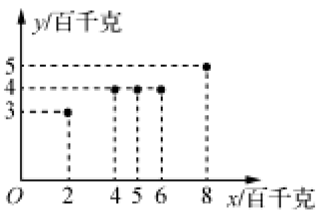
$\sum_{i=1}^{4}x_i^2=22^2+34^2+25^2+20^2=2\ 665$, $4\overline{x}'^2=4\times25.25\times25.25=2\ 550.25$，

所以 $\hat{b}=\dfrac{8\ 446-8\ 332.5}{2\ 665-2\ 550.25}\approx1$，

$\hat{a}=\overline{y}'-\hat{b}\,\overline{x}'=82.5-25.25\approx77.3$.

所以经验回归方程为 $\hat{y}=x+77.3$.

**变式** 根据统计，某蔬菜基地西红柿亩产量的增加量 $y$(单位：百千克)与某种液体肥料每亩使用量 $x$(单位：百千克)之间的对应数据的散点图如图所示.




(1) 依据数据的散点图可以看出，可用线性回归模型拟合 $y$ 与 $x$ 的关系，请计算相关系数 $r$ 并加以说明(若$|r|>0.75$，则线性相关程度很高，可用线性回归模型

拟合)；

(2) 求 $y$ 关于 $x$ 的经验回归方程.

参考数据：$\sqrt{0.3}\approx0.55$，$\sqrt{0.9}\approx0.95$.

【解答】(1) 由已知数据可得 $\overline{x}=\dfrac{2+4+5+6+8}{5}=5$，$\overline{y}=\dfrac{3+4+4+4+5}{5}=4$，

所以 $\sum\limits_{i=1}^{5}(x_i-\overline{x})(y_i-\overline{y})=(-3)\times(-1)+(-1)\times0+0\times0+1\times0+3\times1=6$，

$\sqrt{\sum\limits_{i=1}^{5}(x_i-\overline{x})^2}=\sqrt{(-3)^2+(-1)^2+0^2+1^2+3^2}=2\sqrt{5}$，$\sqrt{\sum\limits_{i=1}^{5}(y_i-\overline{y})^2}=\sqrt{(-1)^2+0^2+0^2+0^2+1^2}=\sqrt{2}$，

所以相关系数 $r=\dfrac{\sum\limits_{i=1}^{5}(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\sum\limits_{i=1}^{5}(x_i-\overline{x})^2}\sqrt{\sum\limits_{i=1}^{5}(y_i-\overline{y})^2}}=\dfrac{6}{2\sqrt{5}\times\sqrt{2}}=\sqrt{\dfrac{9}{10}}$ $\approx0.95>0.75$，所以可用线性回归模型拟合 $y$ 与 $x$ 的关系.

(2) $\hat{b}=\dfrac{\sum\limits_{i=1}^{5}(x_i-\overline{x})(y_i-\overline{y})}{\sum\limits_{i=1}^{5}(x_i-\overline{x})^2}=\dfrac{6}{20}=0.3$，$\hat{a}=4-5\times0.3=2.5$，所以经验回归方程为 $\hat{y}=0.3x+2.5$.

【规律总结】 求经验回归方程的三个步骤：

(1) 算：根据数据计算 $\overline{x}$，$\overline{y}$，$\sum\limits_{i=1}^{n}x_i^2$，$\sum\limits_{i=1}^{n}x_iy_i$.

(2) 代：代入公式求 $\hat{b}$，$\hat{a}$ 的具体数值.

(3) 求：由上面的计算结果求方程 $\hat{y}=\hat{b}x+\hat{a}$.

课堂评价▶基础达标

1. 某同学为了了解气温对热饮销售的影响，经过统计分析，得到一个卖出的

热饮杯数 $y$ 与当天气温 $x$ 的经验回归方程$\hat{y}=-2.352x+147.767$.下列选项正确的是(　B　)

A. $x$ 与 $y$ 线性正相关

B. $x$ 与 $y$ 线性负相关

C. $y$ 随 $x$ 增大而增大

D. $y$ 随 $x$ 减小而减小

【解析】 由回归方程$\hat{y}=-2.352x+147.767$，可得 $x$ 与 $y$ 线性负相关，且 $y$ 随 $x$ 增大而减小.

2. 随机变量 $x$ 与 $y$ 的数据如下表，其中缺少了一个数值，已知 $y$ 关于 $x$ 的经验回归方程为$\hat{y}=0.9x+3$，则缺少的数值为(　A　)

| $x$ | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| $y$ | 5 | 6 | ▲ | 7 | 9 |

A. 6　　　　B. 6.6

C. 7.5　　　　D. 8

【解析】 设缺少的数值为 $m$，由于回归方程为$\hat{y}=0.9x+3$ 过样本中心点$(\overline{x},\overline{y})$，且$\overline{x}=\frac{2+3+4+5+6}{5}=4$，代入$\hat{y}=0.9\times4+3=6.6$，所以$\overline{y}=\frac{5+6+7+9+m}{5}=6.6$，解得 $m=6$.

3. 已知变量 $x$ 与 $y$ 负相关，且由观测数据算得样本平均数$\overline{x}=2$，$\overline{y}=2.5$，则由该观测数据算得的经验回归方程可能是(　D　)

A. $\hat{y}=0.4x+1.7$　　　　B. $\hat{y}=2x-1.2$

C. $\hat{y}=-3x+7.5$　　　　D. $\hat{y}=-2x+6.5$

【解析】 由于变量 $x$ 与 $y$ 负相关，排除 A，B，把(2,2.5)代入直线$\hat{y}=-2x+6.5$，得 $2.5=-2\times2+6.5$ 成立，所以(2,2.5)在直线上.

4. 在一次考试中，5 名学生的数学和物理成绩如下表：(已知学生的数学和物理成绩具有线性相关关系)

| 学生的编号 $i$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|

| 数学成绩 $x$ | 80 | 75 | 70 | 65 | 60 |
|---|---|---|---|---|---|
| 物理成绩 $y$ | 70 | 66 | 68 | 64 | 62 |

现已知其经验回归方程为$\hat{y}=0.36x+\hat{a}$，则根据此经验回归方程估计数学得 90 分的同学的物理成绩为__73__.(四舍五入到整数)

**【解析】** $\overline{x}=\dfrac{60+65+70+75+80}{5}=70$，$\overline{y}=\dfrac{62+64+66+68+70}{5}=66$，

所以 $66=0.36\times 70+\hat{a}$，即$\hat{a}=40.8$，即线性回归方程为$\hat{y}=0.36x+40.8$.当 $x=90$ 时，$\hat{y}=0.36\times 90+40.8=73.2\approx 73$.

5. 某测试团队为了研究“饮酒”对“驾车安全”的影响，随机选取 100 名驾驶员先后在无酒状态、酒后状态下进行“停车距离”测试．测试的方案：电脑模拟驾驶，以某速度匀速行驶，记录下驾驶员的“停车距离”(驾驶员从看到意外情况到车子停下所需要的距离)，无酒状态与酒后状态下的试验数据分别如下表，且表 1 数据的中位数估计值为 26.

表 1

| 停车距离 $d$/m | (10,20] | (20,30] | (30,40] | (40,50] | (50,60] |
|---|---|---|---|---|---|
| 频数 | 26 | $a$ | $b$ | 8 | 2 |

表 2

| 平均每毫升血液酒精含量 $x$/mg | 10 | 30 | 50 | 70 | 90 |
|---|---|---|---|---|---|
| 平均停车距离 $y$/m | 30 | 50 | 60 | 70 | 90 |

(1) 求 $a$，$b$ 的值，并估计驾驶员无酒状态下停车距离的平均数；

(2) 根据最小二乘法，由表 2 的数据计算 $y$ 关于 $x$ 的经验回归方程$\hat{y}=\hat{b}x+\hat{a}$；

(3) 该测试团队认为：若驾驶员酒后驾车的平均“停车距离”$y$ 大于(1)中无酒状态下的停车距离平均数的 3 倍，则认定驾驶员是“醉驾”．请根据(2)中的经验回归方程，预测当每毫升血液酒精含量大于多少时为“醉驾”．

附：$\hat{b}=\dfrac{\sum\limits_{i=1}^{n}x_iy_i-n\overline{x}\,\overline{y}}{\sum\limits_{i=1}^{n}x_i^2-n\overline{x}^2}$，$\hat{a}=\overline{y}-\hat{b}\,\overline{x}$.

【解答】 (1) 依题意，得$\frac{26-20}{10}=\frac{50-26}{a}$，解得 $a=40$.又 $a+b+36=100$，

解得 $b=24$，

故停车距离的平均数为 $15\times\frac{26}{100}+25\times\frac{40}{100}+35\times\frac{24}{100}+45\times\frac{8}{100}+55\times\frac{2}{100}=$ 27.

(2) 依题意，可知 $\overline{x}=50$，$\overline{y}=60$，

$\sum_{i=1}^{5}x_iy_i=10\times30+30\times50+50\times60+70\times70+90\times90=17\ 800$，

$\sum_{i=1}^{5}x_i^2=10^2+30^2+50^2+70^2+90^2=16\ 500$，

所以$\hat{b}=\frac{17\ 800-5\times50\times60}{16\ 500-5\times50^2}=0.7$，$\hat{a}=60-0.7\times50=25$，

所以经验回归方程为$\hat{y}=0.7x+25$.

(3) 由(1)知当 $y>81$ 时，认定驾驶员是“醉驾”.

令$\hat{y}>81$，得 $0.7x+25>81$，解得 $x>80$，

则当每毫升血液酒精含量大于 80mg 时认定为“醉驾”.

## 第 2 课时　回归分析模型及其应用

预习评价▸新知初探

### 学习目标

1. 了解残差与残差分析.

2. 针对实际问题，会用一元线性回归模型(非线性回归模型)进行预测.

### 问题导引

1. 线性回归模型怎样用表达式表示？产生随机误差的原因是什么？

2. 刻画回归效果的方式有哪些？

### 现学现用

1. 概念辨析(正确的打“√”，错误的打“×”)

(1) 在残差图中，纵坐标为残差，横坐标可以选为样本编号.(　√　)

(2) 在残差图中，残差点比较均匀落在水平的带状区域中，即可说明选用的

模型比较合适，与带状区域宽度无关．( × )

(3) 在线性回归模型中，决定系数 $R^2$ 表示解释变量对于预报变量变化的贡献率，$R^2$ 越接近于 1，拟合的效果越好．( √ )

2. 已知经验回归方程$\hat{y}=\hat{b}x+0.6$ 相对于点(3,6.5)的残差为$-0.1$，则$\hat{b}$的值为( B )

A. 1　　　　B. 2

C. $-0.5$　　　　D. $-3$

**【解析】** 因为相对于点(3,6.5)的残差为-0.1，所以 $6.5-\hat{y}=-0.1$，所以 $6.5+0.1=3\hat{b}+0.6$，解得$\hat{b}=2$.

3. 已知一系列样本点$(x_i，y_i)(i=1,2,3，\cdots，n)$的经验回归方程为$\hat{y}=2x+a$，若样本点$(r,1)$与$(1，s)$的残差相同，则有( C )

A. $r=s$　　　　B. $s=2r$

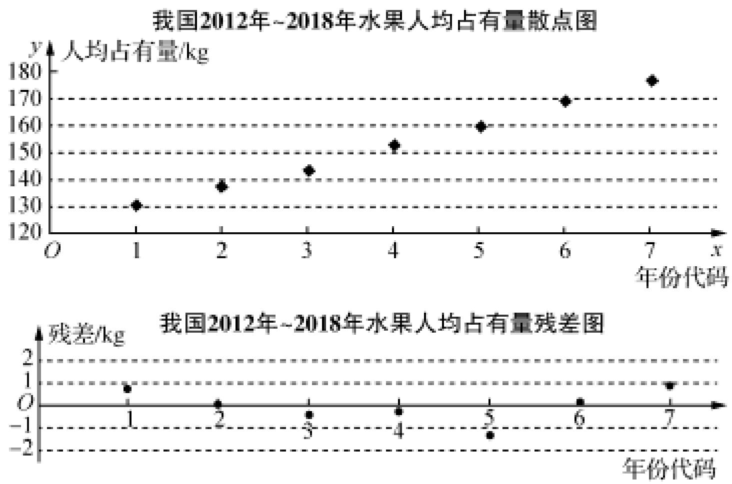
C. $s=-2r+3$　　　　D. $s=2r+1$

**【解析】** 样本点$(r,1)$的残差为 $2r+a-1$，样本点$(1，s)$的残差为 $2+a-s$，依题意 $2r+a-1=2+a-s$，故 $s=-2r+3$.

## 对点练 ▶ 先掌握基础

### 类型 1　回归分析中的“决定系数 $R^2$ 与残差图”

**例1**　下面给出了我国 2012 年～2018 年水果人均占有量 $y$(单位：kg)和年份代码 $x$ 绘制的散点图和经验回归方程的残差图(2012 年～2018 年的年份代码 $x$ 分别为 1～7)．



(1) 根据散点图分析 $y$ 与 $x$ 之间的相关关系；

(2) 根据散点图相应数据计算得 $\sum\limits_{i=1}^{7}y_{i=1}\ 074$，$\sum\limits_{i=1}^{7}x_iy_i=4\ 517$，求 $y$ 关于 $x$ 的经验回归方程；

(3) 根据经验回归方程的残差图，分析经验回归方程的拟合效果．(精确到 0.01)

【解答】 (1) 由散点图可以看出，点大致分布在某一直线的附近，且当 $x$ 由小变大时，$y$ 也由小变大，从而 $y$ 与 $x$ 之间是正相关关系．

(2) 由题得 $\overline{x}=\frac{1}{7}\times(1+2+3+4+5+6+7)=4$，$\overline{y}=\frac{1}{7}\times1\ 074=\frac{1\ 074}{7}$，故 $\hat{b}=\frac{\sum\limits_{i=1}^{7}x_iy_i-7\overline{x}\ \overline{y}}{\sum\limits_{i=1}^{7}x_i^2-7\overline{x}^2}=\frac{4\ 517-7\times\frac{1\ 074}{7}\times4}{1^2+2^2+3^2+4^2+5^2+6^2+7^2-7\times4^2}=\frac{221}{28}$，所以 $\hat{a}=\overline{y}-\hat{b}\overline{x}=\frac{1\ 074}{7}-\frac{221}{28}\times4=\frac{853}{7}$，故所求 $y$ 关于 $x$ 的经验回归方程为 $\hat{y}=\frac{221}{28}x+\frac{853}{7}$.

(3) 由残差图可以看出，残差对应的点均匀地落在水平带状区域内，且宽度较窄，说明拟合效果较好．

**变式** 为研究质量 $x$(单位：g)对弹簧长度 $y$(单位：cm)的影响，对不同质量的 6 个物体进行测量，数据如下表所示：

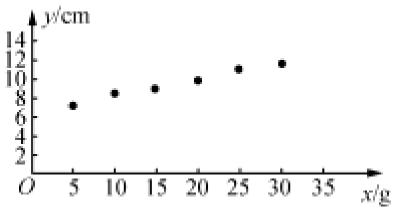
| $x$ | 5 | 10 | 15 | 20 | 25 | 30 |
|---|---|---|---|---|---|---|
| $y$ | 7.25 | 8.12 | 8.95 | 9.90 | 10.9 | 11.8 |

(1) 作出散点图并求经验回归方程；

(2) 求 $R^2$；

(3) 进行残差分析．

【解答】 (1) 散点图如图．




$\overline{x}=\frac{1}{6}\times(5+10+15+20+25+30)=17.5$，

$\overline{y}=\frac{1}{6}\times(7.25+8.12+8.95+9.90+10.9+11.8)\approx9.487$，$\sum\limits_{i=1}^{6}x_i^2=2\ 275$，$\sum\limits_{i=1}^{6}x_iy_i=1\ 076.2$，

$\hat{b}=\frac{\sum\limits_{i=1}^{6}x_iy_i-6\overline{x}\ \overline{y}}{\sum\limits_{i=1}^{6}x_i^2-6\overline{x}^2}\approx0.183$，$\hat{a}=\overline{y}-\hat{b}\overline{x}\approx6.285$，

故所求经验回归方程为$\hat{y}=0.183x+6.285$.

(2) 列表如下：

| | | | | | | |
|---|---|---|---|---|---|---|
| $y_i-\hat{y}_i$ | 0.05 | 0.005 | $-0.08$ | $-0.045$ | 0.04 | 0.025 |
| $y_i-y$ | $-2.24$ | $-1.37$ | $-0.54$ | 0.41 | 1.41 | 2.31 |

所以$\sum\limits_{i=1}^{6}(y_i-\hat{y}_i)^2\approx0.013\ 18$，

$\sum\limits_{i=1}^{6}(y_i-\overline{y})^2=14.678\ 4$.

所以 $R^2=1-\frac{0.013\ 18}{14.678\ 4}\approx0.999\ 1$，回归模型的拟合效果较好.

(3) 由残差表中的数值可以看出第 3 个样本点的残差比较大，需要确认在采集这个数据的时候是否有人为的错误，如果有的话，需要纠正数据，重新建立回归模型；由表中数据可以看出残差点比较均匀地落在不超过 0.15 的狭窄的水平带状区域中，说明选用的线性回归模型的精度较高，由以上分析可知，弹簧长度与拉力呈线性关系.

**【规律总结】** “相关指数 $R^2$、残差图”在回归分析中的作用：

(1) 相关指数 $R^2$ 是用来刻画回归效果的，$R^2$ 越大，意味着残差平方和越小，也就是说模型的拟合效果就越好.

(2) 残差图也是用来刻画回归效果的，判断依据是：残差点比较均匀地分布在水平带状区域中，带状区域越窄，说明模型拟合精度越高，回归方程预报的精度也越高.

类型 2　线性回归分析

**例2** 每年春暖以后至寒冬前，昆虫大量活动与繁殖，易于采集各种药用昆虫．已知一只药用昆虫的产卵数 $y$(单位：个)与一定范围内的温度 $x$(单位：℃)有关，于是科研人员在 3 月份的 31 天中随机挑选了 5 天进行研究，现收集了该种药用昆虫的 5 组观测数据如下表：

| 日期 | 2 日 | 7 日 | 15 日 | 22 日 | 30 日 |
| --- | --- | --- | --- | --- | --- |
| 温度 $x$/℃ | 10 | 11 | 13 | 12 | 8 |
| 产卵数 $y$/个 | 23 | 25 | 30 | 26 | 16 |

科研人员确定的研究方案是：先从这 5 组数据中任选 2 组，用剩下的 3 组数据建立 $y$ 关于 $x$ 的经验回归方程，再对被选取的 2 组数据进行检验．

(1) 若选取的是 3 月 2 日与 30 日这 2 组的数据，请根据 3 月 7 日、15 日和 22 日这 3 组的数据，求出 $y$ 关于 $x$ 的经验回归方程；

(2) 若由经验回归方程得到的估计数据与所选出的检验数据的误差均不超过 2 个，则认为得到的经验回归方程是可靠的，试问(1)中所得的经验回归方程是否可靠？

【解答】 (1) 由题得 $\overline{x}=12$，$\overline{y}=27$，$\sum\limits_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})=5$，$\sum\limits_{i=1}^{n}(x_i-\overline{x})^2=2$，所以 $\hat{b}=\dfrac{\sum\limits_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sum\limits_{i=1}^{n}(x_i-\overline{x})^2}=\dfrac{5}{2}$，$\hat{a}=\overline{y}-\dfrac{5}{2}\overline{x}=27-\dfrac{5}{2}\times12=-3$，所以 $y$ 关于 $x$ 的经验回归方程为 $\hat{y}=\dfrac{5}{2}x-3$.

(2) 由(1)知，$y$ 关于 $x$ 的经验回归方程为 $\hat{y}=\dfrac{5}{2}x-3$.

当 $x=10$ 时，$\hat{y}=\dfrac{5}{2}\times10-3=22$，$|22-23|<2$，

当 $x=8$ 时，$\hat{y}=\dfrac{5}{2}\times8-3=17$，$|17-16|<2$.

所以(1)中所得的经验回归方程 $\hat{y}=\dfrac{5}{2}x-3$ 是可靠的．

【规律总结】 线性回归分析的两种策略：

(1) 利用回归方程进行预测：把回归直线方程看作一次函数，求函数值．

(2) 利用回归直线判断正、负相关：决定正相关还是负相关的是回归系数 $b$.

类型 3　非线性回归分析

例3　中国茶文化博大精深，已知茶水的口感与茶叶类型以及水温有关．经验表明，某种绿茶用 85 ℃的水泡制，再等到茶水温度降至 60 ℃时饮用，可以产生最佳口感．某学习研究小组通过测量，得到了下面表格中的数据(室温是 20 ℃).

| 泡制时间 $x$/min | 0 | 1 | 2 | 3 | 4 |
| --- | --- | --- | --- | --- | --- |
| 水温 $y$/℃ | 85 | 79 | 74 | 71 | 65 |

(1) 小组成员根据上面表格中的数据绘制散点图，并根据散点图分布情况，考虑到茶水温度降到室温(即 20 ℃)就不能再降的事实，决定选择函数模型 $y=kc^x+20(x\geqslant 0)$来刻画．

①令 $z=\ln(y-20)$，求 $z$ 关于 $x$ 的经验回归方程；

②利用①的结论，求 $y=kc^x+20(x\geqslant 0,\ c>0)$中的 $k$ 与 $c$.

(2) 你认为该品种绿茶用 85 ℃的水大约泡制多久后饮用，可以产生最佳口感？

参考数据：$\ln 65\approx 4.2$，$\lg 59\approx 4.1$，$\ln 54\approx 4.0$，$\ln 51\approx 3.9$，$\ln 45\approx 3.8$，$\log_{0.9}0.6\approx 4.8$，$e^{-0.1}\approx 0.9$，$e^{4.2}\approx 66.7$，$\frac{400}{667}\approx 0.6$.

【解答】 (1) ①由已知得出 $x$ 与 $z$ 的关系，如下表：

| | 0 | 1 | 2 | 3 | 4 |
| --- | --- | --- | --- | --- | --- |
| $z$ | 4.2 | 4.1 | 4.0 | 3.9 | 3.8 |

设经验回归方程为$\hat{z}=\hat{b}x+\hat{a}$，由题意得$\overline{x}=\frac{1}{5}\times(0+1+2+3+4)=2$，$\overline{z}=\frac{1}{5}\times(4.2+4.1+4.0+3.9+3.8)=4$，

故$\sum\limits_{i=1}^{5}(x_i-\overline{x})(z_i-\overline{z})=(-2)\times 0.2+(-1)\times 0.1+1\times(-0.1)+2\times(-0.2)=-1$，

$\sum\limits_{i=1}^{5}(x_i-\overline{x})^2=(-2)^2+(-1)^2+1^2+2^2=10$，

则$\hat{b}=\dfrac{\sum\limits_{i=1}^{5}(x_i-\overline{x})(z_i-\overline{z})}{\sum\limits_{i=1}^{5}(x_i-\overline{x})^2}=-\dfrac{1}{10}=-0.1$，

$\hat{a}=\overline{z}-\hat{b}\overline{x}=4+0.1\times2=4.2$，

则 z 关于 x 的经验回归方程为$\hat{z}=-0.1x+4.2$.

②由 $y=kc^x+20(x\geqslant0)$，

得 $y-20=kc^x(x\geqslant0)$，

两边取对数得 $\ln(y-20)=\ln k+x\ln c$，

利用①的结论得 $\ln c=-0.1$，$\ln k=4.2$，

所以 $c=e^{-0.1}\approx0.9$，$k=e^{4.2}\approx66.7$.

(2) 由(1)得 $y=66.7\times0.9^x+20(x\geqslant0)$，

令 $y=60$，得 $x\approx\log_{0.9}0.6\approx4.8$.

所以该品种绿茶用 85 ℃的水泡制 4.8 min 后饮用，口感最佳.

**【规律总结】** 求非线性回归方程的步骤：

(1) 确定变量，作出散点图.

(2) 根据散点图，选择恰当的拟合函数.

(3) 变量置换，通过变量置换把非线性回归问题转化为线性回归问题，并求出经验回归方程.

(4) 分析拟合效果：通过计算相关指数或画残差图来判断拟合效果.

(5) 根据相应的变换，写出非经验回归方程.

课堂评价▶基础达标

1. 对于给定的样本点所建立的模型 $A$ 和模型 $B$，它们的残差平方和分别是 $a_1$，$a_2$，$R^2$ 的值分别为 $b_1$，$b_2$，下列说法正确的是( C )

A. 若 $a_1<a_2$，则 $b_1<b_2$，$A$ 的拟合效果更好

B. 若 $a_1<a_2$，则 $b_1<b_2$，$B$ 的拟合效果更好

C. 若 $a_1<a_2$，则 $b_1>b_2$，$A$ 的拟合效果更好

D. 若 $a_1<a_2$，则 $b_1>b_2$，$B$ 的拟合效果更好

2. 已知$\hat{y}=0.85x-85.7$ 是根据女大学生的身高预报体重的经验回归方程(其中 $x$，$y$ 的单位分别是 cm，kg)，则该方程在样本(165,57)处的残差是 2.45.

【解析】 当 $x=165$ 时，$\hat{y}=0.85\times165-85.7=54.55$，所以方程在样本(165,57)处的残差是 $57-54.55=2.45$.

3. 若关于 $x$ 与 $y$ 的数据如下，且有如下两个模型：①$\hat{y}=6.5x+17.5$；②$\hat{y}=7x+17$.

| $x$ | 2 | 4 | 5 | 6 | 8 |
| --- | --- | --- | --- | --- | --- |
| $y$ | 30 | 40 | 60 | 50 | 70 |

通过残差分析发现第①个经验回归模型比第②个拟合效果好，则 $R_1^2$__>__$R_2^2$，$Q_1$__<__$Q_2$.(用“>”或“<”填空，$R$，$Q$ 分别是决定系数和残差平方和)

【解析】 由 $R^2$ 的性质可得，$R^2$ 越大模型的拟合效果越好，所以 $R_1^2>R_2^2$.由残差的性质可得，残差平方和越小模型的拟合效果越好，所以 $Q_1<Q_2$.

4. 在研究两个变量的相关关系时，观察散点图发现样本点集中于某一条指数曲线 $y=e^{bx+a}$ 的周围，令 $z=\ln y$，求得经验回归方程为$\hat{z}=0.25x-2.58$，则该模型的回归方程为$\hat{y}=e^{0.25x-2.58}$.

【解析】 由 $z=\ln y$，$\hat{z}=0.25x-2.58$，得 $\ln\hat{y}=0.25x-2.58$，所以$\hat{y}=e^{0.25x-2.58}$，故该模型的经验回归方程为$\hat{y}=e^{0.25x-2.58}$.

## 习题课　经验回归方程的应用

题型 1　经验回归方程的理解

例1　根据最小二乘法由一组样本点$(x_i，y_i)$(其中 $i=1,2，\cdots，500$)，求得的经验回归方程是$\hat{y}=\hat{b}x+\hat{a}$，则下列说法不正确的是(　B　)

A. 样本点可能全部都不在回归直线$\hat{y}=\hat{b}x+\hat{a}$上

B. 若所有样本点都在回归直线$\hat{y}=\hat{b}x+\hat{a}$上，则变量间的相关系数为 1

C. 若所有的样本点都在回归直线$\hat{y}=\hat{b}x+\hat{a}$上，则$\hat{b}x_i+\hat{a}$的值与 $y_i$ 相等

D. 若回归直线$\hat{y}=\hat{b}x+\hat{a}$的斜率$\hat{b}<0$，则变量 $x$ 与 $y$ 呈负相关

【解析】回归直线必过样本数据中心点，但样本点可能全部不在回归直线上，故 A 正确；所有样本点都在回归直线$\hat{y}=\hat{b}x+\hat{a}$上，则变量间的相关系数可能为±1，故 B 错误；若所有的样本点都在回归直线$\hat{y}=\hat{b}x+\hat{a}$上，则$\hat{b}x+\hat{a}$的值与 $y_i$ 相等，故 C 正确；若回归直线$\hat{y}=\hat{b}x+\hat{a}$的斜率$\hat{b}<0$，则 $r<0$，则变量 $x$ 与 $y$ 呈负相关，故 D 正确.

题型 2　经验回归模型的应用

例2　随着电商事业的快速发展，网络购物交易额也快速提升，特别是每年的“双十一”，天猫的交易额数目惊人.2020 年天猫公司的工作人员为了迎接天猫“双十一”年度购物狂欢节，加班加点做了大量准备活动，截止 2020 年 11 月 11 日 24 时，2020 年的天猫“双十一”交易额定格在 3 700 多亿元，天猫总公司所有员工对于新的战绩皆大欢喜，同时又对 2021 年充满了憧憬，因此公司工作人员反思从 2014 年至 2020 年每年“双十一”总交易额(取近似值)，进行分析统计如下表：

| 年份 | 年份代码($t$) | 总交易额 $y$(单位：百亿) |
|---|---|---|
| 2014 | 1 | 5.7 |
| 2015 | 2 | 9.1 |
| 2016 | 3 | 12.1 |
| 2017 | 4 | 16.8 |
| 2018 | 5 | 21.3 |
| 2019 | 6 | 26.8 |
| 2020 | 7 | 37 |

(1) 通过分析，发现可用线性回归模型拟合总交易额 $y$ 与年份代码 $t$ 的关系，请用相关系数加以说明；

(2) 利用最小二乘法建立 $y$ 关于 $t$ 的经验回归方程(系数精确到 0.1)，推测 2022

年天猫“双十一”的总交易额.

参考数据：$\sum_{i=1}^{7}(t_i-\overline{t})(y_i-\overline{y})=138.5$，$\sqrt{\sum_{i=1}^{7}(y_i-\overline{y})^2}=26.7$，$\sqrt{7}\approx2.646$.

【解答】 (1) $\overline{t}=4$，$\sum_{i=1}^{7}(t_i-\overline{t})^2=28$，

$\sum_{i=1}^{7}(t_i-\overline{t})(y_i-\overline{y})=138.5$，$\sqrt{\sum_{i=1}^{7}(y_i-\overline{y})^2}=26.7$，

所以 $r=\dfrac{\sum_{i=1}^{n}(t_i-\overline{t})(y_i-\overline{y})}{\sqrt{\sum_{i=1}^{n}(t_i-\overline{t})^2\sum_{i=1}^{n}(y_i-\overline{y})^2}}$

$\approx\dfrac{138.5}{2\times2.646\times26.7}\approx0.98$.

因为总交易额 $y$ 与年份代码 $t$ 的相关系数近似为 0.98，

说明总交易额 $y$ 与年份代码 $t$ 的线性相关性很强，

从而可用线性回归模型拟合总交易额 $y$ 与年份代码 $t$ 的关系.

(2) 因为 $\overline{y}=18.4$，$\sum_{i=1}^{7}(t_i-\overline{t})^2=28$，

所以 $\hat{b}=\dfrac{\sum_{i=1}^{7}(t_i-\overline{t})(y_i-\overline{y})}{\sum_{i=1}^{7}(t_i-\overline{t})^2}=\dfrac{138.5}{28}\approx4.9$，

$\hat{a}=\overline{y}-\hat{b}\,\overline{t}\approx18.4-4.9\times4=-1.2$，

所以 $y$ 关于 $t$ 的经验回归方程为 $\hat{y}=4.9t-1.2$.

又将 2022 年对应的 $t=9$ 代入回归方程得 $\hat{y}=4.9\times9-1.2=42.9$，所以推测 2022 年天猫“双十一”的总交易额约为 42.9 百亿.

**变式** 某地随着经济的发展，居民收入逐年增长，该地一建设银行连续五年的储蓄存款(年底余额)如下表 1：

表 1

| 年份 $x$ | 2013 | 2014 | 2015 | 2016 | 2017 |
|---|---|---|---|---|---|

| 储蓄存款 $y$(千亿元) | 5 | 6 | 7 | 8 | 10 |
| --- | --- | --- | --- | --- | --- |

为了研究计算的方便，工作人员将上表的数据进行了处理，$t=x-2\ 012$，$z=y-5$ 得到如下表 2：

表 2

| 时间代号 $t$ | 1 | 2 | 3 | 4 | 5 |
| --- | --- | --- | --- | --- | --- |
| $z$ | 0 | 1 | 2 | 3 | 5 |

(1) 求 $z$ 关于 $t$ 的经验回归方程；

(2) 通过(1)中的方程，求 $y$ 关于 $x$ 的经验回归方程；

(3) 用所求经验回归方程推测到 2022 年年底，该地储蓄存款额.

【解答】 (1) 因为 $\overline{t}=3$，$\overline{z}=2.2$，$\sum\limits_{i=1}^{5}t_iz_i=45$，

$\sum\limits_{i=1}^{5}t_i^2=55$，$\hat{b}=\dfrac{45-5\times3\times2.2}{55-5\times9}=1.2$，

$\hat{a}=\overline{z}-\hat{b}\,\overline{t}=2.2-3\times1.2=-1.4$，

所以$\hat{z}=1.2t-1.4$.

(2) 将 $t=x-2\ 012$，$z=y-5$，代入$\hat{z}=1.2t-1.4$，

得 $y-5=1.2(x-2\ 012)-1.4$，即$\hat{y}=1.2x-2\ 410.8$.

(3) 因为$\hat{y}=1.2\times2\ 022-2\ 410.8=15.6$，

所以推测到 2022 年年底，该地储蓄存款额可达 15.6 千亿元.

题型 3　非线性回归模型的应用

例3　某企业为了了解年研发资金投入额 $x$(单位：亿元)对年盈利额 $y$(单位：亿元)的影响，研究了“十二五”和“十三五”规划发展期间近 10 年年研发资金投入额 $x_i$ 和年盈利额 $y_i$ 的数据．通过对比分析，建立了两个函数模型：①$y=\alpha+\beta x^2$，②$y=e^{\lambda x+t}$，其中 $\alpha$，$\beta$，$\lambda$，$t$ 为常数，e 为自然对数的底数．令 $u_i=x_i^2$，$v_i=\ln y_i(i=1,2,\cdots,10)$，经计算得如下数据：$\overline{x}=26$，$\overline{y}=215$，$\sum\limits_{i=1}^{10}(x_i-\overline{x})^2=65$，

$\sum\limits_{i=1}^{10}(y_i-\overline{y})^2=2$，$\overline{u}=680$，$\overline{v}=5.36$，

$\sum\limits_{i=1}^{10}(u_i-\overline{u})^2=11\ 250$，$\sum\limits_{i=1}^{10}(u_i-\overline{u})(y_i-\overline{y})=130$，$\sum\limits_{i=1}^{10}(v_i-\overline{v})^2=2.6$，$\sum\limits_{i=1}^{10}(x_i$

$-\overline{x})(v_i-\overline{v})=12$.

(1) 请从相关系数的角度，分析哪一个模型拟合程度更好？

(2) ①根据(1)的选择及表中数据，建立 $y$ 关于 $x$ 的经验回归方程；(系数精确到 0.01)

②若希望 2021 年的年盈利额 $y$ 为 250 亿元，请推测 2021 年的年研发资金投入额 $x$.(结果精确到 0.01)

参考数据：$\ln 2\approx 0.693$，$\ln 5\approx 1.609$.

【解答】 (1) 设 $u_i$ 和 $y_i$ 的相关系数为 $r_1$，$x_i$ 和 $v_i$ 的相关系数为 $r_2$，由题意，

$$r_1=\frac{\sum_{i=1}^{10}(u_i-\overline{u})(y_i-\overline{y})}{\sqrt{\sum_{i=1}^{10}(u_i-\overline{u})^2\sum_{i=1}^{10}(y_i-\overline{y})^2}}$$

$$=\frac{130}{\sqrt{11\ 250\times 2}}=\frac{13}{15}\approx 0.87,$$

$$r_2=\frac{\sum_{i=1}^{10}(x_i-\overline{x})(v_i-\overline{v})}{\sqrt{\sum_{i=1}^{10}(x_i-\overline{x})^2\sum_{i=1}^{10}(v_i-\overline{v})^2}}=\frac{12}{\sqrt{65\times 2.6}}=\frac{12}{13}\approx 0.92，则 |r_1|<|r_2|,$$

因此从相关系数的角度，模型 $y=e^{\lambda x+t}$ 的拟合程度更好.

(2) ①先建立 $v$ 关于 $x$ 的经验回归方程，

由 $y=e^{\lambda x+t}$，得 $\ln y=t+\lambda x$，即 $v=t+\lambda x$，

$$\hat{\lambda}=\frac{\sum_{i=1}^{10}(x_i-\overline{x})(v_i-\overline{v})}{\sum_{i=1}^{10}(x_i-\overline{x})^2}=\frac{12}{65}\approx 0.18,$$

$$\hat{t}=\overline{v}-\hat{\lambda}\overline{x}=5.36-\frac{12}{65}\times 26=0.56,$$

所以 $v$ 关于 $x$ 的经验回归方程为$\hat{v}=0.18x+0.56$，

所以 $\ln\hat{y}=0.18x+0.56$，则$\hat{y}=e^{0.18x+0.56}$.

②2021 年的年盈利额 $y=250$(亿元)，

所以 $250=e^{0.18x+0.56}$，则 $0.18x+0.56=\ln 250$.

因为 $\ln 250=3\ln 5+\ln 2\approx 3\times 1.609+0.693=5.52$，

所以 $x\approx\dfrac{5.52-0.56}{0.18}\approx 27.56$.

所以推测 2021 年的年研发资金投入量约为 27.56 亿元.

**变式** 某种新产品投放市场一段时间后，经过调研获得了时间 $x$(天数)与销售单价 $y$(元)的一组数据，且做了一定的数据处理(如下表)，并作出了如图散点图．表中 $w_i=\dfrac{1}{x_i}$，$\overline{w}=\dfrac{1}{10}\sum\limits_{i=1}^{10}w_i$.

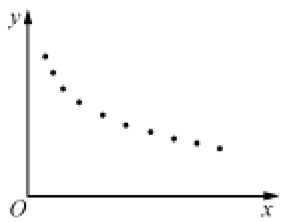
| $\overline{x}$ | $\overline{y}$ | $\overline{\omega}$ | $\sum\limits_{i=1}^{10}(x_i-\overline{x})^2$ | $\sum\limits_{i=1}^{10}(\omega_i-\overline{\omega})^2$ |
|---|---|---|---|---|
| 1.63 | 37.8 | 0.89 | 5.15 | 0.92 |

| $\sum\limits_{i=1}^{10}(x_i-\overline{x})(y_i-\overline{y})$ | $\sum\limits_{i=1}^{10}(\omega_i-\overline{\omega})(y_i-\overline{y})$ |
|---|---|
| $-20.6$ | 18.40 |

(1) 根据散点图判断$\hat{y}=\hat{a}+\hat{b}x$ 与$\hat{y}=\hat{c}+\dfrac{\hat{d}}{x}$哪一个更适合作为价格 $y$ 关于时间 $x$ 的回归方程类型；(不必说明理由)

(2) 根据判断结果和表中数据，建立 $y$ 关于 $x$ 的经验回归方程；

(3) 若该产品的日销售量 $g(x)$(件)与时间 $x$ 的函数关系为 $g(x)=\dfrac{-100}{x}+120(x\in\mathbf{N}^*)$，问该产品投放市场第几天的销售额最高？最高为多少元？



【解答】 (1) 根据散点图知$\hat{y}=\hat{c}+\frac{\hat{d}}{x}$更适合作为价格 $y$ 关于时间 $x$ 的回归方程类型.

(2) 令 $w=\frac{1}{x}$，则$\hat{y}=\hat{c}+\hat{d}w$，

而$\hat{d}=\frac{\sum_{i=1}^{10}(w_i-\overline{w})(y_i-\overline{y})}{\sum_{i=1}^{10}(w_i-\overline{w})^2}=\frac{18.40}{0.92}=20$，

$\hat{c}=\overline{y}-\hat{d}\,\overline{w}=37.8-20\times 0.89=20$，即有$\hat{y}=20\left(1+\frac{1}{x}\right)$.

(3) 由题意结合(2)知日销售额为

$f(x)=\hat{y}\cdot g(x)=20\left(1+\frac{1}{x}\right)\left(120-\frac{100}{x}\right)$，

所以$f(x)=20\left(1+\frac{1}{x}\right)\left(120-\frac{100}{x}\right)=400\left(6+\frac{1}{x}-\frac{5}{x^2}\right)$.

若 $t=\frac{1}{x}$，令 $h(t)=6+t-5t^2=-5\left(t-\frac{1}{10}\right)^2+\frac{121}{20}$，

所以 $t=\frac{1}{10}$时，$h(t)_{max}=h\left(\frac{1}{10}\right)=\frac{121}{20}$，即 $x=10$ 时，

$f(x)_{max}=f(10)=400\times\frac{121}{20}=2\ 420$ 元.

所以该产品投放市场第 10 天的销售额最高，最高销售额为 2 420 元.

【规律总结】 (1) 正确理解计算$\hat{b}$，$\hat{a}$的公式和准确的计算是求经验回归方程的关键.

(2) 经验回归方程$\hat{y}=\hat{b}x+\hat{a}$必过样本点中心$(\overline{x}，\overline{y})$.

(3) 在分析两个变量的相关关系时，可根据样本数据作出散点图来确定两个变量之间是否具有相关关系，若具有线性相关关系，则可通过经验回归方程来估计和预测.

(4) 对于非线性回归分析问题，应先进行变量代换， 求出代换后的经验回归

方程，再求非线性回归方程．

## 8.3　列联表与独立性检验

### 第 1 课时　分类变量与列联表

预习评价▸新知初探

**学习目标**

1. 通过实例，理解 2×2 列联表的统计意义．

2. 掌握 2×2 列联表和等高堆积条形图的实际应用．

**问题导引**

1. 分类变量的概念是什么？什么是列联表？什么是 2×2 列联表？

2. 如何利用等高条形图判断两个变量之间的关系？

**现学现用**

1. 概念辨析(正确的打“√”，错误的打“×”)

(1) 列联表中的数据是两个分类变量的频数. (　√　)

(2) 从等高条形图中可以精确地判断两个分类变量是否有关系. (　×　)

(3) 2×2 列联表主要用于研究两个事件之间相互独立的还是存在某种关联性，它适用于分析两个事件之间的关系．(　√　)

2. 可以粗略地判断两个分类变量是否有关的是(　B　)

A. 散点图　　B. 等高条形图

C. 2×2 列联表　　D. 以上均不对

3. 下面是一个 2×2 列联表

| | $y_1$ | $y_2$ | 合计 |
|---|---|---|---|
| $x_1$ | $a$ | 21 | 73 |
| $x_2$ | 2 | 25 | 27 |
| 总计 | $b$ | 46 | 100 |

则表中 $a$，$b$ 处的值分别为(　C　)

A. 94,96　　B. 52,50

C. 52,54　　D. 54,52

4. 研究某医院某段时间内婴儿出生的时间与性别的关系，得到如下的数据：出生时间在晚上的男婴为 24 人，女婴为 8 人；出生时间在白天的男婴为 31 人，

女婴为 26 人．将下面的 2×2 列联表补充完整.

| 性别＼出生时间 | 晚上 | 白天 | 总计 |
|---|---|---|---|
| 男婴 | | | |
| 女婴 | | | |
| 总计 | | | |

【解答】

| 性别＼出生时间 | 晚上 | 白天 | 总计 |
|---|---|---|---|
| 男婴 | 24 | 31 | 55 |
| 女婴 | 8 | 26 | 34 |
| 总计 | 32 | 57 | 89 |

**对点练 ▶ 先掌握基础**

**类型 1　建立 2×2 列联表**

**例1**　针对某新型病毒，某科研机构已研发出甲、乙两种疫苗，为比较两种疫苗的效果，选取 100 名志愿者，将他们随机分成两组，每组 50 人．第一组志愿者注射甲种疫苗，第二组志愿者注射乙种疫苗，经过一段时间后，对这 100 名志愿者进行该新型病毒抗体检测，发现有$\frac{1}{10}$的志愿者未产生该新型病毒抗体，在未产生该新型病毒抗体的志愿者中，注射甲种疫苗的志愿者占$\frac{1}{5}$.

| | 产生抗体 | 未产生抗体 | 合计 |
|---|---|---|---|
| 甲 | | | |
| 乙 | | | |
| 合计 | | | |

根据题中数据，完成列联表.

【解答】 由题意可得未产生该新型病毒抗体的志愿者的人数为 $100\times\frac{1}{10}=$

10，则注射甲种疫苗的志愿者中未产生抗体的人数为 $10\times\frac{1}{5}=2$，产生抗体的人数为 $50-2=48$；注射乙种疫苗的志愿者中未产生抗体的人数为 $10-2=8$，产生抗体的人数为 $50-8=42$.列联表如下：

| | 产生抗体 | 未产生抗体 | 合计 |
|---|---|---|---|
| 甲 | 48 | 2 | 50 |
| 乙 | 42 | 8 | 50 |
| 合计 | 90 | 10 | 100 |

【规律总结】 $2\times2$ 列联表有助于直观地观测数据之间的关系，主要用于研究两个事件之间相互独立的还是存在某种关联性，它适用于分析两个事件之间的关系.

类型 2　$2\times2$ 列联表的简单运用

例2　在某测试中，卷面满分为 100 分，60 分为及格，为了调查午休对本次测试前两个月复习效果的影响，特对复习中进行午休和不进行午休的考生进行了测试成绩的统计，数据如下表所示：

| 分数段 | 29～40 | 41～50 | 51～60 | 61～70 | 71～80 | 81～90 | 91～100 |
|---|---|---|---|---|---|---|---|
| 午休考生人数 | 23 | 47 | 30 | 21 | 14 | 31 | 14 |
| 不午休考生人数 | 17 | 51 | 67 | 15 | 30 | 17 | 3 |

(1) 根据上述表格完成下面列联表：

| | 及格人数 | 不及格人数 | 总计 |
|---|---|---|---|
| 午休 | | | |
| 不午休 | | | |
| 总计 | | | |

(2) 根据列联表可以得出什么样的结论？对今后的复习有什么指导意义？

【解答】 (1) 根据题表中数据可以得到列联表如下:

| | 及格人数 | 不及格人数 | 总计 |
|---|---|---|---|
| 午休 | 80 | 100 | 180 |
| 不午休 | 65 | 135 | 200 |
| 总计 | 145 | 235 | 380 |

(2) 计算可知，午休的考生及格率为 $P_1=\frac{80}{180}=\frac{4}{9}$，不午休的考生的及格率为 $P_2=\frac{65}{200}=\frac{13}{40}$，则 $P_1>P_2$，因此，可以粗略判断午休与考生考试及格有关系，并且午休的及格率高，所以在以后的复习中考生应尽量适当午休，以保持最佳的学习状态.

**变式** (多选)随着我国经济结构调整和方式转变，社会对高质量人才的需求越来越大，因此考研现象在我国不断升温．某大学甲、乙两个本科专业，研究生的报考和录取情况如下表，则(　　)

| 性别 | 甲专业报考人数 | 乙专业报考人数 | 性别 | 甲专业录取率 | 乙专业录取率 |
|---|---|---|---|---|---|
| 男 | 100 | 400 | 男 | 25% | 45% |
| 女 | 300 | 100 | 女 | 30% | 50% |

A. 甲专业比乙专业的录取率高

B. 乙专业比甲专业的录取率高

C. 男生比女生的录取率高

D. 女生比男生的录取率高

【解析】 由题意可得甲专业录取了男生 25 人，女生 90 人；乙专业录取了男生 180 人，女生 50 人．甲专业的录取率为$\frac{25+90}{100+300}=28.75\%$，乙专业的录取率为$\frac{180+50}{400+100}=46\%$，所以乙专业比甲专业的录取率高．男生的录取率为$\frac{25+180}{100+400}=$

41%，女生的录取率为$\frac{90+50}{300+100}=35\%$，所以男生比女生的录取率高．

【规律总结】 可以通过 2×2 列联表中的$\frac{a}{a+b}$和$\frac{c}{c+d}$值的大小粗略地判断两个事件之间有无关系，一般其值越大，两个事件有关系的可能性就越大．

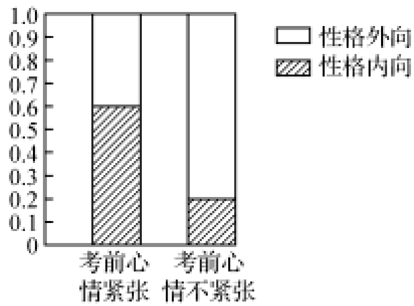
类型 3　等高条形图的应用

例3　某校对高三学生作了一项调查，发现在平时的模拟考试中，性格内向的学生 426 人中 332 人在考前心情紧张，性格外向的学生 594 人中有 213 人在考前心情紧张，作出等高条形图，利用图形判断考前心情紧张与性格类别是否有关系．

【解答】 作列联表如下：

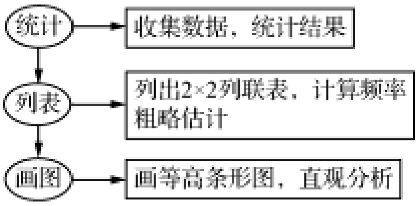
| | 性格内向 | 性格外向 | 总计 |
|---|---|---|---|
| 考前心情紧张 | 332 | 213 | 545 |
| 考前心情不紧张 | 94 | 381 | 475 |
| 总计 | 426 | 594 | 1 020 |

相应的等高条形图如图所示：



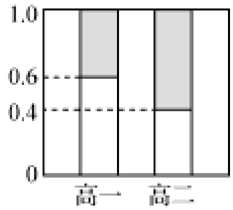

图中阴影部分表示考前心情紧张与考前心情不紧张中性格内向的比例，从图中可以看出，考前心情紧张的样本中性格内向占的比例，比考前心情不紧张样本中性格内向占的比例大，可以认为考前紧张与性格类别有关．

【规律总结】 利用等高条形图判断两个分类变量是否相关的步骤




课堂评价▶基础达标

1. 如图是调查某学校高一、高二年级学生参加社团活动的等高条形图，阴影部分的高表示参加社团的频率．已知该校高一、高二年级学生人数均为 600 人(所有学生都参加了调查)，现从参加社团的同学中按分层随机抽样的方式抽取 45 人，则抽取的高二学生人数为(　C　)

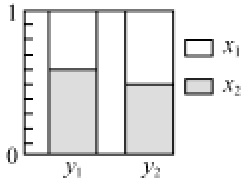



A. 9　　　　B. 18

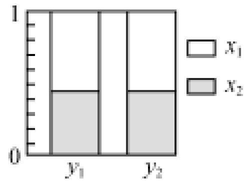
C. 27　　　　D. 36

【解析】 根据等高条形图可知，参加社团的高一和高二的人数比为 2∶3，由分层随机抽样的性质可得，抽取的高二学生人数为 $45\times\frac{3}{5}=27$.

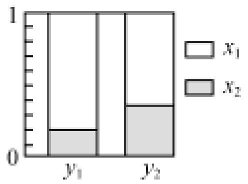
2. 观察如图所示的等高条形图，其中最有把握认为两个分类变量 $x$，$y$ 之间有关系的是(　D　)



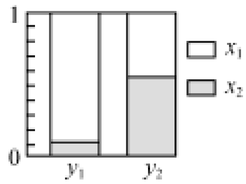

A




B




C




D

【解析】 在等高条形图中，$x_1$，$x_2$ 所占比例相差越大，分类变量 $x$，$y$ 有关系的把握越大．

3. 某村庄对该村内 50 名老年人、年轻人每年是否体检的情况进行调查，统计数据如下表

| | 每年体检 | 每年未体检 | 合计 |
|---|---|---|---|
| 老年人 | $a$ | 7 | $c$ |
| 年轻人 | 6 | $b$ | $d$ |
| 合计 | $e$ | $f$ | 50 |

已知抽取的老年人、年轻人各 25 名，则对列联表数据的分析错误的是(　D　)

A. $a=18$　　B. $b=19$

C. $c+d=50$　　D. $e-f=2$

【解析】 由题意得 $a+7=c=25,6+b=d=25$，$a+6=e$，$7+b=f$，$e+f=50$，所以 $a=18$，$b=19$，$c+d=50$，$e=24$，$f=26$，则 $e-f=-2$.

4. 在调查中发现 480 名男人中有 38 名患有色盲，520 名女人中有 6 名患有色盲．下列说法正确的是(　C　)

A. 男人、女人中患色盲的频率分别为 0.038 和 0.006

B. 男、女患色盲的概率分别为$\frac{19}{240}$，$\frac{3}{260}$

C. 男人中患色盲的比例比女人中患色盲的比例大，可以认为患色盲与性别是有关的

D. 调查人数太少，不能说明色盲与性别有关

【解析】 男人中患色盲的比例为$\frac{38}{480}\approx0.079$，女人中患色盲的比例$\frac{6}{520}\approx0.012$，其差值为$\left|\frac{38}{480}-\frac{6}{520}\right|\approx0.067\ 6$，差值较大，说明患色盲与性别有关.

5. 在对人们的休闲方式的一次调查中，共调查了 110 人，其中女性 50 人，男性 60 人．女性中有 30 人主要的休闲方式是看电视，另外 20 人主要的休闲方式是运动；男性中有 20 人主要的休闲方式是看电视，另外 40 人主要的休闲方式是运动．

(1) 根据以上数据建立一个 2×2 列联表；

(2) 判断性别与休闲方式是否有关系．

【解答】 (1) 2×2 列联表如下：

| | 看电视 | 运动 | 合计 |
|---|---|---|---|
| 女 | 30 | 20 | 50 |
| 男 | 20 | 40 | 60 |
| 合计 | 50 | 60 | 110 |

(2) 根据列联表中的数据，可得女性中休息方式为看电视的频率为$\frac{30}{50}=0.6$，男性中休息方式为看电视的频率为$\frac{20}{60}\approx 0.333$，二者差别较大，可知性别与休闲方式有关系.

## 第 2 课时　独立性检验

预习评价▸新知初探

### 学习目标

1. 了解独立性检验的基本思想、方法及其简单应用.

2. 能利用 2×2 列联表进行独立性检验，提升利用图表进行数据分析的能力.

3. 理解判断两个分类变量是否有关系的常用方法、独立性检验中 $\chi^2$ 的含义及其实施步骤.

### 问题导引

1. 什么是独立性检验？

2. 怎样进行独立性检验？

### 现学现用

1. 概念辨析(正确的打“√”，错误的打“×”)

(1) 事件 $A$ 与 $B$ 的独立性检验无关，即两个事件互不影响. (　×　)

(2) $\chi^2$ 的大小是判断事件 $A$ 与 $B$ 是否相关的统计量. (　√　)

(3) 事件 $X,Y$ 关系越密切，则由观测数据计算得到的 $\chi^2$ 的观测值越大. (　√　)

2. 给出下列实际问题：①一种药物对某种病的治愈率；②两种药物治疗同一种病是否有区别；③吸烟者得肺病的概率；④吸烟是否与性别有关系；⑤网吧与青少年的犯罪是否有关系. 其中用独立性检验可以解决的问题有(　B　)

A. ①②③　　　　B. ②④⑤

C. ②③④⑤　　　　　　　　D. ①②③④⑤

【解析】 ①③都是概率问题，不能用独立性检验解决．

3. 下列说法中不正确的是(　B　)

A. 独立性检验是检验两个分类变量是否有关的一种统计方法

B. 独立性检验得到的结论一定是正确的

C. 独立性检验的样本不同，其结论可能不同

D. 独立性检验的基本思想是带有概率性质的反证法

4. 为了考察高中生的性别与是否喜欢数学课程之间的关系，利用 2×2 列联表进行检验，经计算 $\chi^2$ 的观测值 $x=7.069$，参考下表，则认为“性别与是否喜欢数学课程有关”犯错误的概率不超过(　B　)

| $P(\chi^2\geqslant x_0)$ | 0.100 | 0.050 | 0.025 | 0.010 | 0.001 |
| --- | --- | --- | --- | --- | --- |
| $x_0$ | 2.706 | 3.841 | 5.024 | 6.635 | 10.828 |

A. 0.001　　　　　　　　B. 0.01

C. 0.99　　　　　　　　D. 0.999

【解析】 $x=7.069>6.635$，对照表格，则认为“性别与是否喜欢数学课程有关”犯错误的概率不超过 0.01.

## 对点练 ▶ 先掌握基础

### 类型 1　有关“相关的检验”

例1　某校对学生课外活动进行调查，结果整理成下表．请依据小概率值 $\alpha=0.005$ 的独立性检验，判断“喜欢体育还是文娱”与“性别”的关系.

| | 体育 | 文娱 | 总计 |
| --- | --- | --- | --- |
| 男生 | 21 | 23 | 44 |
| 女生 | 6 | 29 | 35 |
| 总计 | 27 | 52 | 79 |

附：临界值 $\chi_{0.005}=7.879$.

【解答】 判断方法如下：假设 $H_0$“喜欢体育还是喜欢文娱与性别没有关系”，若 $H_0$ 成立，则 $\chi^2$ 应该很小.

因为 $a=21$，$b=23$，$c=6$，$d=29$，$n=79$，

所以 $\chi^2=\dfrac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$

$=\dfrac{79\times(21\times29-23\times6)^2}{44\times35\times27\times52}$

$\approx8.106>7.879$，

故可以认为“喜欢体育还是喜欢文娱与性别有关”.

**变式** “中国式过马路”存在很大的交通安全隐患，某调查机构为了了解路人对“中国式过马路”的态度是否与性别有关，从马路旁随机抽取30名路人进行了问卷调查，得到了如下的2×2列联表．已知在这30人中随机抽取1人，抽到反感“中国式过马路”的路人的概率是$\frac{3}{5}$.

(1) 求2×2列联表中的 $b$，$c$ 的值；

| | 男性 | 女性 | 合计 |
|---|---|---|---|
| 反感 | 10 | $b$ | $10+b$ |
| 不反感 | $c$ | 8 | $c+8$ |
| 合计 | $10+c$ | $b+8$ | 30 |

(2) 根据列联表中的数据，依据小概率值 $\alpha=0.05$ 的独立性检验，分析“中国式过马路”与“性别”的关系.

附：临界值 $\chi_{0.05}=3.841$.

【解答】 (1) 因为在这30人中随机抽取1人，抽到反感“中国式过马路”的路人的概率是$\frac{3}{5}$，所以$\dfrac{10+b}{30}=\dfrac{3}{5}$，解得 $b=8$.

又 $10+b+c+8=30$，解得 $c=4$.

故 $b=8$，$c=4$.

(2) 补充完整的2×2列联表如下：

| | 男性 | 女性 | 合计 |
|---|---|---|---|

| | | | |
|---|---|---|---|
| 反感 | 10 | 8 | 18 |
| 不反感 | 4 | 8 | 12 |
| 合计 | 14 | 16 | 30 |

所以$\chi^2=\frac{30\times(10\times8-4\times8)^2}{14\times16\times18\times12}\approx1.429<3.841$，

故不可以认为反感“中国式过马路”与性别有关．

【规律总结】

1．利用$\chi^2=\frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$求出$\chi^2$的观测值$x_0$的值．再利用临界值的大小来判断假设是否成立．

2．解题时应注意准确代数与计算，不可错用公式，准确进行比较与判断．

类型 2　有关“无关的检验”

例2　为了探究学生选报文、理科是否与对外语的兴趣有关，某同学调查了 361 名高二在校学生，调查结果如下：理科对外语有兴趣的有 138 人，无兴趣的有 98 人，文科对外语有兴趣的有 73 人，无兴趣的有 52 人．依据小概率值 $\alpha=0.1$ 的独立性检验，判断学生选报文、理科与对外语的兴趣是否有关．附：临界值$\chi_{0.1}=2.706$.

【解答】　列出 2×2 列联表如下：

| | 理 | 文 | 总计 |
|---|---|---|---|
| 有兴趣 | 138 | 73 | 211 |
| 无兴趣 | 98 | 52 | 150 |
| 总计 | 236 | 125 | 361 |

所以$\chi^2=\frac{361\times(138\times52-73\times98)^2}{236\times125\times211\times150}\approx1.871\times10^{-4}$.

因为 $1.871\times10^{-4}<2.706$，所以可以认为学生选报文、理科与对外语的兴趣

无关．

【规律总结】 独立性检验的思想来自统计上的假设检验思想，它与反证法类似，它们都是先假设结论不成立，然后根据是否能推出“矛盾”来判定结论是否成立．但二者“矛盾”的含义不同，反证法中的“矛盾”是指不符合逻辑的事件发生；而假设检验中的“矛盾”是指不符合逻辑的小概率事件发生，即在结论不成立的假设下推出有利于结论成立的小概率事件的发生．

类型 3 独立性检验的综合问题

例3 新生儿某疾病要接种三次疫苗免疫(即 0,1,6 月龄)，假设每次接种之间互不影响，每人每次接种成功的概率相等．为了了解新生儿该疾病疫苗接种剂量与接种成功之间的关系，现进行了两种接种方案的临床试验：10 μg/次剂量组与 20 μg/次剂量组，试验结果如下表：

| | 接种成功 | 接种不成功 | 总计(人) |
|---|---|---|---|
| 10 μg/次剂量组 | 900 | 100 | 1 000 |
| 20μg/次剂量组 | 973 | 27 | 1 000 |
| 总计(人) | 1 873 | 127 | 2 000 |

(1) 根据数据判断哪种方案接种效果好，依据小概率值 $\alpha=0.001$ 的独立性检验，分析该疾病疫苗接种成功是否与两种接种方案有关．

(2) 以频率代替概率，若选用接种效果好的方案，参与该试验的 1 000 人的成功人数比此剂量只接种一次的成功人数平均提高多少人？

附：临界值 $\chi_{0.001}=10.828$.

【解答】 (1) 由于两种接种方案都是 1 000 人接受临床试验，接种成功人数 10 μg/次剂量组 900 人，20 μg/次剂量组 973 人，且 973＞900，所以方案 20 μg/次剂量组接种效果好．

计算 $\chi^2=\dfrac{2000\times(900\times27-100\times973)^2}{1\ 000\times1\ 000\times1\ 873\times127}\approx44.806>10.828$，故可认为该疾病疫苗接种成功与两种接种方案有关．

(2) 假设 20 μg/次剂量组临床试验接种一次成功的概率为 $p$，由数据知，三次接种成功的概率为$\frac{973}{1\ 000}=0.973$，

不成功的概率为$\frac{27}{1\ 000}=0.027$.

由于三次接种之间互不影响，每人每次接种成功的概率相等，所以$(1-p)^3=0.027$，解得 $p=0.7$.

设参与试验的 1 000 人此剂量只接种一次成功的人数为 $X$，则 $X\sim B(1\ 000, 0.7)$，$E(X)=1\ 000\times 0.7=700$，

参与试验的 1 000 人此剂量只接种一次成功的人数平均为 700 人，且 $973-700=273$，

所以试验选用 20 μg/次剂量组方案，参与该试验的 1 000 人比此剂量只接种一次的成功人数平均提高 273 人.

【规律总结】 独立性检验的关键：

(1) 根据 2×2 列联表准确计算 $\chi^2$，若 2×2 列联表没有列出来，要先列出此表.

(2) $\chi^2$ 的观测值 $x_0$ 越大，对应假设事件 $H_0$ 成立的概率越小，$H_0$ 不成立的概率越大.

课堂评价▶基础达标

1. (多选)对于分类变量 $X$ 与 $Y$ 的 $\chi^2$ 的值，下列说法错误的是(　ACD　)

A. $\chi^2$ 越大，“$X$ 与 $Y$ 有关系”的可信程度越小

B. $\chi^2$ 越小，“$X$ 与 $Y$ 有关系”的可信程度越小

C. $\chi^2$ 越接近于 0，“$X$ 与 $Y$ 没有关系”的可信程度越小

D. $\chi^2$ 越大，“$X$ 与 $Y$ 没有关系”的可信程度越大

【解析】 根据独立性检验的基本思想可知，分类变量 $X$ 与 $Y$ 的 $\chi^2$ 的观测值

越大，“$X$与$Y$没有关系”的可信程度越小，则“$X$与$Y$有关系”的可信程度越大；$\chi^2$越小，“$X$与$Y$有关系”的可信程度越小，“$X$与$Y$没有关系”的可信程度越大.

2. 如果根据性别与是否爱好运动的列联表得到$\chi^2\approx3.852>3.841$，所以判断性别与运动有关，那么这种判断犯错的可能性不超过(　D　)

| $P(\chi^2\geqslant x_0)$ | 0.050 | 0.010 | 0.001 |
| --- | --- | --- | --- |
| $x_0$ | 3.841 | 6.635 | 10.828 |

A. 2.5%　　　　B. 0.5%

C. 1%　　　　D. 5%

**【解析】** 因为$\chi^2\approx3.852>3.841$，所以由临界值表可得有95%的把握说性别与运动有关，即有$1-95\%=5\%$的犯错误的可能性.

3. 在吸烟与患肺病这两个分类变量的计算中，下列说法正确的是(　C　)

附表：

| $P(\chi^2\geqslant x_0)$ | 0.15 | 0.10 | 0.05 | 0.025 | 0.010 | 0.005 | 0.001 |
| --- | --- | --- | --- | --- | --- | --- | --- |
| $x_0$ | 2.072 | 2.706 | 3.841 | 5.024 | 6.635 | 7.879 | 10.828 |

A. 若$\chi^2$的观测值$x=6.635$，我们有99%的把握认为吸烟与患肺病有关系，那么在100个吸烟的人中必有99人患有肺病

B. 从独立性检验可知有99%的把握认为吸烟与患肺病有关系时，我们说某人吸烟，那么他有99%的可能性患有肺病

C. 从统计量中求出有95%的把握认为吸烟与患肺病有关系，是指有5%的可能性使得判断出现错误

D. 以上三种说法都不正确

4. 某大型企业人力资源部，为了研究企业员工工作积极性和对待企业改革态度的关系，随机抽取了189名员工进行调查，所得数据如下表所示.

| | 积极支持企业改革 | 不太赞成企业改革 | 总计 |
| --- | --- | --- | --- |
| 工作积极 | 54 | 40 | 94 |
| 工作一般 | 32 | 63 | 95 |
| 总计 | 86 | 103 | 189 |

对于人力资源部的研究项目，根据上述数据能得出什么结论？李明和张宇都对该题进行了独立性检验的分析，李明的结论是“在犯错误的概率不超过 0.01 的前提下，认为企业员工的工作积极性和对待企业改革的态度有关系”；张宇的结论是“在犯错误的概率不超过 0.005 的前提下，认为企业员工的工作积极性和对待企业改革的态度有关系.”

他们两人的结论正确吗？他们的结论为什么不一样？

【解答】 由题知 $\chi^2=\dfrac{189\times(54\times63-40\times32)^2}{94\times95\times86\times103}\approx10.759$，

10．759>7.879>6.635，

若以 $x_0=7.879$ 为临界值，则在犯错误的概率不超过 0.005 的前提下，认为企业员工的工作积极性和对待企业改革的态度有关系；

若以 $x_0=6.635$ 为临界值，则在犯错误的概率不超过 0.01 的前提下，认为企业员工的工作积极性和对待企业改革的态度有关系．所以李明和张宇的结论都正确．

造成结论不一样的原因是，他们两人采用了两种不同的判断规则，即所选用的临界值不同．

## 章复习　能力整合与素养提升

要点回顾▶连点成面

1. 两个变量的线性相关

(1) 从散点图上看，如果点分布在从左下角到右上角的区域内，那么两个变量的这种相关关系称为__正相关__；如果点分布在从左上角到右下角的区域内，那么两个变量的这种相关关系称为__负相关__.

(2) $r=\dfrac{\sum\limits_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\sum\limits_{i=1}^{n}(x_i-\overline{x})^2\sum\limits_{i=1}^{n}(y_i-\overline{y})^2}}$，当 $r>0$ 时，表示两个变量正相关；当 $r<0$ 时，表示两个变量负相关. $r$ 的绝对值越接近__1__，表示两个变量的线性相关性越强；$r$ 的绝对值越接近__0__，表示两个变量的线性相关性越弱. 通常当 $r$ 的绝对值大于 0.75 时，认为两个变量具有很强的线性相关关系.

2. 最小二乘法与一元线性回归模型

两个具有线性相关关系的变量的一组数据：$(x_1，y_1)$，$(x_2，y_2)$，…，$(x_n，y_n)$，其经验回归方程为$\hat{y}=\hat{b}x+\hat{a}$，则

$$\begin{cases}\hat{b}=\dfrac{\sum\limits_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sum\limits_{i=1}^{n}(x_i-\overline{x})^2}=\dfrac{\sum\limits_{i=1}^{n}x_iy_i-n\overline{x}\ \overline{y}}{\sum\limits_{i=1}^{n}x_i^2-n\overline{x}^2},\\ \hat{a}=\overline{y}-\hat{b}\overline{x},\end{cases}$$

其中$\overline{x}=\frac{1}{n}\sum\limits_{i=1}^{n}x_i$，$\overline{y}=\frac{1}{n}\sum\limits_{i=1}^{n}y_i$，__$(\overline{x}，\overline{y})$__称为样本点的中心.

3. 残差与残差分析

(1) 线性回归模型用$y=bx+a+e$表示，其中$a$和$b$为模型的未知参数，$e$称为随机误差.它的均值满足$E(e)=$__0__，$D(e)=\sigma^2$，$\sigma^2$越小，精度越__高__.

(2) 残差：$\hat{e}_i=y_i-\hat{y}_i$称为相应于点$(x_i，y_i)$的残差，残差平方和为$\sum\limits_{i=1}^{n}(y_i-\hat{y}_i)^2$.

(3) 决定系数$R^2=1-\dfrac{\sum\limits_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum\limits_{i=1}^{n}(y_i-\overline{y})^2}$. $R^2$越大，说明残差平方和__越小__，即模型的拟合效果__越好__；$R^2$越小，残差平方和__越大__，即模型的拟合效果__越差__.

4. 独立性检验

假设有两个分类变量$X$和$Y$，它们的可能取值分别为$\{x_1，x_2\}$和$\{y_1，y_2\}$，其样本频数列联表(称为2×2列联表)为

| | $y_1$ | $y_2$ | 总计 |
|---|---|---|---|
| $x_1$ | $a$ | $b$ | $a+b$ |
| $x_2$ | $c$ | $d$ | $c+d$ |
| 总计 | $a+c$ | $b+d$ | $a+b+c+d$ |

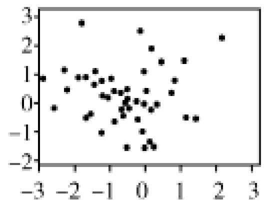
构造一个随机变量$\chi^2=\dfrac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$，如果$\chi^2$的观测值$x\geq x_0$，就认为“两个分类变量之间有关系”；否则就认为“两个分类变量之间没有关系”.我们称这样的$x_0$为一个判断规则的临界值. 按照上述规则，把“两个分类变量之

间没有关系”错误地判断为“两个分类变量之间有关系”的概率不超过$P(\chi^2\geq x_0)$．

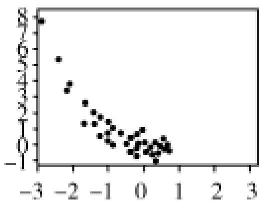
考法聚焦▶核心突破

考法 1　变量的相关性

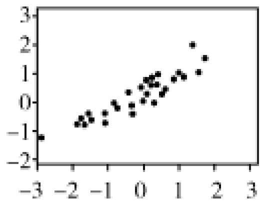
例1　下面 4 个散点图中，不适合线性回归模型拟合的两个变量是(　A　)



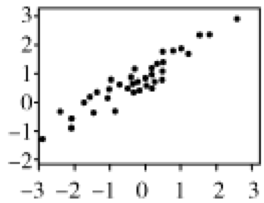
A



B



C



D

**【解析】** 根据题意，适合用线性回归拟合其中两个变量的散点图，必须是散点分布比较集中，且大体接近某一条直线，分析选项可知，A 中的散点杂乱无章，最不符合条件．

**【类题固法】**

1．已知变量 $x$，$y$，$z$ 都是正数，$y$ 与 $x$ 的经验回归方程为$\hat{y}=\hat{b}x+3$，且 $x$ 每增加 1 个单位，$y$ 减少 2 个单位，$y$ 与 $z$ 的经验回归方程为$\hat{y}=2z^2$，则(　D　)

A. $y$ 与 $x$ 正相关，$z$ 与 $x$ 正相关

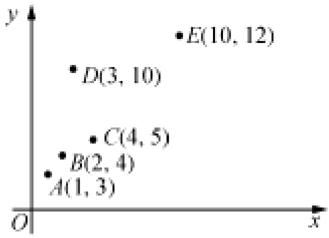
B. $y$ 与 $x$ 正相关，$z$ 与 $x$ 负相关

C. $y$ 与 $x$ 负相关，$z$ 与 $x$ 正相关

D. $y$ 与 $x$ 负相关，$z$ 与 $x$ 负相关

**【解析】** 因为 $x$ 每增加 1 个单位，$y$ 减少 2 个单位，所以$\hat{b}=-2$，所以 $y$ 与 $x$ 负相关．又 $y$，$z$ 都是正数且$\hat{y}=2z^2$，所以 $y$ 与 $z$ 正相关，所以 $z$ 与 $x$ 负相关．

2．如图，5 组数据$(x，y)$中去掉点 $D(3,10)$后，下列说法正确的是(　D　)




A. 偏差平方和变大

B. 相关系数 $r$ 变小

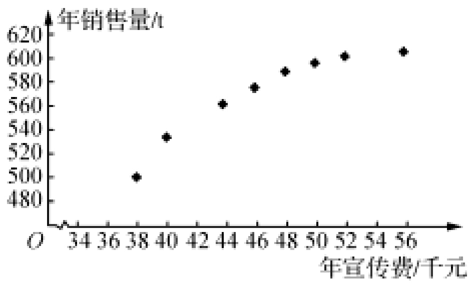
C. 负相关变为正相关

D. 解释变量 $x$ 与预报变量 $y$ 的相关性变强

**【解析】** **由散点图知，去掉点 $D(3,10)$后，$y$ 与 $x$ 的线性相关性加强，且为正相关，所以 $r$ 变大，$R^2$ 变大，偏差平方和变小．**

考法 2　回归分析

**例2**　某公司为确定下一年度投入某种产品的宣传费，需了解年宣传费 $x$(单位：千元)对年销售量 $y$(单位：t)和年利润 $z$(单位：千元)的影响，对近 8 年的年宣传费 $x_i$ 和年销售量 $y_i(i=1,2，…，8)$数据作了初步处理，得到下面的散点图及一些统计量的值．




| $\overline{x}$ | $\overline{y}$ | $\overline{w}$ | $\sum\limits_{i=1}^{8}(x_i-\overline{x})^2$ |
| --- | --- | --- | --- |
| 46.6 | 563 | 6.8 | 289.8 |

| $\sum\limits_{i=1}^{8}(w_i-\overline{w})^2$ | $\sum\limits_{i=1}^{8}(x_i-\overline{x})\cdot(y_i-\overline{y})$ | $\sum\limits_{i=1}^{8}(w_i-\overline{w})\cdot(y_i-\overline{y})$ |
| --- | --- | --- |
| 1.6 | 1 469 | 108.8 |

表中 $w_i=\sqrt{x_i}$，$\overline{w}=\frac{1}{8}\sum\limits_{i=1}^{8}w_i$.

(1) 根据散点图判断，$y=a+bx$ 与 $y=c+d\sqrt{x}$哪一个适宜作为年销售量 $y$ 关

于年宣传费 $x$ 的回归方程类型；(给出判断即可，不必说明理由)

(2) 根据(1)的判断结果及表中数据，建立 $y$ 关于 $x$ 的经验回归方程；

(3) 已知这种产品的年利润 $z$ 与 $x$，$y$ 的关系为 $z=0.2y-x$. 根据(2)的结果回答下列问题：

①当年宣传费 $x=49$ 时，年销售量及年利润的预报值是多少？

②当年宣传费 $x$ 为何值时，年利润的预报值最大？

**【解答】** (1) 由散点图可以判断，$y=c+d\sqrt{x}$适宜作为年销售量 $y$ 关于年宣传费 $x$ 的回归方程类型.

(2) 令 $w=\sqrt{x}$，先建立 $y$ 关于 $w$ 的经验回归方程，由于

$$\hat{d}=\frac{\sum_{i=1}^{n}(w_i-\overline{w})\cdot(y_i-\overline{y})}{\sum_{i=1}^{n}(w_i-\overline{w})^2}=\frac{108.8}{1.6}=68,$$

$$\hat{c}=\overline{y}-\hat{d}\,\overline{w}=563-68\times6.8=100.6,$$

所以 $y$ 关于 $w$ 的经验回归方程为$\hat{y}=100.6+68w$，因此 $y$ 关于 $x$ 的回归方程为$\hat{y}=100.6+68\sqrt{x}$.

(3) ①由(2)知，当 $x=49$ 时，年销售量 $y$ 的预报值

$$\hat{y}=100.6+68\sqrt{49}=576.6,$$

年利润 $z$ 的预报值$\hat{z}=576.6\times0.2-49=66.32$.

②根据(2)的结果知，年利润 $z$ 的预报值

$$\hat{z}=0.2(100.6+68\sqrt{x})-x=-x+13.6\sqrt{x}+20.12,$$

所以当$\sqrt{x}=\frac{13.6}{2}=6.8$，即 $x=46.24$ 时，$\hat{z}$取得最大值.

故年宣传费为 46.24 千元时，年利润的预报值最大.

**【类题固法】**

1. 为了了解某社区居民的家庭年收入与年支出的关系，随机调查了该社区 5 户家庭，得到如下统计数据表：

| 收入 $x$/万元 | 8.2 | 8.6 | 10.0 | 11.3 | 11.9 |
|---|---|---|---|---|---|
| 支出 $y$/万元 | 6.2 | 7.5 | 8.0 | 8.5 | 9.8 |

根据上表可得经验回归方程$\hat{y}=\hat{b}x+\hat{a}$，其中$\hat{b}=0.76$，$\hat{a}=\overline{y}-\hat{b}\overline{x}$.据此估计，该社区一户年收入为 15 万元家庭的年支出为(　B　)

A. 11.4 万元　　　　B. 11.8 万元

C. 12.0 万元　　　　D. 12.2 万元

**【解析】** 回归直线一定过样本点中心(10,8)，因为$\hat{b}=0.76$，所以$\hat{a}=0.4$.由$\hat{y}=0.76x+0.4$，得当 $x=15$ 万元时，$\hat{y}=11.8$ 万元.

2. 根据如下样本数据得到的经验回归方程为$\hat{y}=bx+a$，则(　B　)

| $x$ | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|
| $y$ | 4.0 | 2.5 | $-0.5$ | 0.5 | $-2.0$ | $-3.0$ |

A. $a>0$，$b>0$　　　　B. $a>0$，$b<0$

C. $a<0$，$b>0$　　　　D. $a<0$，$b<0$

**【解析】** 把样本数据中的 $x$，$y$ 分别当作点的横、纵坐标，在平面直角坐标系 $xOy$ 中作出散点图，由图可知 $b<0$，$a>0$.

3. 为了解春季昼夜温差大小与某种子发芽多少之间的关系，现在从 4 月份的 30 天中随机挑选了 5 天进行研究，且分别记录了每天昼夜温差与每天 100 颗种子浸泡后的发芽数，得到如下表格：

| 日期 | 4 月 1 日 | 4 月 7 日 | 4 月 15 日 | 4 月 21 日 | 4 月 30 日 |
|---|---|---|---|---|---|
| 温差 $x$/℃ | 10 | 11 | 13 | 12 | 8 |
| 发芽数 $y$/颗 | 23 | 25 | 30 | 26 | 16 |

(1) 从这 5 天中任选 2 天，记发芽的种子数分别为 $m$，$n$，求事件“$m$，$n$ 均不小于 25”的概率；

(2) 从这 5 天中任选 2 天，若选取的是 4 月 1 日与 4 月 30 日的两组数据，请根据这 5 天中的另 3 天的数据，求出 $y$ 关于 $x$ 的经验回归方程$\hat{y}=\hat{b}x+\hat{a}$；

(3) 若由经验回归方程得到的估计数据与所选出的检验数据的误差均不超过 2 颗，则认为得到的经验回归方程是可靠的，试问(2)中所得的经验回归方程是否

可靠？

【解答】 (1) 所有的样本点为(23,25)，(23,30)，(23,26)，(23,16)，(25,30)，(25,26)，(25,16)，(30,26)，(30,16)，(26,16)，共 10 个.

设“$m$，$n$ 均不小于 25”为事件 $A$，则事件 $A$ 包含的样本点为(25,30)，(25,26)，(30,26)，共 3 个，所以 $P(A)=\frac{3}{10}$.

(2) 由数据得，另 3 天的平均数 $\overline{x}=12$，$\overline{y}=27$，$3\overline{x}\,\overline{y}=972$，

$3\overline{x}^2=432$，$\sum\limits_{i=1}^{n}x_iy_i=977$，$\sum\limits_{i=1}^{n}x_i^2=434$，

所以 $\hat{b}=\frac{977-972}{434-432}=\frac{5}{2}$，$\hat{a}=27-\frac{5}{2}\times12=-3$，

所以 $y$ 关于 $x$ 的经验回归方程为 $\hat{y}=\frac{5}{2}x-3$.

(3) 依题意得，当 $x=10$ 时，$\hat{y}=22$，$|22-23|<2$；

当 $x=8$ 时，$\hat{y}=17$，$|17-16|<2$，

所以(2)中所得到的经验回归方程是可靠的.

考法 3　独立性检验

例3　某高中学校为了解高二年级学生在 2021 年高考和中考期间居家学习的自制力，随机抽取了 100 名学生，请他们的家长(每名学生请一位家长)对学生打分，满分为 10 分. 下表是家长所打分数 $x$ 的频数统计：

| 分数 $x$ | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|
| 频数 | 5 | 15 | 20 | 25 | 20 | 15 |

(1) 求家长所打分数的平均值 $\overline{x}$；

(2) 在抽取的 100 位学生中，男同学共 50 人，其中打分不低于 8 分的男同学为 20 人，填写列联表. 若打分不低于 8 分认为“自制力强”，打分低于 8 分认为“自制力一般”，依据小概率值 $\alpha=0.001$ 的独立性检验，判断高二年级学生的性别与自制力的强弱是否有关联；如果结论是性别与自制力的强弱有关联，请解释它们如何相互影响.

【解答】 (1) 家长所打分数的平均值为 $\overline{x}=\frac{1}{100}\times(5\times5+6\times15+7\times20+8\times25+9\times20+10\times15)=7.85$.

(2) 列联表如下：

| 性别 $X$ | 自制力 $Y$ | | 合计 |
|---|---|---|---|
| | 不低于 8 分 | 低于 8 分 | |
| 男 | 20 | 30 | 50 |
| 女 | 40 | 10 | 50 |
| 合计 | 60 | 40 | 100 |

设 $H_0$：分类变量 $X$ 与 $Y$ 相互独立，即性别与自制力的强弱之间无关联．根据列联表中的数据，

得 $\chi^2=\frac{100\times(20\times10-30\times40)^2}{60\times40\times50\times50}=\frac{50}{3}>10.828=x_{0.001}$.

依据小概率值 $\alpha=0.001$ 的独立性检验，推断 $H_0$ 不成立，即认为性别与自制力的强弱之间有关联，该推断犯错误的概率不超过 0.001.

男生中"自制力强"和"自制力一般"的频率分别为 $\frac{20}{50}=0.4$ 和 $\frac{30}{50}=0.6$；

女生中"自制力强"和"自制力一般"的频率分别为 $\frac{40}{50}=0.8$ 和 $\frac{10}{50}=0.2$.

由 $\frac{0.8}{0.4}=2$，$\frac{0.6}{0.2}=3$，可见，女生"自制力强"的频率是男生的 2 倍，男生"自制力一般"的频率是女生的 3 倍，于是，根据频率稳定于概率的原理，可以认为女生自制力强的概率明显大于男生自制力强的概率，即女生自制力更强．

【类题固法】

1. 在吸烟与患肺癌这两个分类变量的独立性检验的计算中，下列说法正确的是(　C　)

A. 若 $\chi^2$ 的观测值为 $x=6.635$，在犯错误的概率不超过 0.01 的前提下认为吸

烟与患肺癌有关系，那么在 100 个吸烟的人中必有 99 人患有肺癌

B. 由独立性检验可知，在犯错误的概率不超过 0.01 的前提下认为吸烟与患肺癌有关系时，我们说某人吸烟，那么他有 99%的可能患有肺癌

C. 若从统计量中求出在犯错误的概率不超过 0.01 的前提下认为吸烟与患肺癌有关系，是指有 1%的可能性使得判断出现错误

D. 以上三种说法都不正确

2. 某研究性学习小组调查研究学生使用智能手机对学习的影响，部分统计数据如下表：

| | 使用智能手机 | 不使用智能手机 | 合计 |
|---|---|---|---|
| 成绩优秀 | 4 | 8 | 12 |
| 成绩不优秀 | 16 | 2 | 18 |
| 合计 | 20 | 10 | 30 |

| $P(\chi^2 \geqslant x_0)$ | 0.15 | 0.10 | 0.05 | 0.025 | 0.010 | 0.005 | 0.001 |
|---|---|---|---|---|---|---|---|
| $x_0$ | 2.072 | 2.706 | 3.841 | 5.024 | 6.635 | 7.879 | 10.828 |

则下列选项正确的是(  A  )

A. 有 99.5%的把握认为使用智能手机对学习有影响

B. 有 99.5%的把握认为使用智能手机对学习无影响

C. 有 99.9%的把握认为使用智能手机对学习有影响

D. 有 99.9%的把握认为使用智能手机对学习无影响

【解析】 依题意 $\chi^2=\dfrac{30\times(4\times2-16\times8)^2}{12\times18\times20\times10}=10>7.879$，故有 99.5%的把握认为使用智能手机对学习有影响．

3. 为了判断高三年级学生是否选修历史与性别的关系，现随机抽取 50 名学生，得到如下 2×2 列联表：

| | 理科 | 文科 | 总计 |
|---|---|---|---|
| 男 | 13 | 10 | 23 |
| 女 | 7 | 20 | 27 |
| 总计 | 20 | 30 | 50 |

已知 $P(\chi^2\geqslant 3.841)\approx 0.05$，$P(\chi^2\geqslant 5.024)\approx 0.025$.根据表中数据，得到 $\chi^2$ 的观测值 $x=\dfrac{50\times(13\times 20-10\times 7)^2}{23\times 27\times 20\times 30}\approx 4.844$.则认为选修历史与性别有关系出错的可能性为 5%.

**【解析】** $\chi^2$ 的观测值为 4.844，这表明小概率事件发生．根据假设检验的基本原理，应该断定“是否选修历史与性别之间有关系”成立，并且这种判断出错的可能性约为 5%.

4. 心理学家分析发现视觉和空间想象能力与性别有关，某数学兴趣小组为了验证这个结论，从所在学校中按分层随机抽样的方法抽取 50 名同学(男 30，女 20)，给所有同学几何题和代数题各一题，让各位同学自由选择一道题进行解答．选题情况如下表：(单位：人)

| | 几何题 | 代数题 | 总计 |
| --- | --- | --- | --- |
| 男同学 | 22 | 8 | 30 |
| 女同学 | 8 | 12 | 20 |
| 总计 | 30 | 20 | 50 |

根据上述数据，推断视觉和空间想象能力与性别有关系，则这种推断犯错误的概率不超过 0.025.

附表：

| $P(\chi^2\geqslant x_0)$ | 0.15 | 0.10 | 0.05 | 0.025 | 0.010 | 0.005 | 0.001 |
| --- | --- | --- | --- | --- | --- | --- | --- |
| $x_0$ | 2.072 | 2.706 | 3.841 | 5.024 | 6.635 | 7.879 | 10.828 |

**【解析】** 由列联表计算 $\chi^2$ 的观测值 $x=\dfrac{50\times(22\times 12-8\times 8)^2}{30\times 20\times 20\times 30}\approx 5.556>5.024$，所以推断犯错误的概率不超过 0.025.