数学高考一轮复习专题之十四

概率与统计

黄梅一中 王卫华 上海行知实验中学 蔡敏珍

概率统计试题通常是通过对课本原题进行改编，通过对基础知识的重新组合、变式和拓展，从而加工为立意高、情境新、设问巧、并赋予时代气息、贴近学生实际的问题.

例1 为了分析某个高三学生的学习状态，对其下一阶段的学习提供指导性建议. 现对他前7次考试的数学成绩 x 、物理成绩 y 进行分析. 下面是该生7次考试的成绩.

数学 x	88	83	117	92	108	100	112
物理 y	94	91	108	96	104	101	106

(1)他的数学成绩与物理成绩哪个更稳定？请给出你的证明；

(2)已知该生的物理成绩 y 与数学成绩 x 是线性相关的，若该生的物理成绩达到115分，请你估计他的数学成绩大约是多少？并请你根据物理成绩与数学成绩的相关性，给出该生在学习数学、物理上的合理建议.

分析 成绩的稳定性用样本数据的方差判断，由物理成绩估计数学成绩可用回归直线方程解决.

解 (1) $\bar{x}=100+\dfrac{-12-17+17-8+8+12}{7}=100$；

$\bar{y}=100+\dfrac{-6-9+8-4+4+1+6}{7}=100$；

$\therefore S^2_{数学}=\dfrac{994}{7}=142$，$\therefore S^2_{物理}=\dfrac{250}{7}$，

从而 $S^2_{数学}>S^2_{物理}$，所以物理成绩更稳定.

(2)由于 x 与 y 之间具有线性相关关系，根据回归系数公式得到 $\hat{b}=\dfrac{497}{994}=0.5$，$\hat{a}=100-0.5\times100=50$，

$\therefore$ 线性回归方程为 $y=0.5x+50$.

当 $y=115$ 时，$x=130$.

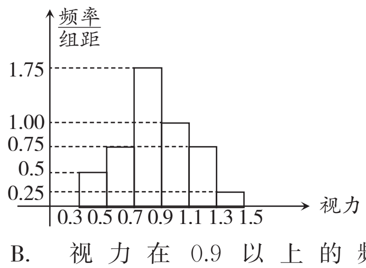
建议：进一步加强对数学的学习，提高数学成绩的稳定性，将有助于物理成绩的进一步提高.

点拨 《考试大纲》在必修部分的统计中明确指出"①会作两个有关联变量的数据的散点图，会利用散点图认识变量间的相关关系. ②了解最小二乘法的思想，能根据给出的线性回归方程系数公式建立线性回归方程"，在复习备考时不可掉以轻心.

例2 从某校高三年级随机抽取一个班，对该班50名学生的高校招生体检表中视力情况进行统计，其结果的频率分布直方图如右图：若某高校 A 专业对视力的要求在0.9以上，则该班学生中能报 A 专业的人数为(　　)

A. 10　　B. 20　　C. 8　　D. 16

分析 根据图找出视力在0.9以上的人数的频率即可.



解 B. 视力在0.9以上的频率为 $(1+0.75+0.25)\times0.2=0.4$，人数为 $0.4\times50=20$.

点拨 在解决频率分布直方图问题时，容易出现的错误是认为直方图中小矩形的高就是各段的频率，实际上小矩形的高是频率除以组距.

例3 在调查的480名上网的学生中有38名学生睡眠不好，520名不上网的学生中有6名学生睡眠不好，利用独立性检验的方法来判断是否能以99%的把握认为"上网和睡眠是否有关系". 附：

$K^2=\dfrac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$；参考数据如下：

$P(K^2\geqslant k_0)$	0.50	0.40	0.25	0.15	0.10	0.05	0.025	0.010	0.005	0.001
k_0	0.455	0.708	1.323	2.072	2.706	3.841	5.024	6.635	7.879	10.828

$42064\times2496=104991744$，$1688^2=2849344$.

分析　这是一个独立性检验应用题，处理本题时要根据列联表及 K^2 的计算公式，计算出 K^2 的值，并代入临界值表中进行比较，不难得到答案.

解　由题中数据列表如下：

	睡眠好	睡眠不好	总计
不上网	514	6	520
上网	442	38	480
总计	956	44	1000

由上表 $a=514, b=6, c=442, d=38$，

则 $K^2=\dfrac{1000(514\times38-442\times6)^2}{520\times480\times44\times956}\approx27.14$.

由参考数据得 $P(K^2\geqslant6.635)=0.010$，

而 $27.14>6.635$，

故可以以 99 %的把握认为“上网和睡眠有关系”.

点拨　本题关键在于理解临界值表的意义.

例4　某篮球运动员在一个赛季的40场比赛中的得分的茎叶图如图所示，则这组数据的中位数是______；众数是__________.

0	8 9
1	1 2 2 2 3 3 4 6 7 8 9
2	0 1 1 1 3 3 3 3 5 5 7 8 8
3	0 1 2 2 3 4 4 8 9
4	0 1 3 5 6

分析　根据茎叶图和中位数、众数的概念解决.

解　由于中位数是把样本数据按照由小到大的顺序排列起来，处在中间位置的一个（或是最中间两个数的平均数），故从茎叶图可以看出中位数是23；而众数是样本数据中出现次数最多的数，故众数也是23.

点拨　一表（频率分布表）、三图（频率分布直方图、频率折线图、茎叶图）、三数（众数、中位数、平均数）和标准差，是高考考查统计的一个主要考点.

例5　某种比赛的规则是5局3胜制，甲、乙两人在比赛中获胜的概率分别是 $\frac{2}{3}$ 和 $\frac{1}{3}$.（1）若有3局中乙以2∶1领先，求乙获胜的概率；（2）若胜一局得2分，负一局得分，求甲得分 ξ 的数学期望.

分析　乙获胜的可能有两种：（1）3∶1，乙只需用胜第4场即可；（2）3∶2，乙需第4场失败，第5场获胜，第（2）问先分析 ξ 的取值，注意在计算各种情况的得分时要将正分加上负分.

解　（1）依题意，前三局乙以2∶1领先，∴乙获用的可能有两种：①乙在第4局获胜，概率为 $\frac{1}{3}$；②乙在第4局失败，在第5局获胜，概率为 $\frac{2}{3}\times\frac{1}{3}=\frac{2}{9}$，而这两种情况互斥，∴乙获胜的概率为 $\frac{1}{3}+\frac{2}{9}=\frac{5}{9}$.

（2）将甲获胜的场数写在前面，则比赛结果有以下几种：①0∶3；②1∶3；③2∶3；④3∶0；⑤3∶1；⑥3∶2. 在①中 $\xi=-3$，②中 $\xi=-1$，③中 $\xi=1$，④中 $\xi=6$，⑤中 $\xi=5$，⑥中 $\xi=4$. ∴ ξ 的取值为 $-3,-1,1,4,5,6$.

$P(\xi=-3)=(\frac{1}{3})^3=\frac{1}{27}$，

$P(\xi=-1)=C_3^1\cdot\frac{2}{3}\cdot(\frac{1}{3})^3=\frac{2}{27}$，

$P(\xi=1)=C_4^2\cdot(\frac{2}{3})^2\cdot(\frac{1}{3})^3=\frac{8}{81}$，

$P(\xi=4)=C_4^2\cdot(\frac{1}{3})^2\cdot(\frac{2}{3})^3=\frac{16}{81}$，

$P(\xi=5)=C_3^1\cdot\frac{1}{3}\cdot(\frac{2}{3})^3=\frac{8}{27}$，

$P(\xi=6)=(\frac{2}{3})^3=\frac{8}{27}$.

∴ ξ 的分布列如下所示：

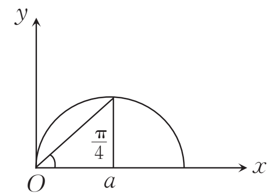
ξ	-3	-1	1	4	5	6
P	$\frac{1}{27}$	$\frac{2}{27}$	$\frac{8}{81}$	$\frac{16}{81}$	$\frac{8}{27}$	$\frac{8}{27}$

∴ $E\xi=-3\times\frac{1}{27}+(-1)\times\frac{2}{27}+1\times\frac{8}{81}+4\times\frac{16}{81}+5\times\frac{8}{27}+6\times\frac{8}{27}$.

∴甲得分 ξ 的数学期望为 $\frac{107}{27}$.

点拨　在实际生活的概率问题，常常是“互斥、独立、独立重复”的混合问题，解题时必须仔细分析题意，正确判断属于何种概型，灵活运用公式，选择较为简便的方法解决问题.

例5　随机地向半圆 $0<y<\sqrt{2ax-x^2}$（a 为正常数）内掷一点，点落在圆内任何区域的概率与区域的面积成正比，求原点与该点的连线与 x 轴的夹角小于 $\frac{\pi}{4}$ 的概率.



分析 就是圆的面积和正方形面积的比值.

解 半圆域如图,设 A="原点与该点连线与 x 轴夹角小于 $\frac{\pi}{4}$",由几何概率的定义得

$$P(A)=\frac{A\text{的面积}}{\text{半园的面积}}=\frac{\frac{1}{4}\pi a^2+\frac{1}{2}a^2}{\frac{1}{2}\pi a^2}=\frac{1}{2}+\frac{1}{\pi}.$$

点拨 高考对几何概型的考查一般有两个方面,一是以选择题、填空题的方式有针对性地考查,二是作为综合解答题的一部分和其他概率计算一起进行综合考查.

【专题训练十四】

1. 以下五个命题:①从匀速传递的产品生产流水线上,质检员每10分钟从中抽取一件产品进行某项指标检测,这样的抽样是分层抽样;②样本方差反映了样本数据与样本平均值的偏离程度;③在回归分析模型中,残差平方和越小,说明模型的拟合效果越好;④在回归直线方程 $\hat{y}=0.1x+10$ 中,当解释变量 x 每增加一个单位时,预报变量 $\hat{y}$ 增加0.1个单位;⑤在一个 2×2 列联表中,由计算得 $K^2=13.079$,则其两个变量间有关系的可能性是90%以上. 其中正确的是(　　)

A. ②③④⑤　　B. ①③④

C. ①③⑤　　D. ②④

2. 在区间 $[-2,2]$ 内任取两数 a、b,使函数 $f(x)=x^2+2bx+a^2$ 有两相异零点的概率是(　　)

A. $\frac{1}{6}$　　B. $\frac{1}{4}$　　C. $\frac{1}{3}$　　D. $\frac{1}{2}$

3. 连续2次抛掷一枚骰子(六个面上分别标有数字1,2,3,4,5,6),记"两次向上的数字之和等于 m"为事件 A,则 $P(A)$ 最大时,m=________.

4. 一个盒子里装有相同大小的黑球10个,红球12个,白球4个,从中任取两个,其中白球的个数记为 ξ,则下列算式中等于 $\frac{C_{26}^1C_4^1+C_{22}^2}{C_{26}^2}$ 的是(　　)

A. $P(0<\xi\leqslant2)$　　B. $P(\xi\leqslant1)$

C. $E\xi$　　D. $D\xi$

5. 已知 $k\in\mathbf{Z}$,$\overrightarrow{AB}=(k,1)$,$\overrightarrow{AC}=(2,4)$,若 $|A|\leqslant4$,则 $\triangle ABC$ 是直角三角形的概率为________.

6. 设 $X\sim N(\mu,\sigma^2)$,且总体密度曲线的函数表达式为:$f(x)=\frac{1}{2\sqrt{\pi}}e^{-\frac{x^2-2x+1}{4}}$,$x\in\mathbf{R}$.(1)求 μ,σ;(2)求 $p(|x-1|<\sqrt{2})$ 及 $P(1-\sqrt{2}<x<1+2\sqrt{2})$ 的值.

7. 为让学生了解"奥运"知识,某中学举行了一次"奥运知识竞赛",共有800名学生参加了这次竞赛,为了解本次竞赛成绩情况,从中抽取了部分学生的成绩(得分均为整数,满分为100分)进行统计. 请你根据尚未完成并有局部污损的频率分布表,解答下列问题:

分组	频数	频率
60.5 ~ 70.5		0.16
70.5 ~ 80.5	10	
80.5 ~ 90.5	18	0.36
90.5 ~ 100.5		
合计	50	

(1)若用系统抽样的方法抽取50个个体,现将所有学生随机地编号为000,001,002,…,799,试写出第二组第一位学生的编号;(2)填充频率分布表的空格(将答案直接填在表格内),并作出频率分布直方图;(3)若成绩在85. 5 ~ 95. 5分的学生为二等奖,问参赛学生中获得二等奖的学生约为多少人?

8. 设三组实验数据 (x_1,y_1)、(x_2,y_2)、(x_3,y_3) 的回归直线方程是 $\hat{y}=bx+a$,使代数式 $[y_1-(bx_1+a)]^2+[y_2-(bx_2+a)]^2+[y_3-(bx_3+a)]^2$ 的值最小时,$a=\bar{y}-b\bar{x}$,$b=\frac{x_1y_1+x_2y_2+x_3y_3-3\bar{x}\bar{y}}{x_1^2+x_2^2+x_3^2-3\bar{x}^2}$($\bar{x}$、$\bar{y}$ 分别是这三组数据的横、纵坐标的平均数). 若有七组数据列表如下:

x	2	3	4	5	6	7	8
y	4	6	5	6.2	8	7.1	8.6

(1)求上表中前三组数据的回归直线方程;

(2)若 $|y_i-(bx_i+a)|\leqslant0.2$,即称 (x_i,y_i) 为(1)中回归直线的拟和"好点",求后四组数据中拟和"好点"的概率.

【参考答案】

1. A　2. D　3. 7　4. B　5. $\frac{3}{7}$

6. $X\sim N(1,2)$;0.8185.

7. 编号为016. 获二等奖的大约有256人

8. (1) $\hat{y}=\frac{1}{2}x+\frac{7}{2}$　(2) $\frac{1}{2}$