例析概率与统计十大考点

■中山市东升高中　高建彪

概率与统计是近几年新教材及新课标高考的命题热点，题目的背景密切联系实际生活中的问题，较好地整合了统计方法与概率计算，着重考查考生的运算能力、逻辑思维能力以及分析问题和解决问题的能力. 下面结合几个实例，剖析其中较为重要的考点.

一、统计图表与概率计算

例1 某校从参加高一年级期末考试的学生中抽出 60 名学生，将其成绩（均为整数）分成六段 [40, 50),[50,60)…[90,100]后画出如下部分频率分布直方图. 观察图形的信息,回答下列问题：

（1）求第四小组的频率，并补全这个频率分布直方图；

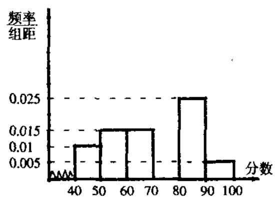
（2）估计这次考试的及格率（60 分及以上为及格）和平均分；

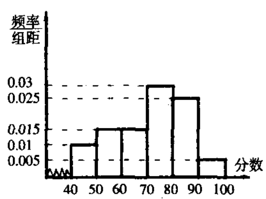
（3）从成绩是 80 分以上（包括 80 分）的学生中选两人，求他们选在同一组的概率.

解析（1）因为各组的频率和等于 1，故第四组的频率：

$f_4=1-(0.025+0.015*2+0.01+0.005)*10=0.03$.

直方图如下所示：





（2）依题意，60 及以上的分数所在的第三、四、五、六组，频率和为：

$(0.015+0.03+2+0.025+0.005)*10=0.75$.

所以，抽样学生成绩的合格率是 75%.

利用组中值估算抽样学生的平均分：

$45\cdot f_1+55\cdot f_2+65\cdot f_3+75\cdot f_4+85\cdot f_5+95\cdot f_6$

$=45\times0.1+55\times0.15+65\times0.15+75\times0.3+85\times0.25+95\times0.05=71$.

所以估计这次考试的平均分是 71 分.

（3）[80,90)，[90, 100] 的人数是 15，3. 所以从成绩是 80 分以上（包括 80 分）的学生中选两人，他们选在同一组的概率为：

$$P=\frac{\frac{15\times14}{2}+\frac{3\times2}{2}}{\frac{18\times17}{2}}=\frac{12}{17}.$$

点评 本题前两问着重考查了对统计图表的数据分析与处理能力，第 3 问的设计中，借助选取的途径，融入了简单的古典概型概率计算，注意领会基本事件计数中，可以简单地运用分类加与分步乘法，从而避免了列举法的繁琐.

二、古典概型与几何概型

例2 在一次商贸交易会上，商家在柜台开展促销抽奖活动，甲、乙两人相约同一天上午去该柜台参与抽奖.

（1）若抽奖规则是从一个装有 6 个红球和 4 个白球的袋中无放回地取出 2 个球，当两个球同色时则中奖，求中奖概率；

（2）若甲计划在 9:00~9:40 之间赶到，乙计划在 9：20~10：00 之间赶到，求甲比乙提前到达的概率.

解析（1）从袋中 10 个球中摸出 2 个，试验的结果共有 $\frac{10\times9}{2}=45$（种）.

中奖的情况分为两种：

（i）2 个球都是红色，包含的基本事件数为 $\frac{6\times5}{2}=15$；

（ii）2 个球都是白色，包含的基本事件数为 $\frac{4\times3}{2}=6$.

所以，中奖这个事件包含的基本事件数为 15+6=21. 因此，中奖概率为 $\frac{21}{45}=\frac{7}{15}$.

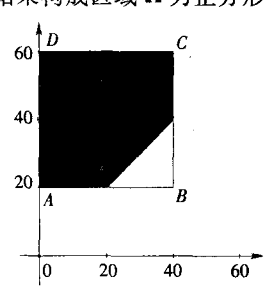
（2）设两人到达的时间分别为 9 点到 10 点之间的 x 分钟、y 分钟. 用（x, y）表示每次试验的结果，则所有可能结果为：

$\Omega=\{(x,y)=|0\leqslant x\leqslant40,20\leqslant y\leqslant60\}$.

记甲比乙提前到达为事件 A，则事件 A 的可能结果为：

$A=\{(x,y)|x<y,0\leqslant x\leqslant40,20\leqslant y\leqslant60\}$.

如图所示，试验全部结果构成区域 Ω 为正方形 $ABCD$,而事件 A 所构成区域是正方形内的阴影部分.



根据几何概型公式，得到：

$$P(A)=\frac{S_{阴影}}{S_{正方形}}=\frac{40^2-\frac{1}{2}\times20^2}{40^2}=\frac{7}{8}.$$

所以，甲比乙提前到达的概率为$\frac{7}{8}$.

点评 要能正确区分古典概型及几何概型的特征，运用相关公式来计算. 求解古典概型，关键是分析出基本事件的个数；求解几何概型，关键是正确构造出随机事件对应的几何图形，利用图形的几何度量之比来求随机事件的概率，体现了数形结合思想的应用.

三、分布列与期望方差

例3 在一次有奖竞猜活动中，有A、B两个相互独立的问题，现规定：答对问题A可获奖金1000元，答对问题B可获奖金2000元，先答哪个题可自由选择，但只有第一个问题答对，才能再答第二个问题，否则终止答题. 若你参加答题，且假设答对问题A、B的概率分别为$\frac{1}{2}$、$\frac{1}{4}$.

(1) 记先回答问题A获得的奖金数为随机变量ξ，求ξ的分布列；

(2) 记先回答问题B获得的奖金数为随机变量η，求η的期望与方差；

(3) 你将选择先回答哪个问题？请说明理由.

解析 (1)随机变量ξ的可能取值分别为0,1000,3000,则有$P(\xi=0)=1-\frac{1}{2}=\frac{1}{2}$;$P(\xi=1000)=\frac{1}{2}\times(1-\frac{1}{4})=\frac{3}{8}$;$P(\xi=3000)=\frac{1}{2}\times\frac{1}{4}=\frac{1}{8}$.

所以，先回答问题A获得的奖金数ξ的分布列为：

ξ	0	1000	3000
P	$\frac{1}{2}$	$\frac{3}{8}$	$\frac{1}{8}$

(2)随机变量y的可能取值分别为0,2000,3000,则有：

$P(\eta=0)=1-\frac{1}{4}=\frac{3}{4}$;$P(\eta=2000)=\frac{1}{4}\times(1-\frac{1}{2})=\frac{1}{8}$;$P(\eta=3000)=\frac{1}{4}\times\frac{1}{2}=\frac{1}{8}$,

$\therefore E\eta=0\times\frac{3}{4}+2000\times\frac{1}{8}+3000\times\frac{1}{8}=\frac{5000}{8}=625$,

$D\eta=(0-625)^2\times\frac{3}{4}+(2000-625)^2\times\frac{1}{8}+(3000-625)^2\times\frac{1}{8}=1234375$.

(3) $E\xi=0\times\frac{1}{2}+1000\times\frac{3}{8}+3000\times\frac{1}{8}=\frac{6000}{8}=750$(元)，$E\eta=625$(元)，

故先回答问题A获得的奖金期望较多，所以选择先回答问题A.

点评 求离散型随机变量的分布列，必须先明确变量的取值，再按照相关概率计算公式求得随机变量各个取值的概率，并运用性质“$P_i\leqslant1$”与“$P_1+P_2+\cdots+P_i+\cdots=1$”进行检验. 而期望与方差直接按照公式计算即可，注意期望反映了离散型随机变量取值的平均水平，方差则反映了随机变量取值的稳定与波动、集中与离散的程度.

四、超几何分布与二项分布

例4 某校设计了一个实验学科的实验考查方案：考生从6道备选题中一次性随机抽取3题，按照题目要求独立完成全部实验操作. 规定：至少正确完成其中2题的便可通过. 已知6道备选题中考生甲有4题能正确完成，2题不能完成；考生乙每题正确完成的概率都为$\frac{2}{3}$，且每题正确完成与否互不影响.

(1) 写出甲正确完成题数的概率分布列，并计算其数学期望；

(2) 写出乙正确完成题数的概率分布列，并求乙至少正确完成2题的概率；

(3) 试用统计知识分析比较两考生的实验操作能力.

解析 (1) 设考生甲正确完成实验操作的题数为ξ，则ξ取值为1，2，3，

$P(\xi=1)=\frac{C_4^1C_2^2}{C_6^3}=\frac{1}{5}$,$P(\xi=2)=\frac{C_4^2C_2^1}{C_6^3}=\frac{3}{5}$,$P(\xi=3)=\frac{C_4^3C_2^0}{C_6^3}=\frac{1}{5}$,

$\therefore$ 考生甲正确完成题数的概率分布列为：

ξ	1	2	3
P	$\frac{1}{5}$	$\frac{3}{5}$	$\frac{1}{5}$

$E\xi=1\times\frac{1}{5}+2\times\frac{3}{5}+3\times\frac{1}{5}=2$.

(2) 设考生乙正确完成实验操作的题数为η，则η取值为0，1，2.

根据题意，易知考生乙做对题数η服从二项分布，即$\eta\sim B(3,\frac{2}{3})$,

$\therefore P(\eta=0)=C_3^0(1-\frac{2}{3})^3=\frac{1}{27}$，$P(\eta=1)=C_3^1(1-\frac{2}{3})^2\cdot\frac{2}{3}=\frac{6}{27}$，$P(\eta=2)=C_3^2(1-\frac{2}{3})\cdot(\frac{2}{3})^2=\frac{12}{27}$，$P(\eta=3)=C_3^3(\frac{2}{3})^3=\frac{8}{27}$.

得到考生乙正确完成题数的概率分布列为：(见后表)

所以，乙至少正确完成2题的概率为：

ξ	0	1	2	3
P	$\frac{1}{27}$	$\frac{6}{27}$	$\frac{12}{27}$	$\frac{8}{27}$

$P(\eta\geqslant 2)=P(\eta=2)+P(\eta=3)=\frac{12}{27}+\frac{8}{27}=\frac{20}{27}$.

(3) 由 $E\eta=np=3\times\frac{2}{3}=2$，得到 $E\xi=E\eta$.

又 $D\xi=(2-1)^2\times\frac{1}{5}+(2-2)^2\times\frac{3}{5}+(2-3)^2\times\frac{1}{5}=\frac{2}{5}$，

$D\eta=np(1-p)=3\times\frac{2}{3}\times\frac{1}{3}=\frac{2}{3}$， $\therefore D\xi<D\eta$.

$\because P(\xi\geqslant 2)=\frac{3}{5}+\frac{1}{5}=0.8$，$P(\eta\geqslant 2)=\frac{12}{27}+\frac{8}{27}\approx$ 0.74，$\therefore P(\xi\geqslant 2)>P(\eta\geqslant 2)$.

从做对题数的数学期望考察，两人水平相当；从做对题数的方差考察，甲较稳定；从至少完成 2 题的概率考察，甲获得通过的可能性大，因此判断甲的实验操作能力较强.

点评 此题甲答对题数属于典型的超几何分布，关键是根据计数原理，完成随机变量各取值的概率计算，注意运用分布列的性质来检验答案是否正确. 而第 (2) 问乙答对题数服从二项分布，从而可利用二项分布的相关公式来求解；第 (3) 问则在期望值相同的情况下，进一步通过方差的大小来比较甲、乙正确完成题目的稳定性.

五、独立性检验与回归分析

例 5 为考察初中生的数学成绩与物理成绩的关系，对全市中考成绩进行抽样分析：

(1) 从中抽 105 个样本，按照考试成绩优秀和不优秀统计成绩后，得到如下列联表：

请问数学成绩与物理成绩在多大程度上有关系？

数学成绩与物理绩的列联表

	优秀	不优秀	总　计
数学成绩	10	45	55
物理成绩	20	30	50
总计	30	75	105

(2) 从中抽 8 个样本，这 8 位同学的数学、物理分数对应如下表：

学生编号	1	2	3	4	5	6	7	8
数学分数 x	60	65	70	75	80	85	90	95
物理分数 y	72	77	80	84	88	90	93	95

试用变量 y 与 x 的相关系数说明物理与数学的相关程度，并求 y 与 x 的线性回归方程（系数精确到 0.01），并用相关指数描述所求回归模型的拟合效果.

参考数据：$\bar{x}=77.5$，$\bar{y}=85$，$\sum\limits_{i=1}^{8}(x_i-\bar{x})^2\approx 1050$，

$\sum\limits_{i=1}^{8}(y_i-\bar{y})^2\approx 456$，$\sum\limits_{i=1}^{8}(x_i-\bar{x})(y_i-\bar{y})\approx 688$，

$\sum\limits_{i=1}^{8}(y_i-\hat{y}_i)^2\approx 7$，$\sqrt{1050}\approx 32.4$，$\sqrt{456}\approx 21.4$.

解析 (1) 假设"数学成绩与物理成绩之间没有关系".

而随机变量 K^2 的观测值 $k=\frac{105\times(10\times 30-20\times 45)^2}{55\times 50\times 30\times 75}$ $\approx 6.109>5.024$.

由 $P(K^2\geqslant 5.024)\approx 0.025$，得到"数学成绩与物理成绩之间有关系"这一结论是错误的可能性约为 0.025，即有 97.5%的把握认为"数学成绩与物理成绩之间有关系".

(2) 变量 y 与 x 的相关系数：

$$r=\frac{\sum\limits_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum\limits_{i=1}^{n}(x_i-\bar{x})^2\sum\limits_{i=1}^{n}(y_i-\bar{y})^2}}=\frac{688}{32.4\times 21.4}\approx 0.99,$$

所以物理与数学是高度正相关.

设 y 与 x 的线性回归方程分别 $\hat{y}_i=bx+a$，根据所给的数据，得到：

$$b=\frac{\sum\limits_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum\limits_{i=1}^{n}(x_i-\bar{x})^2}=\frac{688}{1050}=0.65,$$

$a=\bar{y}-b\bar{x}=85-0.65\times 77.5=34.63$，

所以 y 与 x 的回归方程是 $\hat{y}_i=0.65x+34.63$.

又 y 与 x、z 与 x 的相关指数是：

$$R^2=1-\frac{\sum\limits_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum\limits_{i=1}^{n}(y_i-\bar{y})^2}=1-\frac{7}{456}\approx 0.98.$$

故解释变量数学分数对预报变量物理分数解释了 98%，模型拟合效果较好.

点评 本题考查了独立性检验、回归分析相关统计量及最小二乘法求线性回归方程等知识，突出了对数据处理能力、运算求解能力、应用数学知识解决实际实际问题的能力的考查. 解答此类问题，要求能熟练运用公式并理解其含义. 此题通过给出参考数值，控制了计算量；题目综合了统计案例章节中的两大案例，但在高考中以其中一个作为考查对象的可能性较大. 在实施课程标准之后，我们一定要注重提高数据处理能力.

责任编校　徐国坚