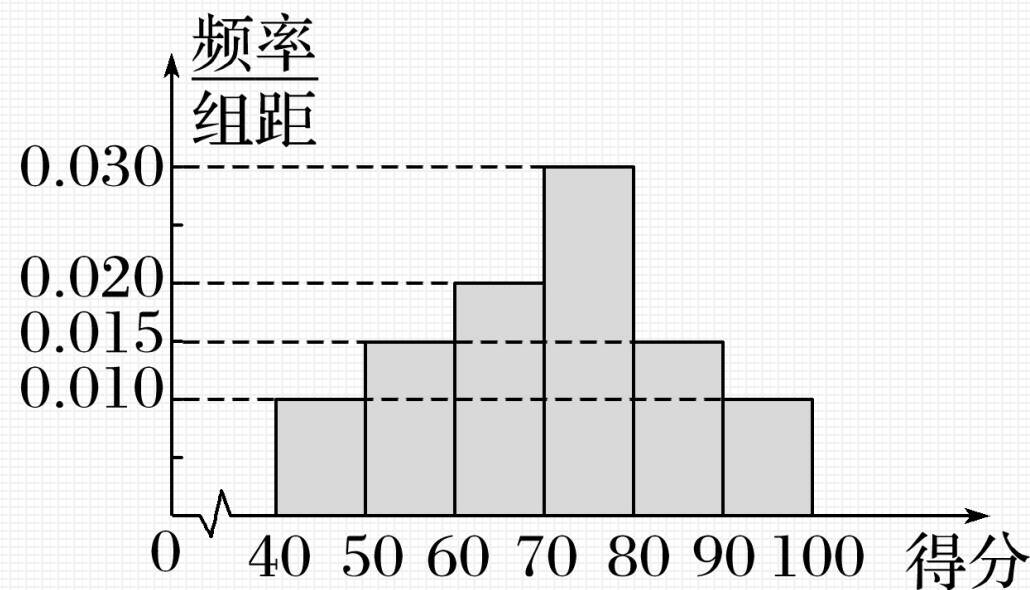


§10.8 概率与统计 的综合问题

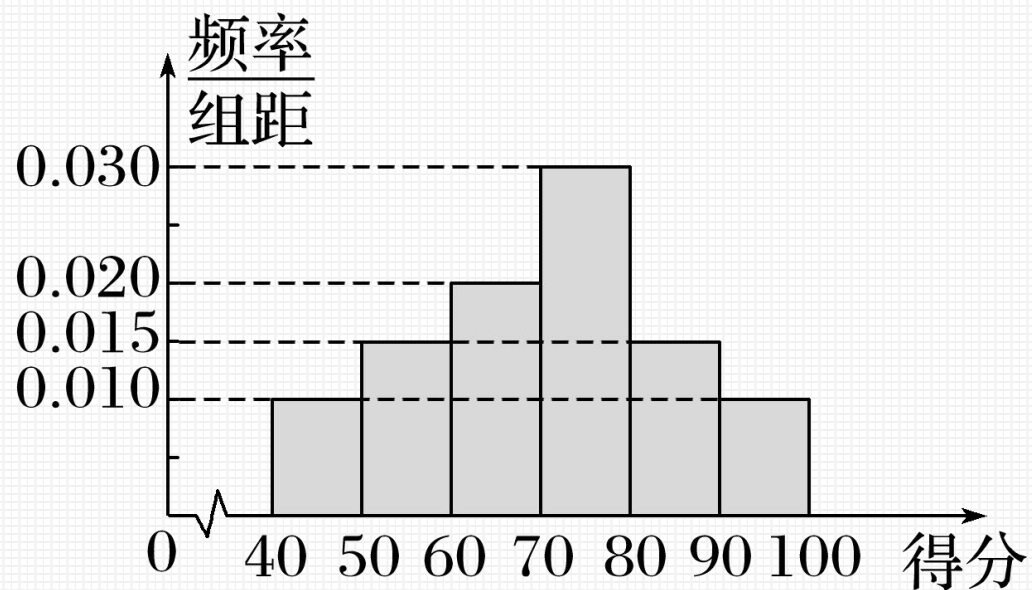
例1 2022年是中国共产主义青年团成立100周年，为引导和带动青少年重温共青团百年光辉历程，某校组织全体学生参加共青团百年历史知识竞赛，现从中随机抽取了100名学生的成绩组成样本，并将得分分成以下6组： $[40,50)$ ， $[50,60)$ ， $[60,70)$ ， $\cdots$ ， $[90,100]$ ，统计结果如图所示.

(1)试估计这100名学生得分的平均数；

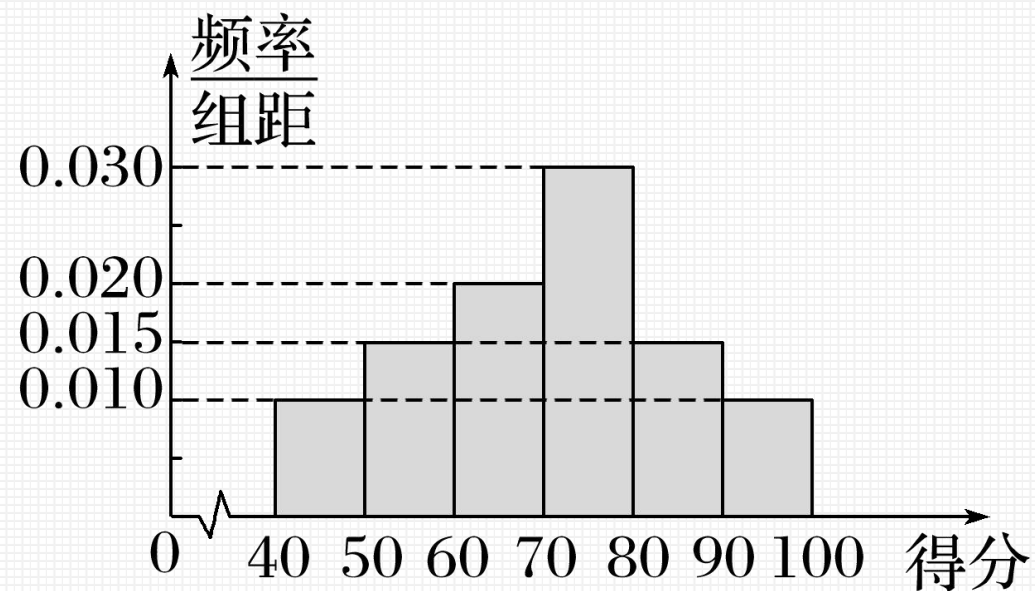


解

估计这100名学生得分的平均数为 $10 \times (45 \times 0.010 + 55 \times 0.015 + 65 \times 0.020 + 75 \times 0.030 + 85 \times 0.015 + 95 \times 0.010) = 70.5$.



(2)从样本中得分不低于70分的学生中，用比例分配的分层随机抽样的方法选取11人进行座谈，若从座谈名单中随机抽取3人，记其得分在 $[90,100]$ 的人数为 ξ ，试求 ξ 的分布列和均值；



解

从样本中得分不低于70分的学生中，用比例分配的分层随机抽样的方法选取11人进行座谈，

其中得分在[90,100]的人数为 $\frac{0.010}{0.015+0.030+0.010} \times 11 = 2$.

若从座谈名单中随机抽取3人，记其得分在[90,100]的人数为 ξ ，则 ξ 的所有可能取值为0,1,2.

$$P(\xi=0) = \frac{C_9^3}{C_{11}^3} = \frac{28}{55},$$

解

$$P(\xi=1)=\frac{C_9^2C_2^1}{C_{11}^3}=\frac{24}{55},$$

$$P(\xi=2)=\frac{C_9^1C_2^2}{C_{11}^3}=\frac{3}{55},$$

则 ξ 的分布列为

ξ	0	1	2
P	$\frac{28}{55}$	$\frac{24}{55}$	$\frac{3}{55}$

$$\text{所以 } E(\xi)=0\times\frac{28}{55}+1\times\frac{24}{55}+2\times\frac{3}{55}=\frac{6}{11}.$$

(3)以样本估计总体，根据频率分布直方图，可以认为参加知识竞赛的学生的得分 X 近似地服从正态分布 $N(\mu, \sigma^2)$ ，其中 μ 近似为样本平均数， σ^2 近似为样本方差 s^2 ，经计算 $s^2=42.25$.现从所有参加知识竞赛的学生中随机抽取500人，若这500名学生的得分相互独立，试问得分高于77分的人数最有可能是多少？

参考数据：若随机变量 $X \sim N(\mu, \sigma^2)$ ，则 $P(\mu - \sigma \leq X \leq \mu + \sigma) \approx 0.6827$ ， $P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \approx 0.9545$ ， $P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) \approx 0.9973$.

解

由题意知, $\mu=70.5$, $\sigma^2=s^2=42.25$, $\sigma=6.5$.

$$P(X>77)=P(X>\mu+\sigma)=\frac{1-P(\mu-\sigma\leqslant X\leqslant \mu+\sigma)}{2}\approx 0.158\ 65,$$

所以这500名学生得分高于77分的人数最有可能为 $0.158\ 65\times 500\approx 79$.

高考常将频率分布直方图与分布列等交汇在一起进行考查，解题时要正确理解频率分布直方图，能利用频率分布直方图正确计算出各组数据.概率问题以计算为主，往往和实际问题相结合，要注意理解实际问题的意义，使之和相应的概率计算对应起来.

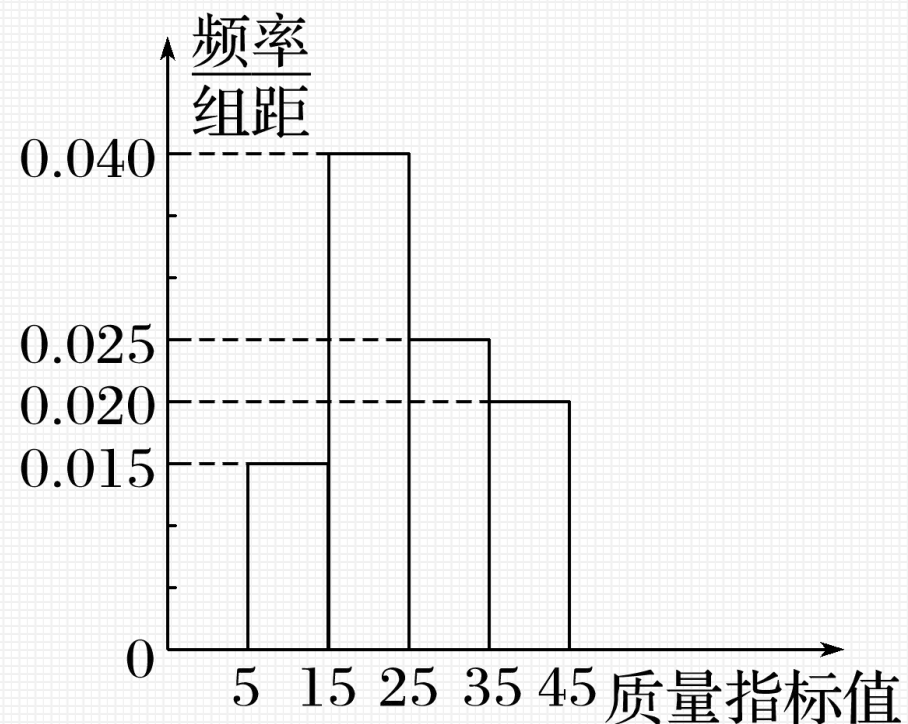
跟踪训练1 (2023·济南模拟)从某企业的某种产品中随机抽取100件，测量这些产品的一项质量指标值，由测量结果制成如图所示的频率分布直方图.

(1)求这100件产品质量指标值的样本平均数 $\bar{x}$ (同一组数据用该区间的中点值作代表);

解

$$\text{由已知得, } \bar{x} = 10 \times 0.015 \times 10 + 20 \times 0.040 \times 10 + 30 \times 0.025 \times 10 + 40 \times 0.020 \times 10 = 25.$$

(2) 已知某用户从该企业购买了3件该产品，用 X 表示这3件产品中质量指标值位于 $[35, 45]$ 内的产品件数，用频率估计概率，求 X 的分布列.



解

因为购买一件产品，其质量指标值位于 $[35,45]$ 内的概率为0.2，

所以 $X \sim B(3,0.2)$ ，因为 X 的所有可能取值为0,1,2,3，

所以 $P(X=0)=(1-0.2)^3=0.512$ ，

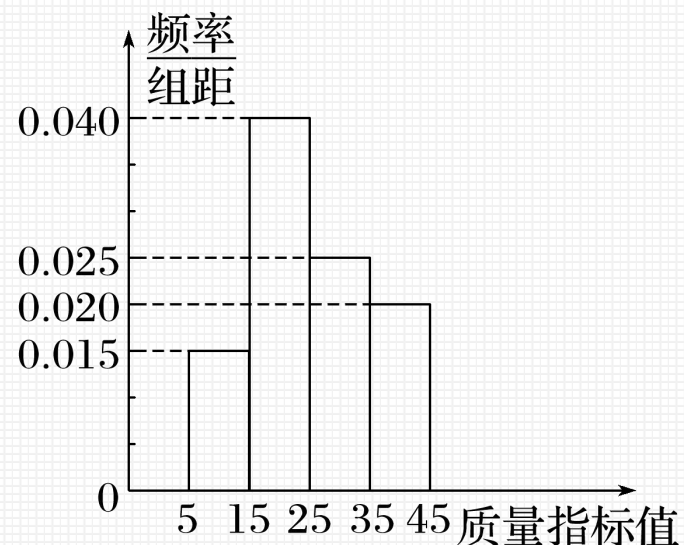
$P(X=1)=C_3^1 \times 0.2 \times (1-0.2)^2=0.384$ ，

$P(X=2)=C_3^2 \times 0.2^2 \times (1-0.2)=0.096$ ，

$P(X=3)=0.2^3=0.008$ ，

所以 X 的分布列为

X	0	1	2	3
P	0.512	0.384	0.096	0.008



例2 (2022·德州模拟)工信部发布的《“十四五”促进中小企业发展规划》中明确提出建立“百十万千”的中小企业梯度培育体系，引导中小企业走向“专精特新”“小巨人”“隐形冠军”的发展方向，“专精特新”是指具备专业化、精细化、特色化、新颖化优势的中小企业.下表是某地2017—2021年新增企业数量的有关数据：

年份(年)	2017	2018	2019	2020	2021
年份代码(x)	1	2	3	4	5
新增企业数量(y)	8	17	29	24	42

年份(年)	2017	2018	2019	2020	2021
年份代码(x)	1	2	3	4	5
新增企业数量(y)	8	17	29	24	42

(1)请根据表中所给的数据，求出y关于x的经验回归方程，并预测2023年此地新增企业的数量；

参考公式：经验回归方程 $\hat{y}=\hat{a}+\hat{b}x$ 中，

斜率和截距最小二乘估计公式分别为 $\hat{b}=\frac{\sum_{i=1}^n (x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^n (x_i-\bar{x})^2}$ ， $\hat{a}=\bar{y}-\hat{b}\bar{x}$.

解

$$\bar{x} = \frac{1+2+3+4+5}{5} = 3,$$

$$\bar{y} = \frac{8+17+29+24+42}{5} = 24,$$

$$\sum_{i=1}^5 (x_i - \bar{x})(y_i - \bar{y}) = (-2) \times (-16) + (-1) \times (-7) + 0 \times 5 + 1 \times 0 + 2 \times 18 = 75,$$

$$\sum_{i=1}^5 (x_i - \bar{x})^2 = 4 + 1 + 0 + 1 + 4 = 10,$$

解

$$\text{所以 } \hat{b} = \frac{\sum_{i=1}^5 (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^5 (x_i - \bar{x})^2} = 7.5, \text{ 则 } \hat{a} = \bar{y} - \hat{b}\bar{x} = 1.5,$$

$$\text{所以 } \hat{y} = 1.5 + 7.5x,$$

预测 2023 年, 即当 $x=7$ 时, 由经验回归方程可得 $\hat{y}=54$,

所以估计 2023 年此地新增企业的数量约为 54 家.

(2)若在此地进行考察，考察企业中有4个为“专精特新”企业，3个企业中为普通企业，现从这7个企业中随机抽取3个，用 X 表示抽取的3个企业中为“专精特新”企业的个数，求随机变量 X 的分布列与均值.

解

由题意可知, X 的所有可能取值为0,1,2,3,

$$\text{则 } P(X=0)=\frac{C_3^3}{C_7^3}=\frac{1}{35}, \quad P(X=1)=\frac{C_4^1 C_3^2}{C_7^3}=\frac{12}{35},$$

$$P(X=2)=\frac{C_4^2 C_3^1}{C_7^3}=\frac{18}{35}, \quad P(X=3)=\frac{C_4^3}{C_7^3}=\frac{4}{35},$$

所以 X 的分布列为

X	0	1	2	3
P	$\frac{1}{35}$	$\frac{12}{35}$	$\frac{18}{35}$	$\frac{4}{35}$

$$\text{所以 } E(X)=0 \times \frac{1}{35} + 1 \times \frac{12}{35} + 2 \times \frac{18}{35} + 3 \times \frac{4}{35} = \frac{12}{7}.$$

高考常将回归模型与分布列等交汇在一起进行考查，求经验回归方程时要充分利用已知数据，合理利用公式减少运算.求解概率问题时要注意概率模型的应用，明确所求问题所属的事件类型是关键.

跟踪训练2 (2023·三明模拟)2022年，中国新能源汽车销售火爆， A 省相关部门调查了该省2022年1月份至10月份的新能源汽车销量情况，得到一组样本数据 $(x_i, y_i)(i=1,2, \cdots, 10)$ ，其中 x_i 表示第 i 个月， y_i 表示第 i 个月 A 省新能源汽车的销量(单位：万辆)，由样本数据的散点图可知， y 与 x 具有线性相关关系，并将这10个月的数据作了初步处理，得到下面一些统计量的值：

$\overline{y}$	$\sum_{i=1}^{10} x_i y_i$	$\sum_{i=1}^{10} x_i^2$	$\sum_{i=1}^{10} y_i$
1.5	89.1	385	15

(1)建立 y 关于 x 的经验回归方程，并估计 A 省12月份新能源汽车的销量；

附：对于一组数据 $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ ，其经验回归方程 $\hat{y} = \hat{a}$

$+ \hat{b}x$ 的斜率和截距的最小二乘估计分别为 $\hat{b} = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2}$ ， $\hat{a} = \bar{y} - \hat{b} \bar{x}$.

解

$$\text{由题意得, } \bar{x} = \frac{1+2+3+\cdots+9+10}{10} = 5.5,$$

$$\text{又 } \bar{y} = 1.5, \quad \sum_{i=1}^{10} x_i y_i = 89.1, \quad \sum_{i=1}^{10} x_i^2 = 385,$$

$$\text{所以 } \hat{b} = \frac{\sum_{i=1}^{10} x_i y_i - 10 \bar{x} \bar{y}}{\sum_{i=1}^{10} x_i^2 - 10 \bar{x}^2} = \frac{89.1 - 10 \times 5.5 \times 1.5}{385 - 10 \times 5.5^2} = 0.08,$$

$$\hat{a} = 1.5 - 0.08 \times 5.5 = 1.06,$$

解

则 y 关于 x 的经验回归方程为 $\hat{y} = 1.06 + 0.08x$,

当 $x = 12$ 时, $\hat{y} = 2.02$,

故 A 省 12 月份新能源汽车的销量约为 2.02 万辆.

(2)为鼓励新能源汽车销售商积极参与调查， A 省汽车行业协会针对新能源汽车销售商开展抽奖活动，所有费用由某新能源汽车厂商赞助.奖项共设一、二、三等奖共三个奖项，其中一、二、三等奖分别奖励2万元、1万元、5千元，抽中一、二、三等奖的概率分别为 $\frac{1}{6}$ ， $\frac{1}{3}$ ， $\frac{1}{2}$. 现有甲、乙两家汽车销售商参加了抽奖活动，假设他们是否中奖相互独立，求这两家汽车销售商所获奖金总额 X (单位：万元)的分布列及均值.

解

这两家汽车销售商所获得的奖金总额 X 的所有可能取值为4,3,2.5,2,1.5,1,

$$P(X=4)=\frac{1}{6}\times\frac{1}{6}=\frac{1}{36},$$

$$P(X=3)=2\times\frac{1}{6}\times\frac{1}{3}=\frac{1}{9},$$

$$P(X=2.5)=2\times\frac{1}{6}\times\frac{1}{2}=\frac{1}{6},$$

$$P(X=2)=\frac{1}{3}\times\frac{1}{3}=\frac{1}{9},$$

$$P(X=1.5)=2\times\frac{1}{3}\times\frac{1}{2}=\frac{1}{3},$$

解

$$P(X=1)=\frac{1}{2}\times\frac{1}{2}=\frac{1}{4},$$

则 X 的分布列为

X	4	3	2.5	2	1.5	1
P	$\frac{1}{36}$	$\frac{1}{9}$	$\frac{1}{6}$	$\frac{1}{9}$	$\frac{1}{3}$	$\frac{1}{4}$

$$E(X)=4\times\frac{1}{36}+3\times\frac{1}{9}+2.5\times\frac{1}{6}+2\times\frac{1}{9}+1.5\times\frac{1}{3}+1\times\frac{1}{4}=\frac{11}{6}.$$

例3 (2023·滨州模拟)新能源汽车是指除汽油、柴油发动机之外的所有其他能源汽车，有利于减少空气污染和缓解能源短缺的压力.在当今提倡全球环保的前提下，新能源汽车越来越受到消费者的青睐，新能源汽车产业也必将成为未来汽车产业发展的导向与目标.某车企随机调查了今年3月份购买本车企生产的汽车的100位车主，经统计其购车种类与性别情况如表所示(单位：人).

性别	购车种类		合计
	购置新能源汽车	购置传统燃油汽车	
男性	50	10	60
女性	25	15	40
合计	75	25	100

(1)根据小概率值 $\alpha=0.05$ 的独立性检验，分析购车种类是否与性别有关；

附：
$$\chi^2=\frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}, \quad n=a+b+c+d.$$

α	0.1	0.05	0.01	0.005	0.001
x_α	2.706	3.841	6.635	7.879	10.828

解

零假设为 H_0 : 购车种类与性别无关,

根据表中数据可得 $\chi^2 = \frac{100 \times (15 \times 50 - 25 \times 10)^2}{75 \times 25 \times 60 \times 40} = \frac{50}{9} \approx 5.556 > 3.841 = \chi_{0.05}$,

根据小概率值 $\alpha=0.05$ 的独立性检验, 我们推断 H_0 不成立, 即认为购车种类与性别有关.

(2)用样本估计总体，用本车企售出汽车样本的频率代替售出汽车的概率，从该车企今年3月份售出的汽车中，随机抽取3辆汽车，设被抽取的3辆汽车中属于传统燃油汽车的辆数为 X ，求 X 的分布列及均值.

解

随机抽取 1 辆汽车属于传统燃油汽车的概率为 $\frac{25}{100} = \frac{1}{4}$,

设被抽取的 3 辆汽车中属于传统燃油汽车的辆数为 X , 则 X 的所有可能取值为 0, 1, 2, 3,

依题意得, $X \sim B\left(3, \frac{1}{4}\right)$,

$$P(X=0) = C_3^0 \times \left(\frac{1}{4}\right)^0 \times \left(\frac{3}{4}\right)^3 = \frac{27}{64},$$

$$P(X=1) = C_3^1 \times \left(\frac{1}{4}\right)^1 \times \left(\frac{3}{4}\right)^2 = \frac{27}{64},$$

解

$$P(X=2)=C_3^2 \times \left(\frac{1}{4}\right)^2 \times \left(\frac{3}{4}\right)^1 = \frac{9}{64},$$

$$P(X=3)=C_3^3 \times \left(\frac{1}{4}\right)^3 \times \left(\frac{3}{4}\right)^0 = \frac{1}{64},$$

所以 X 的分布列为

X	0	1	2	3
P	$\frac{27}{64}$	$\frac{27}{64}$	$\frac{9}{64}$	$\frac{1}{64}$

$$\text{则 } E(X)=3 \times \frac{1}{4} = \frac{3}{4}.$$

高考常将独立性检验与分布列等交汇在一起进行考查，解决独立性检验问题，要注意过好“三关”：假设关、公式关、对比关.解决概率问题要准确地把握题中所涉及的事件，明确所求问题所属的事件类型.

跟踪训练3 (2023·昆明模拟)2022年，举世瞩目的冬奥会在北京举行，冬奥会吉祥物“冰墩墩”和“雪容融”有着可爱的外表和丰富的寓意，自亮相以来就好评不断，深受各国人民的喜爱.某市一媒体就本市小学生是否喜爱这两种吉祥物对他们进行了一次抽样调查，列联表如下(单位：人)：

性别	是否喜爱		合计
	喜爱	不喜爱	
男生	30	20	50
女生	40	10	50
合计	70	30	100

(1)根据小概率值 $\alpha=0.01$ 的独立性检验， 能否推断是否喜爱吉祥物与性别有关？

附： $\chi^2=\frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$ ， 其中 $n=a+b+c+d$.

α	0.1	0.05	0.01
x_α	2.706	3.841	6.635

解

零假设为 H_0 : 喜爱吉祥物与性别无关.

根据表中数据得 $\chi^2 = \frac{100 \times (30 \times 10 - 20 \times 40)^2}{70 \times 30 \times 50 \times 50} = \frac{100}{21} \approx 4.762 < 6.635 = \chi_{0.01},$

所以根据小概率值 $\alpha=0.01$ 的独立性检验, 没有充分证据推断 H_0 不成立, 因此可以认为喜爱吉祥物与性别无关.

(2)现从样本的男生中采用比例分配的分层随机抽样的方法抽取5人，再从这5人中随机抽取3人，记抽取的3人中有 X 人喜爱吉祥物，求 X 的分布列和均值.

解

由题意得，采用比例分配的分层随机抽样方法抽取出的5人中，有3人喜爱吉祥物，有2人不喜爱吉祥物，

则 X 的可能取值为1,2,3,

$$\text{所以 } P(X=1)=\frac{C_3^1 C_2^2}{C_5^3}=\frac{3}{10},$$

$$P(X=2)=\frac{C_3^2 C_2^1}{C_5^3}=\frac{3}{5},$$

$$P(X=3)=\frac{C_3^3}{C_5^3}=\frac{1}{10},$$

解

所以 X 的分布列为

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

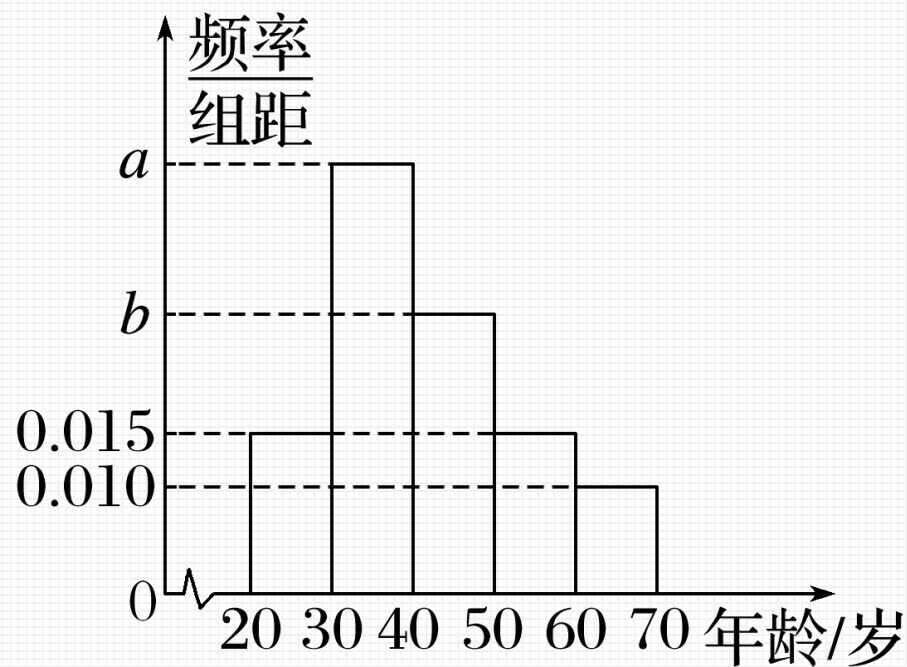
$$\text{则 } E(X) = 1 \times \frac{3}{10} + 2 \times \frac{3}{5} + 3 \times \frac{1}{10} = \frac{9}{5}.$$



课时精练

1.(2023·大理模拟)第24届冬季奥林匹克运动会于2022年2月4日在中国北京开幕,简称“北京冬奥会”.某媒体通过网络随机采访了某市100名关注“北京冬奥会”的市民,并将其年龄数据绘制成如图所示的频率分布直方图.

(1)已知 $[30,40)$, $[40,50)$, $[50,60)$ 三个年龄段的人数依次成等差数列,求 a , b 的值;



以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/945104201213011330>