江苏省仪征中学2022-2023学年度第二学期高三数学学科导学案
统计案例

研制人：葛生芳　　审核人：陈宏强

班级：__________姓名：__________学号：________授课日期：

【考情分析】

高考试题对统计案例的考查与生产和生活紧密相连,紧跟时代创设情境,背景新颖,着重考查随机抽样、用样本估计总体,线性回归方程,独立性检验.有时还与概率知识交汇命题.综合考查考生阅读理解能力、数据分析和处理能力,综合运用数学知识解决实际问题的能力.

【真题感悟】

1.(2022 全国甲卷)甲、乙两城之间的长途客车均由A和B两家公司运营,为了解这两家公司长途客车的运行情况,随机调查了甲、乙两城之间的500个班次,得到下面列联表，

	准点班次数	末准点班次数
A	240	20
B	210	30

(1)根据上表,分别估计这两家公司甲、乙两城之间的长途客车准点的概率;

(2)能否有90%的把握认为甲、乙两城之间的长途客车是否准点与客车所属公司有关?

参考公式和数据:$\chi^2=\frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$,其中$n=a+b+c+d$.

$P(\chi^2\geqslant x_0)$	0.100	0.050	0.010
x_0	2.706	3.841	6.635

2.(2022 全国乙卷·理科)某地经过多年的环境治理,已将荒山改造成了绿水青山.为估计一林区某种树木的总材积量,随机选取了 10 棵这种树木,测量每棵树的根部横截面积(单位:m^2)和材积量(单位:m^3),得到如下数据:

样本号i	1	2	3	4	5	6
根部横截面积x_i	0.04	0.06	0.04	0.08	0.08	0.05
材积量y_i	0.25	0.40	0.22	0.54	0.51	0.34
样本号i	7	8	9	10	总和	
根部横截面积x_i	0.05	0.07	0.07	0.06	0.6	
材积量y_i	0.36	0.46	0.42	0.40	3.9	

并计算得$\sum_{i=1}^{10} x_i^2 = 0.038, \sum_{i=1}^{10} y_i^2 = 1.6158, \sum_{i=1}^{10} x_i y_i = 0.2474$.

(1)估计该林区这种树木平均一棵的根部横截面积与平均一棵的材积量;

(2)求该林区这种树木的根部横截面积与材积量的样本相关系数(精确到0.01);

(3)现测量了该林区所有这种树木的根部横截面积,并得到所有这种树木的根部横截面积总和为$186\mathrm{m}^2$.已知树木的材积量与其根部横截面积近似成正比.利用以上数据给出该林区这种树木的总材积量的估计值.

参考公式和数据：相关系数$r = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_1-\bar{x})^2 \sum_{i=1}^{n}(y_i-\bar{y})^2}}$，$\sqrt{1.896} \approx 1.377$.

【典例导引】

例 1．(2022 江苏南京市模拟)中国是茶的故乡,也是茶文化的发源地.为了弘扬中国茶文化,某酒店推出特色茶饮,按事先拟定的价格进行试销,得到销售数据$(x_i,y_i)(i=1,2,\cdots,6)$,如下表所示:

x试销单价x(元)	20	25	30	35	40	45
销量y(壶)	m	88	86	76	73	68

参考数据:$\bar{y} = \frac{1}{6}\sum_{i=1}^{6} y_i = 80.5, \sum_{i=1}^{6} x_i y_i = 15260, \sum_{i=1}^{6} x_i^2 = 6775, 6 \times 32.5^2 = 6337.5$.

(1)已知变量x, y具有线性相关关系,求销量y(壶)关于试销单价x(元)的线性回归方程$\hat{y} = \hat{b}x + \hat{a}$和$m$的值;

(2)用$\hat{y}_i$表示根据线性回归方程得到的与x_i对应的销量的估计值,当销售数据(x_i, y_i)中y_i与估计值$\hat{y}_i$满足$|\hat{y}_i - y_i| \leqslant 1$ 时,则称该销售数据(x_i, y_i)为一组“理想数据”.现从 6 组销售数据中任取 2 组,求抽取的 2 组销售数据中至少有 1 组是“理想数据”的概率.

参考公式:$\hat{b} = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\cdot\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}$,截距$\hat{a} = \bar{y} - \hat{b}x$.

例 2．为加强环境保护,治理空气污染,环境监测部门对某市空气质量进行调研,随机抽查了 100 天空气中的 PM2.5 和SO_2浓度(单位:$\mu g/m^3$),得下表:

		SO_2		
		[0,50]	(50,150]	(150,475]
PM2.5	[0.35]	32	18	4
	(35.75]	6	8	12
	(75,115]	3	7	10

(1)估计事件“该市一天空气中 PM2.5 浓度不超过 75,且SO_2浓度不超过 150”的概率;

(2)根据所给数据,完成下面的2×2列联表，

		SO_2	
		[0,150]	(150,475]
PM2.5	[0,75]		
	(75,115]		

(3)根据(2)中的列联表,判断是否有99%的把握认为该市一天空气中 PM2.5 浓度与SO_2浓度有关.

参考公式和数据:$\chi^2=\frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$,其中$n=a+b+c+d$.

$P(\chi^2\geqslant x_0)$	0.050	0.010	0.001
x_0	3.841	6.635	10.828

例 3.(2021 江苏南通市三模)面对新一轮科技和产业革命带来的创新机遇,某企业对现有机床进行更新换代,购进一批新机床.设机床生产的零件的直径为X(单位:mm).

(1)现有旧机床生产的零件 10 个,其中直径大于124mm的有 3 个.若从中随机抽取 4 个,记ξ表示取出的零件中直径大于124mm的零件的个数,求ξ的概率分布及数学期望$E(\xi)$;

(2)若新机床生产的零件直径$X\sim N(120,4)$,从生产的零件中随机取出 10 个,求至少有 1 个零件直径大于124mm的概率.

参考数据:若$X\sim N(\mu,\sigma^2)$,则$P(|X-\mu|\leqslant\sigma)\approx 0.6827, P(|X-\mu|\leqslant 2\sigma)\approx 0.9545,$
$P(|X-\mu|\leqslant 3\sigma)\approx 0.9973, 0.97725^{10}\approx 0.7944, 0.9545^{10}\approx 0.6277.$

例 4．(2022 湖南郴州市一模)2021 年东京奥运会,中国举重代表队共 10 人,其中主教练、教练各 1 人,参赛选手 8 人,分别参加 8 个级别的比赛,赛后结果 7 金 1 银,在全世界面前展现了真正的中国力量.举重比赛根据体重进行分级,某次举重比赛中,男子举重按运动员体重分为下列十级:

级别	54 公斤级	59 公斤级	64 公斤级	70 公斤级
体重	554	54.01 ~ 59	59.01 ~ 64	64.01 ~ 70

级别	76 公斤级	83 公斤级	91 公斤级
体重	70.01 ~ 76	76.01 ~ 83	83.01 ~ 91

级别	99 公斤级	108 公斤级	108 公斤级以上
体重	91.01 ~ 99	99.01 ~ 108	⩾ 108

每个级别的比赛分为抓举与挺举两个部分,最后综合两部分的成绩得出总成绩,所举重量最大者获胜,在该次举重比赛中,获得金牌的运动员的体重以及举重成绩如下表:

体重	54	59	64	70	76
举重成绩	291	304	337	353	363
体重	83	91	99	106	
举重成绩	389	406	421	430	

(1)根据表中的数据,求出运动员举重成绩y与运动员的体重x的回归直线方程(保留 1 位小数);

(2)某金牌运动员抓举成绩为 180 公斤,挺举成绩为 218 公斤,则该运动员最有可能是参加的哪个级别的举重?

(3)凯旋后,中央一台记者从团队的 10 人中随机抽取 3 人进行访谈,用ξ表示抽取到的是金牌得主的人数,求ξ的概率分布列与数学期望.

参考公式和数据:$\hat{b}=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$, $\hat{a}=\bar{y}-\hat{b}x$.$\sum_{i=1}^{9}(x_i-\bar{x})^2=2620$,$\sum_{i=1}^{9}(x_i-\bar{x})(y_i-\bar{y})=7076$.

江苏省仪征中学2022-2023学年度第二学期高三数学学科作业

统计案例

研制人： 葛生芳　　审核人：陈宏强

班级：__________ 姓名：__________学号：________时长：60分钟

1.为了研究高中学生对乡村音乐的态度(喜欢和不喜欢两种态度)与性别的关系,运用2×2列联表进行独立性检验,经计算得$\chi^2=7.01$,则认为“喜欢乡村音乐与性别有关系”的把握约为(　　)

$P(\chi^2\geqslant x_0)$	0.1	0.05	0.01	0.005	0.001
x_0	2.706	3.841	6.635	7.879	10.828

A.0.1%　　B.1%　　C.99%　　D.99.9%

2.某省二线城市地铁正式开工建设,地铁时代的到来能否缓解该市的交通拥堵状况呢?某社团进行社会调查,得到的数据如T衷：

	男性市民	女性市民
认为能缓解交通拥堵	48	30
认为不能缓解交通拥堵	12	20

参考公式和数据:$\chi^2=\frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$,其中$n=a+b+c+d$.

$P(\chi^2\geqslant x_0)$	0.05	0.010	0.005	0.001
x_0	3.841	6.635	7.879	10.828

则下列结论正确的是(　　)

A.有95%的把握认为“对能否缓解交通拥堵的认识与性别有关”

B.有95%的把握认为“对能否缓解交通拥堵的认识与性别无关”

C.有99%的把握认为“对能否缓解交通拥堵的认识与性别有关”

D.有99%的把握认为“对能否缓解交通拥堵的认识与性别无关”

3.已知变量x,y的关系可以用模型$y=c\cdot e^{kx}$拟合,设$z=\ln y$,其变换后得到一组数据下:

x	16	17	18	19
z	50	34	41	31

由上表可得线性回归方程$\hat{z}=-4x+\hat{a}$,则$c=$(　　)

A.-4　　B.e^{-4}　　C.109　　D.e^{109}

4.(2021 广东深圳市二模)某产品的宣传费用x(单位:万元)与销售额y(单位:万元)的统计数据如表所示:

x	4	5	6	7	8
y	60	80	90	100	120

根据上表可得回归方程$\hat{y}=14x+\hat{a}$,则宣传费用为9万元时,销售额最接近(　　)

A.123万元　　B.128万元　　C.133万元　　D.138万元

5.(多选题)在发生公共卫生事件期间,有专业机构认为该事件在一段时间内没有发生大规模群体感染的标志为“连续10天,每天新增疑似病例不超过7人”.则根据过去10日新增疑似病例数据信息,一定符合没有发生大规模群体感染标志的是(　　)

A.中位数为2,极差为5　　B.总体平均数为2,众数为2

C.总体平均数为1,总体方差大于0　　D.总体平均数为2,总体方差为3

6.(多选题)某机构在研究性别与是否爱好拳击运动的关系中,通过收集数据得到如下2×2列联表:

	男	女	合计
爱好拳击	35	22	57
不爱好拳击	15	28	43
合计	50	50	100

经计算得$\chi^2 \approx \frac{100\times(35\times28-15\times22)^2}{50\times50\times57\times43} \approx 6.895$.之后又对被研究者的身高进行了统计,得到男、女身高分别近似服从正态分布$N(175,16)$和$N(164,9)$,则下列选项正确的是(　　)

$P(\chi^2 \geqslant x_0)$	0.50	0.05	0.010	0.005	0.001
x_0	0.455	3.841	6.635	7.879	10.828

A.在犯错误的概率不超过1%的前提下,认为“爱好拳击运动与性别有关”
B.在 100 个男生中,至少有 1 个人爱好打拳击
C.男生身高的平均数为 175,男生身高的标准差为 16
D.女生身高的平均数为 164,女生身高的标准差为 3

7.已知y与x之间的线性回归方程为$\hat{y} = 1.6x + 21$,其样本点的中心为$(\bar{x}, 37)$,样本数据中x的取值依次为$2,6,8,16,m$,则$m =$____________.

8.某工厂为研究某种产品产量x(吨)与所需某种原材料y(吨)的相关性,在生产过程中收集 4 组对应数据(x, y)如下表所示:

x	3	4	6	7
y	2.5	3	4	m

根据表中数据,得出y关于x的线性回归方程为$\hat{y} = 0.7x + \hat{a}$.据此计算出在样本$(4,3)$处的残差为-0.15,则表中m的值为____________.

9.(2021 全国甲卷)甲、乙两台机床生产同种产品,产品按质量分为一级品和二级品,为了比较两台机床产品的质量,分别用两台机床各生产了 200 件产品,产品的质量情况统计如下表:

	一级品	二级品	合计
甲机床	150	50	200
乙机床	120	80	200
合计	270	130	400

(1)甲机床、乙机床生产的产品中一级品的频率分别是多少?
(2)能否有99%的把握认为甲机床的产品质量与乙机床的产品质量有差异?

参考公式和数据:$\chi^2 = \frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$,其中$n = a + b + c + d$.

$P(\chi^2 \geqslant x_0)$	0.050	0.010	0.001
x_0	3.841	6.635	10.828

10.(2022 山东济宁二模)为研究某种疫苗A的效果,现对 100 名志愿者进行了实验,得到如下数据:

	未感染病毒B	感染病毒B	合计
接种疫苗A	40	10	50
未接种疫苗A	20	30	50
合计	60	40	100

(1)根据小概率值$\alpha = 0.001$的独立性检验,分析疫苗A是否有效?

(2)现从接种疫苗A的 50 名志愿者中按分层随机抽样方法(各层按比例分配)取出 10 人,再从这 10 人中随机抽取 3 人,求这 3 人中感染病毒B的人数X的分布列和数学期望.

参考公式和数据:$\chi^2 = \frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$,其中$n = a + b + c + d$.$P(\chi^2 \geqslant 10.828) = 0.001$.

11.(2022 山东菏泽市二模)为了培养孩子的终身锻炼习惯,小明与小红的父亲与他们约定周一到周日每天的锻炼时间不能比前一天少.为了监督两人锻炼的情况,父亲记录了他们某周内每天的锻炼时间(单位:min),如下表所示,其中小明周日的锻炼时间a忘了记录,但知道$36 \leqslant a \leqslant 60, a \in \mathbf{Z}$.

	周一	周二	周三	周四	周五	周六	周日
序号x	1	2	3	4	5	6	7
小明的锻炼时间y/min	16	20	20	25	30	36	a
小红的锻炼时间z/min	16	22	25	26	32	35	35

(1)求这一周内小明锻炼的总时间不少于小红锻炼的总时间的概率;

(2)根据小明这一周前 6 天的锻炼时间,求其锻炼时间y关于序号x的线性回归方程,并估计小明周日锻炼时间a的值.

参考公式和数据:回归方程$\hat{y} = \hat{b}x + \hat{a}$中斜率与截距的最小二乘估计公式分别为

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

$1 \times 16 + 2 \times 20 + 3 \times 20 + 4 \times 25 + 5 \times 30 + 6 \times 36 = 582; 1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2 = 91$.