主编 王 静 终审 李双孝
责编 卢晓丽 照排 仇晓凤

(上接第3版)

提升练

时间:45分钟 满分:64分

素养名称	数学抽象	逻辑推理	数学运算	直观想象	数学建模	数据分析
题号	1,2,3,4		5,6,8,9,10			5,7,8,9,10

一、选择题(本大题共6小题,每小题5分,共30分)

1.已知两个变量x,y之间具有相关关系,现选用a,b,c,d四个模型得到相应的回归方程,并计算得到了相应R^2的值分别为$R_a^2=0.80$,$R_b^2=0.98$,$R_c^2=0.93$,$R_d^2=0.86$,那么拟合效果最好的模型为(　　).

A.a　　B.b

C.c　　D.d

2.对具有线性相关关系的两个变量建立的回归方程$\hat{y}=\hat{b}x+\hat{a}$中,$\hat{a}$可以(　　).

A.小于0　　B.大于0

C.不等于0　　D.以上都有可能

3.(2019·宁县第二中学高二期中(文))在回归分析中,相关指数R^2的值越小,说明残差平方和(　　)

A.越小　　B.越大

C.可能大也可能小　　D.以上都不对

4.有下列说法:

①在残差图中,残差点比较均匀地落在水平的带状区域内,说明选用的模型比较合适;②用R^2可以刻画回归的效果,R^2值越小说明模型的拟合效果越好;③比较两个模型的拟合效果,可以比较残差平方和大小,残差平方和越小的模型拟合效果越好.其中说法正确的是(　　).

A.①②　　B.②③

C.①③　　D.①②③

5.(2019·云南高二期中(文))已知表7中数据y与x有较好的线性关系,通过计算得到y关于x的线性回归方程为$\hat{y}=1.05x+\hat{a}$,则相应于下列各点的残差中绝对值最小的是(　　).

表7

x	2	4	6	8	10
y	4	6	9	10	12.5

A.(2,4)　　B.(4,6)

C.(8,10)　　D.(10,12.5)

6.由变量y与x的10组统计数据得到的回归模型中,$R^2=0.95$,又知残差平方和为120.53,那么$\sum_{i=1}^{10}(y_i-\hat{y})^2$的值为(　　).

A.241.06　　B.2410.6

C.253.08　　D.2530.8

二、填空题(本大题共2小题,每小题5分,共10分)

7.(2019·辽宁高二期末(理))若对甲、乙、丙3组不同的数据作线性相关性检验,得到这3组数据的线性相关系数依次为0.83,0.72,-0.90,则线性相关程度最强的一组是_____.(填甲、乙、丙中的一个)

8.(2019·贵州省思南中学高二月考(文))设某大学的女生体重y(单位:kg)与身高x(单位:cm)具有线性相关关系,根据一组样本数据$(x_i,y_i)(i=1,2,\cdots,n)$,用最小二乘法建立的回归方程为$\hat{y}=0.85x-85.71$,那么针对某个体(170,58)的残差是_____.

三、解答题(本大题共2小题,共24分,解答应写出必要的文字说明、证明过程或演算步骤)

9.(12分)某兴趣小组欲研究昼夜温差大小与患感冒人数多少之间的关系,他们分别到气象局与某医院抄录了1至6月份每月10号的昼夜温差情况与因患感冒而就诊的人数,得到表8:

表8

日期	1月10日	2月10日	3月10日	4月10日	5月10日	6月10日
昼夜温差x(℃)	10	11	13	12	8	6
就诊人数y(人)	22	25	29	26	16	12

该兴趣小组确定的研究方案是:先从这六组数据中选取2组,用剩下的4组数据求线性回归方程,再用被选取的2组数据进行检验.

(1)若选取的是1月与6月的两组数据,请根据2至5月的数据,求出y关于x的线性回归方程$\hat{y}=\hat{b}x+\hat{a}$;

(2)若由线性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人,则认为得到的线性回归方程是理想的,试问该小组所得线性回归方程是否理想?

10.(12分)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列于表9、表10.

表9

停车距离d(米)	(10,20]	(20,30]	(30,40]	(40,50]	(50,60]
频数	26	a	b	8	2

表10

平均每毫升血液酒精含量x毫克	10	30	50	70	90
平均停车距离y米	30	50	60	70	90

已知表9数据的中位数估计值为26,回答以下问题.

(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10的数据计算y关于x的回归方程$\hat{y}=\hat{b}x+\hat{a}$;

(3)该测试团队认为: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)中的回归方程,预测当每毫升血液酒精含量大于多少毫克时为"醉驾"?

(附:回归方程$\hat{y}=\hat{b}x+\hat{a}$中,$\hat{b}=\dfrac{\sum_{i=1}^{n}x_iy_i-n\bar{x}\cdot\bar{y}}{\sum_{i=1}^{n}x_i^2-n\bar{x}^2}$,$\hat{a}=\bar{y}-\hat{b}\bar{x}$)

(参考答案见下期)

❖印刷/河南鑫泽印刷有限责任公司 地址/郑州市马寨工业园先锋路3号 ❖本报四开四版 ❖订阅/全国各地邮局(所) ❖每学期24期 单价1.25元 学期定价30.00元

学习报
高中数学

《选修1-2》编辑计划

出版时间:2020.1
期数:4期
内容:
25.回归分析的基本思想及其初步应用
26.独立性检验的基本思想及其初步应用
27.统计案例综合
28.合情推理与演绎推理

出版时间:2020.2
期数:4期
内容:
29.直接证明与间接证明
30.推理与证明综合
31.数系的扩充与复数的概念、复数的几何意义
32.复数代数形式的四则运算

出版时间:2020.3
期数:4期
内容:
33.复数综合
34.流程图与结构图
35~36.选修1-2综合复习,各章节内容回顾、复习导航、能力提高与综合复习测试题

用稿声明

《学习报》是一份多学科、多层次、多角度面向全国大中小学生及教育工作者的教辅类报纸。本报的办报方针是"紧扣课标及教材,针对年级特点和学科特色同步辅导"。本报旨在拓宽学生阅读视野,激发学习兴趣,培养高尚的人格和高雅的情趣,提升综合能力。本报诚征各学科各类教育、教辅稿件,撰稿、荐稿者请注意以下事项:

1.稿件一经采用,即视为同意将网络传播权及电子发行的权利授予本报。本报一次性支付的稿酬已包含上述授权的使用费。

2.推荐或稿件中选用他人文章时,请务必注明原稿件作者姓名和选文出处,否则不予刊登。若原出版者注明"未经许可,不得转载",请勿推荐,来稿一律不收。推荐类稿件一经采用,本报将依据《中华人民共和国著作权法》之规定给原作者和推荐者奉寄稿酬和样报。

3.由于部分作者通讯地址不详,见报后请与我报及时联系,以便奉寄稿酬和样报。

电话:0351-3239626(总编办)或请与各编辑部联系咨询

学习报® 高中数学
全国教辅类一级报纸
人教A(文科)
选修1-2
2019-2020
第25期
总第1601期
2020年1月1日
http://www.xuexibao.com
山西出版传媒集团主管 山西三晋报刊传媒集团主办 学习报社编辑出版 总编辑 苗俊青 国内统一刊号 CN14-0708/(F)

传道授业

直击残差分析

□河南 王恒河

一、概念辨析

1.残差分析

在研究两个变量间的关系时,首先要根据散点图来粗略判断它们是否线性相关,是否可以用线性回归模型来拟合数据.然后,可以通过残差$\hat{e}_1$,$\hat{e}_2$,…,$\hat{e}_n$来判断模型拟合的效果,并判断原始数据中是否存在可疑数据.这方面的分析工作称为残差分析.

残差分析一般有两种方法:

(1)作残差图;

(2)利用相关指数R^2来刻画回归效果.

2.残差与残差平方和

$\hat{e}_i=y_i-\hat{y}_i=y_i-\hat{b}x_i-\hat{a}$,$i=1,2,\cdots,n$,$\hat{e}_i$称为相应于点$(x_i,y_i)$的残差.$\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$称为残差平方和.残差平方和越小,说明模型的拟合效果越好.

3.残差图

利用图形来分析残差特性,作图时纵坐标为残差,横坐标可以选样本编号,或身高数据,或体重的估计值等,这样作出的图形称为残差图.在残差图中有些样本点的残差比较大,需要确认在采集这些样本点的过程中是否有人为错误,若有错误,予以纠正,然后再重新利用线性回归模型拟合数据;如果数据采集没有错误,则需要寻找其他的原因.另外,残差点比较均匀地落在水平的带状区域中,说明选用的模型比较合适.这样的带状区域的宽度越窄,说明模型拟合精度越高,回归方程的预报精度也就越高.

二、实际应用

例1 假定小麦基本苗数x(千棵)与成熟期有效穗y之间存在相关关系,今测得5组数据,如表1:

表1

x	15.0	25.8	30.0	36.6	44.4
y	39.4	42.9	42.9	43.1	49.2

(1)求y与x之间的回归方程,并对于基本苗数56.7(千棵)预报有效穗;

(2)计算各组残差,并计算残差平方和;

(3)求R^2,并说明残差变量对有效穗的影响占百分之几.

解:(1)由已知表1所给数据及散点图(图略)可看出,y与x有比较好的线性相关关系,因此可以用线性回归方程刻画它们之间的关系,可求得回归方程为$\hat{y}=0.291x+34.664$.

因此对于基本苗数56.7(千棵),由回归方程可以预报其有效穗为$\hat{y}=0.291\times56.7+34.664=51.1637$.

(2)由于$\hat{e}_1=y_1-\hat{y}_1$,所以$\hat{e}_1=0.371$,$\hat{e}_2=0.7282$,$\hat{e}_3=-0.494$,$\hat{e}_4=-2.2146$,$\hat{e}_5=1.6156$,所以残差平方和为$\sum_{i=1}^{5}\hat{e}_i^2=0.371^2+0.7282^2+(-0.494)^2+(-22146)^2+1.6156^2\approx8.4266$.

(3)代入公式得$R^2=0.8321$.故解释变量小麦基本苗数对有效穗变化的贡献率为83.21%,残差变量对有效穗的影响占1-83.21%=16.79%.

例2 关于人体的脂肪含量(百分比)和年龄(岁)关系的研究中,研究人员获得了一组数据,如表2所示:

表2

年龄 x	23	27	39	41	45	49	50
脂肪含量 y	9.5	17.8	21.2	25.9	27.5	26.3	28.2
年龄 x	53	54	56	57	58	60	61
脂肪含量 y	29.6	30.2	31.4	30.8	33.5	35.2	34.6

(1)求y与x之间的回归方程;

(2)求R^2,并说明其含义;

(3)请给出37岁人的脂肪含量的预测值.

解:(1)代入公式可求得$\hat{b}=0.577$,$\therefore\hat{a}=-0.448$,

$\therefore$所求的回归方程为$\hat{y}=0.577x-0.448$.

(2)残差平方和为$\sum_{i=1}^{14}\hat{e}_i^2=\sum_{i=1}^{14}(y_i-\hat{y}_i)^2\approx37.1916$.

又$\sum_{i=1}^{14}(y_i-\bar{y})^2\approx644.9921$,

$\therefore R^2=1-\dfrac{37.1916}{644.9921}=0.9423$.

其表明年龄解释了94.23%的脂肪含量的变化.

(3)当$x=37$时,$\hat{y}=0.577\times37-0.448=20.901$.

答疑解惑

回归分析问与答

□江西 王 翊

回归分析这部分知识,是初中数学中统计案例和高中《必修3》第一章统计的深化和拓展.同学们在学习中还有哪些疑惑呢?

问:什么是相关关系?什么是回归分析?

答:自变量取值一定时,因变量取值带有一定随机性的两个变量之间的关系叫作相关关系,相关关系是一种非确定关系;对具有相关关系的两个变量进行统计分析的方法叫回归分析.

问:相关关系和函数关系有何联系和区别?

答:联系:都是指两个变量的关系;区别:函数关系中当一个变量确定时,另一个变量也确定了,而相关关系中一个变量确定后,另一变量不是确定的.

问:散点图和样本点的中心是什么?

答:散点图是表示具有相关关系的两个变量的一组数据的图形;对于n个样本数据(x_1,y_1),(x_2,y_2),…,(x_n,y_n)来说,$(\bar{x},\bar{y})$称为其样本点的中心,回归直线一定通过样本点的中心.

问:判断解释变量x与预报变量y是否具有线性相关关系,先作出散点图,从点的分布特征来判定是否线性相关.那么,如果作图不准,出现误差怎么办?怎样才能更好地判定两个变量相关关系的强弱?

答:作相关性检验,通过作散点图,并观察所给的数据列成的点是否在一条直线的附近来判定,这样做既直观又方便,因而对解决相关性检验问题比较常用.但在作图中,由于存在误差,有时很难说这些点是不是分布在一条直线的附近,这时就很难判断两个变量之间是否具有相关关系.因此,给定样本数据$(x_i,y_i)(i=1,2,\cdots,n)$,单纯地由散点图判定其是否大致在一条直线附近主观性太强.回归分析时还通常用R^2来检验两个变量之间线性相关关系的强弱.

问:如何比较两个不同回归模型的拟合效果?

答:我们可以用R^2来刻画回归交果,$R^2=1-\dfrac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}$中的$\sum_{i=1}^{n}(y_i-\bar{y})^2$是确定的数.$R^2$越大,残差平方和$\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$越小,即模型的拟合效果越好;反之,$R^2$越小,模型的拟合效果越差.在线性回归模型中,$R^2$表示解释变量对于预报变量的贡献率,$R^2$越接近于1,表示回归的效果越好.

问:如何用回归分析的方法对非线性回归问题进行统计分析呢?

答:对于非线性回归问题进行回归分析的方法是:

(1)若问题中已给出经验公式,这时可以将解释变量进行交换(换元),将变量的非线性关系转化为线性关系,从而将问题化为线性回归分析问题来解决.

(2)若问题中没有给出经验公式,需要我们画出已知数据的散点图,通过与各种函数(如指数函数、对数函数、幂函数等)的图象作比较,选择一种与这些散点拟合得最好的函数,然后采用适当的变量变换,将问题转化为线性回归分析问题来解决.

微信扫码,关注学习报公众号,了解教育资讯,伴学快乐成长。

❖广告经营许可证号/1400004000044 ❖本报向全国各省(市)级教研员赠报,欢迎来电咨询 ❖随时征订热线/0351-3239622 邮政订户服务热线/0351-3239623

§1.1 回归分析的基本思想及其初步应用

1.相关系数是度量(　　).

A.两个变量之间线性相关关系的强度

B.散点图是否显示有意义的模型

C.两个变量之间是否存在因果关系

D.两个变量之间是否存在关系

2.在画两个变量的散点图时,下面哪个叙述是正确的(　　).

A.预报变量在x轴上,解释变量在y轴上

B.解释变量在x轴上,预报变量在y轴上

C.可以选择两个变量中任意一个在x轴上

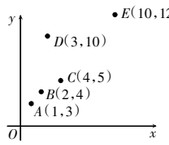
D.可以选择两个变量中任意一个在y轴上

3.若某地财政收入x与支出y满足线性回归模型$y=bx+a+e$(单位:亿元),其中$b=0.8$,$a=2$,$|e|<0.5$,若今年该地区财政收入10亿元,则年支出预计不会超过(　　).

A.10亿　　B.9亿

C.10.5亿　　D.9.5亿

4.(2019·福建高二期末(文))如图所示,5组数据(x,y)中去掉$D(3,10)$后,下列说法错误的是(　　).



A.残差平方和变大

B.相关系数$|r|$变大

C.相关指数R^2变大

D.解释变量x与预报变量y的相关性变强

5.(2019·安徽省太和第一中学高二月考)某工厂某产品产量x(千件)与单位成本y(元)满足回归直线方程$\hat{y}=77.36-1.82x$,则以下说法中正确的是(　　).

A.产量每增加1000件,单位成本约下降1.82元

B.产量每减少1000件,单位成本约下降1.82元

C.产量每增加1000件,单位成本约上升1.82元

D.产量每减少1000件,单位成本约上升1.82元

6.(2019·安徽高二期末)已知变量x,y之间的一组数据如表1:

表1

x	1	3	5	7
y	2	3	4	5

由散点图可知变量x,y线性相关,则y与x的回归直线必经过点(　　).

A.(2,2.5)　　B.(3,3)

C.(4,3.5)　　D.(6,4.8)

7.下列关系中正确的是_____(填序号).

①函数关系是一种确定性关系;

②相关关系是一种非确定性关系;

③回归分析是对具有函数关系的两个变量进行统计分析的一种方法;

④回归分析是对具有相关关系的两个变量进行统计分析的一种常用方法.

8.表2是降耗技术改造后生产甲产品过程中记录的产量x(吨)与相应的生产能耗y(吨标准煤)的几组对应数据,根据表中提供的数据,求出y关于x的线性回归方程$\hat{y}=0.7x+0.35$,那么表2中m的值为_____.

表2

x	3	4	5	6
y	2.5	m	4	4.5

9.回归直线方程$\hat{y}=2-1.5x$,当变量x增加1个单位时,y平均_____个单位.

10.(2019·内蒙古杭锦后旗奋斗中学高二月考(文))在研究两个变量的线性相关关系时,观察散点图发现样本点集中于某一条指数曲线的周围,令$z=\ln y$,求得回归直线方程$\hat{z}=0.25x-2.58$,则该模型的回归方程为_____.

11.(2018·湖北高二期末)一个车间为了规定工作原理,需要确定加工零件所花费的时间,为此进行了5次试验,收集数据如表3:

表3

零件数x(个)	10	20	30	40	50
加工时间y(分钟)	64	69	75	82	90

由表中数据,求得线性回归方程$\hat{y}=0.65x+\hat{a}$,根据回归方程,预测加工70个零件所花费的时间为_____分钟.

12.某单位为了了解用电量y度与气温x℃之间的关系,随机统计了某4天的用电量与当天气温,并制作了表4.

表4

气温(℃)	18	13	10	-1
用电量(度)	24	34	38	64

由表中数据得线性回归方程$\hat{y}=\hat{b}x+\hat{a}$中$\hat{b}=-2$,预测当气温为-4℃时,求用电量的度数约为多少?

13.(2019·贵州高二月考)平顶山市公安局交警支队依据《中华人民共和国道路交通安全法》第90条规定:所有主干道路凡机动车途经十字路口或斑马线,无论转弯或者直行,遇有行人过马路,必须礼让行人,违反者将被处以100元罚款,记3分的行政处罚.如表5是本市一主干路段监控设备所抓拍的5个月内,机动车驾驶员不"礼让斑马线"的行为统计数据:

表5

月份	1	2	3	4	5
违章驾驶员人数	120	105	100	90	85

(1)请利用所给数据求违章人数y与月份x之间的回归直线方程$\hat{y}=\hat{b}x+\hat{a}$;

(2)预测该路段7月份的不"礼让斑马线"违章驾驶员人数.

(参考公式:$\hat{b}=\dfrac{\sum\limits_{i=1}^{n}x_iy_i-n\bar{x}\bar{y}}{\sum\limits_{i=1}^{n}x_i^2-n\bar{x}^2}=\dfrac{\sum\limits_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum\limits_{i=1}^{n}(x_i-\bar{x})^2}$,$\hat{a}=\bar{y}-\hat{b}\bar{x}$.)

(参考答案见下期)

学习报 高中数学

栏目设置

第一版 导·教材解读

本版主要栏目:

传道授业:对本期的知识学生应该怎样去学习才能理解的更到位更准确,(1)讲解一些对课本上难以理解的知识点,(2)课本外的但在课堂上需要补充的知识,(3)关于本期内容前后衔接的一些知识点的介绍。

高屋建瓴:对于基础知识涉及到的基础方法应该怎么用,结合适当的例题讲解,让大多数学生达到轻松掌握基本知识及其解法的目的。

答疑解惑:通过师生问答的形式,将对本期知识中有疑问的地方解惑,使学生能在轻松的气氛中掌握知识。

妙趣横生:从学生的知识结构和思维发展水平的实际出发,拓宽学生的知识视野,以增加数学学习的趣味性。

数学文化:以数学发展史为主,展示丰富多彩的数学文化,使读者体会到数学是人类智慧的结晶,并从中受到启迪,开拓思路。

第二、三版 练·能力提升

本版主要栏目:

互动课堂:通过"基础训练""巩固提升"两个层次让学生了解所学内容在考试中的考查方式,以什么形式、题材、材料来设计问题。

同步分层能力测试题:紧跟教学进度,不可使用历年高考原题,可采用一部分高考或竞赛改编题(改编题请标明出处)。

章节能力测试题:对整章内容的综合测试。

模块综合能力测试题:对对应模块所有知识点的综合测试。

第四版 达·知识迁移

本版主要栏目:

思想方法:针对当期知识点,根据新课标所要求学生掌握的常用数学思想方法进行重点启发与培养,以帮助学生形成良好的数学思维习惯。

茅塞顿开:通过例题的求解过程,阐述解题策略及思路的形成过程。目的在于培养学生自主学习的能力,以及对问题的深入掌握。

一击即破:针对有些问题的一些特殊的巧妙的方法解析,侧重在传统思路基础上的巧与妙,重点在思考上。

举一反三:以课本典型例题或习题为主,对其背景、限制条件、设问方式等进行变换探究,使同学们能够打开视野,触类旁通。

知行合一:立足于数学知识在实际生产和生活中的应用,以及在其他学科领域中的应用,把实际问题转化为数学模型来讲求解,以期达到知行合一的目的。

金榜题名:针对某个知识点在高考中的题型进行回顾、展望,并探求其求解策略及技巧,让学生提前了解并熟悉高考的相关题型。

运筹帷幄:对于本期基础知识涉及到的基础方法应该怎么用,结合适当的例题讲解,化难为易,让大多数学生达到轻松掌握基本知识及其解法的目的。或剖析对某个知识点与其他知识点的交汇综合,追求思路方法的拓展性,目的在于培养学生完整的知识系统。

毫厘千里:对各种典型的错题进行归类,提炼出针对某类错误的解题策略。将常考、易错和能举一反三的典型试题按专题分类,帮学生掌握考点,做少量的题,掌握一类题的典型解法。

同步分层 能力测试题 一

(测试范围:回归分析的基本思想及其初步应用)

基础练

时间:45分钟　满分:86分

素养名称	数学抽象	逻辑推理	数学运算	直观想象	数学建模	数据分析
题号	1,2,3		4,5,6,11			6,7,8,9,10,11,12

一、选择题(本大题共6小题,每小题5分,共30分)

1.有下列关系:①人的年龄与他拥有的财富之间的关系;②曲线上的点与该点的坐标之间的关系;③苹果的产量与气候之间的关系;④森林中同一种树木其横断面直径与高度之间的关系,其中具有相关关系的是(　　).

A.①②③　　B.①②

C.②③　　D.①③④

2.关于残差和残差图,下列说法正确的是(　　).

(1)残差就是随机误差;

(2)残差图的纵坐标是残差;

(3)残差点均匀分布的带状区域的宽度越窄,说明模型拟合精度越高;

(4)残差点均匀分布的带状区域的宽度越窄,说明模型拟合精度越低.

A.(1)(2)　　B.(3)(4)

C.(2)(3)　　D.(2)(4)

3.(2019·广东高二期末(文))下列说法错误的是(　　).

A.相关关系是一种非确定性关系

B.线性回归方程对应的直线$\hat{y}=\hat{b}x+\hat{a}$,至少经过其样本数据点(x_1,y_1),(x_2,y_2),…,(x_n,y_n)中的一个点

C.在残差图中,残差点分布的带状区域的宽度越狭窄,其模型拟合的精度越高

D.在回归分析中,R^2为0.98的模型比R^2为0.80的模型拟合的效果好

4.在一次试验中,测得(x,y)的四组值分别是$A(1,2)$,$B(2,3)$,$C(3,4)$,$D(4,5)$,则y与x之间的回归直线方程为(　　).

A.$\hat{y}=x+1$　　B.$\hat{y}=x+2$

C.$\hat{y}=2x+1$　　D.$\hat{y}=x-1$

5.(2019·广东高二期末(理))从某大学中随机选取8名女大学生,其身高x(单位:cm)与体重y(单位:kg)数据如表1:

表1

x	165	165	157	170	175	165	155	170
y	48	57	50	54	64	61	43	59

若已知y与x的线性回归方程为$\hat{y}=0.85x-85.71$,那么选取的女大学生身高为175cm时,相应的残差为(　　).

A.-0.96　　B.0.96

C.63.04　　D.-4.04

6.某考察团对全国10大城市人均平均工资x与居民人均消费y进行统计调查,发现y与x具有相关关系,且回归方程为$\hat{y}=0.66x+1.562$(单位:千元).若某城市居民人均消费为7.675,估计该城市消费额占人均工资收入的百分比为(　　).

A.66%　　B.72.3%

C.67.3%　　D.83%

二、填空题(本大题共2小题,每小题5分,共10分)

7.某车间为了规定工时定额,需要确定加工零件所花费的时间,为此进行了5次试验,根据收集到的数据(如表2),由最小二乘法求得回归方程$\hat{y}=0.67x+54.9$.

表2

零件数x(个)	10	20	30	40	50
加工时间y(min)	62		75	81	89

现发现表中有一个数据模糊看不清,请你推断出该数据的值为_____.

8.(2019·天津高二期末)为了了解家庭月收入x(单位:千元)与月储蓄y(单位:千元)的关系,从某居民区随机抽取10个家庭,根据测量数据的散点图可以看出x与y之间具有线性相关关系,其回归直线方程为$\hat{y}=0.3x-0.4$,若该居民区某家庭月收入为7千元,据此估计该家庭的月储蓄为_____千元.

三、解答题(本大题共4小题,共46分,解答应写出必要的文字说明、证明过程或演算步骤)

9.(10分)在某地区的12~30岁居民中随机抽取了10个人的身高和体重的统计资料如表3:

表3

身高(cm)	143	156	159	172	165	171	177	161	164	160
体重(kg)	41	49	61	79	68	69	74	69	68	54

根据上述数据,画出散点图并判断居民的身高和体重之间是否有相关关系.

10.(12分)已知表4是x与y之间的一组数据,求关于y与x的线性回归方程$\hat{y}=\hat{a}+\hat{b}x$中的$\hat{a}$与$\hat{b}$的关系.

表4

x	0	1	2	3
y	1	3	5	7

11.(12分)某电脑公司有6名产品推销员,其工作年限与年推销金额数据如表5:

表5

推销员编号	1	2	3	4	5
工作年限x(年)	3	5	6	7	9
推销金额y(万元)	2	3	3	4	5

(1)求年推销金额y关于工作年限x的线性回归方程;

(2)若第6名推销员的工作年限为11年,试估计他的年推销金额.

12.(12分)(2019·吉林蛟河一中高二期中(文))某工厂为了对新研发的一种产品进行合理定价,将该产品按事先拟定的价格进行试销,得到如表6的数据:

表6

单价x(元)	14	16	18	20	22
销量y(件)	12	10	7	5	3

(1)求回归直线方程$\hat{y}=\hat{b}x+\hat{a}$.

(2)利用R^2刻画回归效果.

(下转第4版)

❖报社地址/太原市三墙路坝陵北街盛世华庭A1座21层 邮编/030009 ❖编辑质量监督联系电话/0351-3239626

主编 王 静 终审 李双孝
责编 卢晓丽 照排 仇晓凤

(上接第3版)

提升练

时间:45分钟 满分:64分

素养名称	数学抽象	逻辑推理	数学运算	直观想象	数学建模	数据分析
题号	1,5		2,3,8,10	6		2,3,4,6,7,8,9,10

一、选择题(本大题共6小题,每小题5分,共30分)

1.下列说法正确的有(　　).

①对事件A与B的检验无关时,即两个事件互不影响;

②事件A与B关系越密切,则K^2的观测值k就越大;

③K^2的大小是判定事件A与B是否有关的唯一根据;

④若判定两事件A与B有关,则A发生B一定发生.

A.1个　　B.2个

C.3个　　D.4个

2.假设有两个分类变量X和Y的2×2列联表为表10:

表10

	y_1	y_2	总计
x_1	a	10	a+10
x_2	c	30	c+30
总计	60	40	100

对同一样本,以下数据能说明X与Y有关系的可能性最大的一组为(　　).

A.a=45,c=15　　B.a=40,c=20

C.a=35,c=25　　D.a=30,c=30

3.在某医院,因为患心脏病而住院的665名男性病人中,有214人秃顶,而另外772名不是因为患心脏病而住院的男性病人中有175人秃顶,则K^2的观测值k=(　　).

A.16.373　　B.15.264

C.17.365　　D.9.632

4.(2019·贵州省安顺市第一高级中学高二期末(文))为研究某两个分类变量是否有关系,根据调查数据计算得到$K^2\approx15.968$,因为$P(K^2\geqslant10.828)=0.001$,则断定这两个分类变量有关系,那么这种判断犯错误的概率不超过(　　).

A.0.1　　B.0.001

C.0.01　　D.0.05

5.两个分类变量X,Y,它们的可能取值分别为$\{x_1,x_2\}$和$\{y_1,y_2\}$,其样本频数列联表如表11:

表11

	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

若两个分类变量X,Y独立,则下列结论不成立的是(　　).

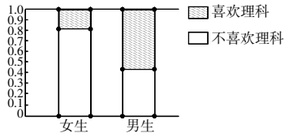
A.$ad\approx bc$

B.$a+c\approx b+d$

C.$\frac{a}{a+b}=\frac{c}{c+d}$

D.$\frac{(a+b+c+d)(ad-bc)^2}{(a+c)(b+d)(a+b)(c+d)}\approx0$

6.如下图所示是调查某地区男、女中学生喜欢理科的等高条形图,阴影部分表示喜欢理科的百分比,从图中可以看出(　　).



A.性别与喜欢理科无关

B.女生中喜欢理科的比例为80%

C.男生比女生喜欢理科的可能性大些

D.男生中不喜欢理科的比例为60%

二、填空题(本大题共2小题,每小题5分,共10分)

7.在独立性检验中,选用K^2的观测值k,用其取值大小来推断独立性是否成立,当k满足条件_____时,在出错的概率不超过0.01的前提下,我们说事件A与B有关.

8.(2018·山东省邹城市第一中学高二期中(文))博鳌亚洲论坛2018年年会于4月8日至11日在海南博鳌举行.为了搞好对外宣传工作,会务组选聘了50名记者担任对外翻译工作如表12的"性别与会俄语"的2×2列联表中,$a-b+d$=_____.

表12

	会俄语	不会俄语	总计
男	a	b	20
女	6	d	
总计	18		50

三、解答题(本大题共2小题,共24分,解答应写出必要的文字说明、证明过程或演算步骤)

9.(12分)考察黄烟经过药物处理跟发生青花病的关系,得到如下数据:在试验的470株黄烟中,经过药物处理的黄烟有25株发生青花病,60株没有发生青花病.未经过药物处理的有185株发生青花病,200株没有发生青花病,试推断经过药物处理跟发生青花病是否有关系.

10.(12分)性格色彩学创始人乐嘉曾是江苏电视台当红节目"非诚勿扰"的特约嘉宾,他的点评视角独特,语言犀利,给观众留下了深刻的印象,某报社为了了解观众对乐嘉的喜爱程度,随机调查了观看了该节目的140名观众,得到表13所示的列联表:(单位:名)

表13

	男	女	总计
喜爱	40	60	100
不喜爱	20	20	40
总计	60	80	140

附表14

表14

$P(K^2\geqslant k_0)$	0.10	0.05	0.025	0.010	0.005
k_0	2.705	3.841	5.024	6.635	7.879

(1)从这60名男观众中按对乐嘉是否喜爱采取分层抽样,抽取一个容量为6的样本,问样本中喜爱与不喜爱的观众各有多少名?

(2)根据以上列联表,问能否在犯错误的概率不超过0.025的前提下认为观众性别与喜爱乐嘉有关?(精确到0.001)

(3)从(1)中的6名男性观众中随机选取2名作跟踪调查,求选到的2名观众都喜爱乐嘉的概率.

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

(参考答案见下期)

学习报 高中数学

《选修1-2》编辑计划

出版时间:2020.1
期数:4期
内容:
25.回归分析的基本思想及其初步应用
26.独立性检验的基本思想及其初步应用
27.统计案例综合
28.合情推理与演绎推理

出版时间:2020.2
期数:4期
内容:
29.直接证明与间接证明
30.推理与证明综合
31.数系的扩充与复数的概念、复数的几何意义
32.复数代数形式的四则运算

出版时间:2020.3
期数:4期
内容:
33.复数综合
34.流程图与结构图
35~36.选修1-2综合复习,各章节内容回顾、复习导航、能力提高与综合复习测试题

用稿声明

《学习报》是一份多学科、多层次、多角度面向全国大中小学生及教育工作者的教辅类报纸。本报的办报方针是"紧扣课标及教材,针对年级特点和学科特色同步辅导"。本报旨在拓宽学生阅读视野,激发学习兴趣,培养高尚的人格和高雅的情趣,提升综合能力。本报诚征各学科各类教育、教辅稿件,撰稿、荐稿者请注意以下事项:

1.稿件一经采用,即视为同意将网络传播权及电子发行的权利授予本报。本报一次性支付的稿酬已包含上述授权的使用费。

2.推荐或稿件中选用他人文章时,请务必注明原稿件作者姓名和选文出处,否则不予刊登。若原出版者注明"未经许可,不得转载",请勿推荐,来稿一律不收。推荐类稿件一经采用,本报将依据《中华人民共和国著作权法》之规定给原作者和推荐者奉寄稿酬和样报。

3.由于部分作者通讯地址不详,见报后请与我报及时联系,以便奉寄稿酬和样报。

电话:0351-3239626(总编办)或请与各编辑部联系咨询

学习报® 高中数学
全国教辅类一级报纸
人教A(文科)
选修1-2
2019-2020
第26期
总第1602期
2020年1月8日
http://www.xuexibao.com
山西出版传媒集团主管　山西三晋报刊传媒集团主办　学习报社编辑出版　总编辑 苗俊青　国内统一刊号 CN14-0708/(F)

传道授业

诠释列联表与"K^2"

□山西　柴有茂

一、列联表:列出的两个分类变量X和Y,它们的取值分别为$\{x_1,x_2\}$和$\{y_1,y_2\}$的样本频数表称为2×2列联表(如表1)

表1

	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

二、K^2

构造随机变量$K^2=\frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$(其中$n=a+b+c+d$),如果$K^2$的观测值$k$较大,说明其发生(无关系)的概率很小,此时不接受假设,也就是两分类变量是有关系的(称小概率事件发生);如果K^2的观测值k较小,此时接受假设,说明两分类变量是无关系的.其思想方法类似于数学上的反证法.

三、得到K^2的观测值k常与以下几个临界值加以比较

如果k>2.706,则在犯错误的概率不超过0.1的前提下认为两个分类变量X和Y有关系;

如果k>3.841,就有则在犯错误的概率不超过0.05的前提下认为两个分类变量X和Y有关系;

如果k>6.635,则在犯错误的概率不超过0.01的前提下认为两个分类变量X和Y有关系;

如果$k\leqslant2.706$,就认为在犯错误的概率不超过0.1的前提下不能推断分类变量X和Y有关系.

像这样利用随机变量K^2来判断"两个分类变量有关系"的方法称为独立性检验.

在实际问题中,我们可以利用独立性检验来考察两个分类变量是否有关系,并且能较为准确地给出这种判断的可靠程度.

四、应用举例

1.利用K^2完成列联表

例1　有两个分类变量X,Y,其2×2列联表如表2所示:

表2

	y_1	y_2
x_1	a	20-a
x_2	15-a	30+a

其中a,15-a均为大于5的整数,则a取何值时,在犯错误的概率不超过0.1的前提下认为X与Y之间有关系?

解:查表可知,要使X与Y之间在犯错误的概率不超过0.1的前提下认为有关系,则$k\geqslant2.706$,由题意,得

$k=\frac{65[a(30+a)-(20-a)(15-a)]^2}{20\times45\times15\times50}=\frac{65(65a-300)^2}{20\times45\times15\times50}=\frac{13(13a-60)^2}{60\times90}$.

由$k\geqslant2.706$,解得$a\geqslant7.19$或$a\leqslant2.04$.

又a>5且15-a>5,$a\in\mathbf{Z}$,∴a=8或a=9.

故当a等于8或9时,在犯错误的概率不超过0.1的前提下认为X与Y之间有关系.

2.建立列联表,用K^2进行判断

例2　为了比较放射性疗法和手术疗法对治疗喉癌的疗效,某医院做了如下试验:从41位喉癌患者中随机选取23位用手术疗法,其余18位用放射性疗法.经过一段时间的治疗后,发现手术疗法中有21位患者的病情得到控制,放射性疗法中有15位患者的病情得到控制.要使结论的可靠性不低于95%,检验手术疗法和放射性疗法的疗效是否相同.

解:先建立2×2列联表如表3:

表3

	病情得到控制	病情未得到控制	总计
手术疗法	21	2	23
放射性疗法	15	3	18
总计	36	5	41

提出统计假设H_0:手术疗法和放射性疗法的疗效相同.

由表中数据可得$k=\frac{41\times(21\times3-2\times15)^2}{23\times18\times36\times5}\approx0.5992$.

当H_0成立时,$k\approx0.5992<2.706$,

所以要使结论的可靠性不低于95%,我们不能否定统计假设,即不能作出手术疗法和放射性疗法的疗效相同的结论.

秉要执本

帮你认清"K^2"

□江西　韩文美

独立性检验是数理统计的一种方法,随机变量K^2是判断分类变量是否相关的重要"统计数据",下面举例说明.

例1　下列说法中正确的个数有(　　).

①K^2的观测值k越大,两个事件的相关性就越大;

②事件A与B关系越密切,则K^2的观测值k就越大;

③K^2的大小是判定事件A与B是否相关的唯一根据;

④K^2是用来判断两个分类变量是否有关系的随机变量,只对两个分类变量适用;

⑤K^2在任何相互独立的问题中都可以用来检验有关系还是无关系.

A.1　　B.2　　C.3　　D.4

解:K^2是用来判断两个分类变量是否有关系的随机变量,所以④正确,⑤错误;K^2越大,只能说明我们能以更大的把握认为二者有关系,但是不能说明两者的相关的关系就越大,因此①错误,但是事件A与B关系越密切,则K^2的观测值k就越大,②正确;判定两事件是否相关除了公式外,还可以用等高条形图等方法来判定,③错误.故选B.

例2　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,所得数据如表1所示:

表1

	积极支持企业改革	不赞成企业改革	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?

解:根据列联表(表1)可以看出"工作积极"的员工赞成企业改革的比例要大于"工作一般"的员工,所以,从列联表可以粗略估计两种类型的员工对改革的态度有区别.

构造数据,取a=54,b=40,c=32,d=63,用字母代替列联表中的数据可得列联表如表2:

表2

	积极支持企业改革	不赞成企业改革	总计
工作积极	a	b	a+b
工作一般	c	d	c+d
总计	a+c	b+d	a+b+c+d

构造随机变量$K^2=\frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$,把数据代入可得其观测值$k=\frac{189\times(54\times63-40\times32)^2}{94\times95\times86\times103}\approx10.76$,

查表可知,$P(K^2\geqslant7.879)=0.005$,即在犯错误的概率不超过0.005的前提下认为两种类型的员工对改革的态度有区别.

微信扫码,关注学习报公众号,了解教育资讯,伴学快乐成长。

❖印刷/河南鑫泽印刷有限责任公司　地址/郑州市马寨工业园先锋路3号　❖本报四开四版　❖订阅/全国各地邮局(所)　❖每学期24期　单价1.25元　学期定价30.00元

❖广告经营许可证号/1400004000044　❖本报向全国各省(市)级教研员赠报,欢迎来电咨询　❖随时征订热线/0351-3239622　邮政订户服务热线/0351-3239623

§1.2 独立性检验的基本思想及其初步应用

1.(2019·广东高二期末(文))为了判定两个分类变量X和Y是否有关系,应用K^2独立性检验法计算得K^2的观测值为5,又已知$P(K^2\geqslant3.841)=0.05$,$P(K^2\geqslant6.635)=0.01$,则下列说法正确的是(　　).

A.有99%以上的把握认为"X和Y有关系"

B.有99%以上的把握认为"X和Y没有关系"

C.有95%以上的把握认为"X和Y有关系"

D.有95%以上的把握认为"X和Y没有关系"

2.(2019·西藏高二期末)通过随机询问50名性别不同的大学生是否爱好某项运动,得到如表1的列联表,由$K^2=\frac{n(ac-bd)^2}{(a+b)(c+d)(a+c)(b+d)}$,得$K^2=\frac{50\times(20\times15-10\times5)^2}{30\times20\times25\times25}\approx8.333$,参照表2,得到的正确结论是(　　).

表1

	爱好	不爱好	合计
男生	20	5	25
女生	10	15	25
合计	30	20	50

表2

$P(K^2\geqslant k)$	0.010	0.005	0.001
k	6.635	7.879	10.828

A.有99.5%以上的把握认为"爱好该项运动与性别有关"

B.有99.5%以上的把握认为"爱好该项运动与性别无关"

C.在犯错误的概率不超过0.1%的前提下,认为"爱好该项运动与性别有关"

D.在犯错误的概率不超过0.1%的前提下,认为"爱好该项运动与性别无关"

3.表3是一个2×2列联表:

表3

	y_1	y_2	总计
x_1	a	21	73
x_2	8	25	33
总计	b	46	

则表中a,b处的值分别为_____,_____.

4.通过对K^2的统计量的研究得到了若干个临界值,当$K^2\leqslant3.841$时,我们认为(　　).

A.在犯错误的概率不超过0.05的前提下推断X与Y有关系

B.在犯错误的概率不超过0.01的前提下推断X与Y有关系

C.在犯错误的概率不超过0.05的前提下不能推断X与Y有关系

D.不能确定

5.为了检验两个事件A,B是否相关,经过计算得K^2的观测值$k=8.283$,则说明事件A和事件B_____.(填"有关系"或"无关系")

6.某市政府在调查市民收入增减与旅游愿望的关系时,采用独立性检验法抽查了3000人,计算发现K^2的观测值$k=6.023$,根据这一数据查表,市政府断言市民收入增减与旅游愿望有关系,这一断言犯错误的概率不超过_____.

7.关于研究生的一个随机样本给出了关于所获取学位类别与学生性别的分类数据如表4所示:

表4

性别	学位		
	硕士	博士	总计
男	162	27	189
女	143	8	151
总计	305	35	340

根据以上数据,则(　　).

A.可认为性别与获取学位类别有关

B.可认为性别与获取学位类别无关

C.可认为性别决定获取学位的类别

D.以上都是错误的

8.有甲、乙两个班级进行数学考试,按照大于或等于85分为优秀,85分以下为非优秀统计成绩,得到如表5的列联表:

表5

	优秀	非优秀	总计
甲班	10	b	
乙班	c	30	
总计			105

已知在全部105人中随机抽取1人,成绩优秀的概率为$\frac{2}{7}$,则下列说法正确的是(　　).

A.列联表中c的值为30,b的值为35

B.列联表中c的值为15,b的值为50

C.根据列联表中的数据,若犯错误的概率不超过0.05要求,能认为"成绩与班级有关系"

D.根据列联表中的数据,若犯错误的概率不超过0.05要求,不能认为"成绩与班级有关系"

9.(2019·福建高二月考(文))"中国式过马路"存在很大的交通安全隐患,某调查机构为了解路人对"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,得到了表6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,并完成表6的2×2列联表;

表6

	男性	女性	合计
反感	10	b	
不反感	c	8	
合计			30

(2)根据列联表中的数据,判断是否有95%把握认为反感"中国式过马路"与性别有关?

临界值表7:

表7

$P(K^2\geqslant k_0)$	0.400	0.250	0.150	0.100	0.050	0.025
k_0	0.780	1.323	2.072	2.706	3.841	5.024

10.(2019·辽宁高二期末(文))某校数学课外兴趣小组为研究数学成绩是否与性别有关,先统计本校高三年级每个学生一学期数学成绩的平均分(采用百分制),剔除平均分在30分以下的学生后,共有男生300名,女生200名.现采用分层抽样的方法,从中抽取了100名学生,按性别分为两组,并将两组学生成绩分为6组,得到如表8所示的频数分布表.

表8

分数段	[40,50)	[50,60)	[60,70)	[70,80)	[80,90)	[90,100]
男	3	9	18	15	6	9
女	6	4	5	10	13	2

(1)规定80分以上为优分(含80分),请你根据已知条件作出表9的2×2列联表.

表9

	优分	非优分	合计
男生			
女生			
合计			100

(2)根据你作出的2×2列联表判断是否有90%以上的把握认为"数学成绩与性别有关".

(参考答案见下期)

学习报 高中数学

上期《堂堂清》参考答案

§1.1 回归分析的基本思想及其初步应用

1.A.　2.B.　3.C.

4.A.　5.A.　6.C.

7.①②④.

8.3.

9.减少1.5.

10.$\hat{y}=e^{0.25x-2.58}$.

11.102.

12.68.

13.(1)$y=-8.5x+125.5$;

(2)66人.

上期《同步分层能力测试题(一)》参考答案

基础练

一、选择题

1~6.DCBABD.

二、填空题

7.68;　8.1.7.

三、解答题

9.略.

10.$\hat{a}=4-1.5\hat{b}$.

11.(1)$\hat{y}=0.5x+0.4$.

(2)5.9万元.

12.(1)$\hat{y}=-1.15x+28.1$.

(2)$R^2\approx0.994$,说明拟合效果比较好.

提升练

一、选择题

1~6.BDBCDB.

二、填空题

7.丙;　8.-0.79.

三、解答题

9.解:(1)由数据,求得$\bar{x}=11$,$\bar{y}=24$.

由公式求得$\hat{b}=\frac{18}{7}$,$\hat{a}=\bar{y}-\hat{b}\bar{x}=-\frac{30}{7}$,

∴y关于x的线性回归方程为$\hat{y}=\frac{18}{7}x-\frac{30}{7}$.

(2)当$x=10$时,$y=\frac{150}{7}$,$|\frac{150}{7}-22|<2$;

当$x=6$时,$y=\frac{78}{7}$,$|\frac{78}{7}-12|<2$,

∴该小组所得线性回归方程是理想的.

10.解:(1)依题意,得$\frac{6}{10}a=50-26$,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)$\hat{y}=0.7x+25$.

(3)由(1)知,当$y>81$时认定驾驶员是"醉驾".令$y>81$,得$0.7x+25>81$,解得$x>80$.当每毫升血液酒精含量大于80毫克时认定为"醉驾".

同步分层 能力测试题(一)

(测试范围:独立性检验的基本思想及其初步应用)

基础练

时间:45分钟　满分:86分

素养名称	数学抽象	逻辑推理	数学运算	直观想象	数学建模	数据分析
题号	1,2,3,4		8			3,4,5,6,7,9,10,11,12

一、选择题(本大题共6小题,每小题5分,共30分)

1.以下关于独立性检验的说法中,正确的个数为(　　).

①独立性检验依赖小概率原理;

②独立性检验得到的结论一定正确;

③样本不同,独立性检验的结论可能有差异;

④独立性检验不是判定两事物是否相关的唯一方法.

A.1　B.2　C.3　D.4

2.下列关于等高条形图的说法正确的是(　　).

A.等高条形图表示高度相对的条形图

B.等高条形图表示的是分类变量的频数

C.等高条形图表示的是分类变量的频率

D.等高条形图表示的是分类变量的实际高度

3.(2019·福建高二期末)在一次独立性检验中,其把握性超过99%但不超过99.5%,则K^2的可能值为(　　).

参考数据:如表1的独立性检验临界值表:

表1

$P(K^2\geqslant k_0)$	0.100	0.050	0.025	0.010	0.005	0.001
k_0	2.706	3.841	5.024	6.635	7.879	10.828

A.5.424　B.6.765

C.7.897　D.11.897

4.在研究吸烟与患肺癌的关系中,通过收集数据、整理分析数据得到"吸烟与患肺癌有关"的结论,并且有99%以上的把握认为这个结论是成立的,则下列说法中正确的是(　　).

A.100个吸烟者中至少有99人患有肺癌

B.1个人吸烟,那么这人有99%的概率患有肺癌

C.在100个吸烟者中一定有患肺癌的人

D.在100个吸烟者中可能一个患肺癌的人也没有

5.分类变量X与Y的列联表如表2:

表2

	y_1	y_2	总计
x_1	a	b	$a+b$
x_2	c	d	$c+d$
总计	$a+c$	$b+d$	$a+b+c+d$

则以下判断正确的是(　　).

A.$ad-bc$越小,说明X与Y的关系越弱

B.$ad-bc$越大,说明X与Y的关系越强

C.$(ad-bc)^2$越大,说明X与Y的关系越强

D.$(ad-bc)^2$越接近于0,说明X与Y的关系越强

6.(2019·广西高二期末)某校对甲、乙两个文科班的数学考试成绩进行分析,规定:大于或等于120分为优秀,120分以下为非优秀.统计成绩后,得到如表3的2×2的列联表.根据列联表的数据判断有多少的把握认为"成绩与班级有关系"(　　).

表3

	优秀	非优秀	合计
甲班	10	50	60
乙班	20	30	50
合计	30	80	110

临界值表4:

表4

$P(K^2\geqslant k)$	0.100	0.050	0.025	0.010	0.001
k	2.706	3.841	5.024	6.635	10.828

A.90%　B.95%　C.99%　D.99.9%

二、填空题(本大题共2小题,每小题5分,共10分)

7.(2018·海林市海南朝鲜族乡中学高二课时练习)某医疗研究所为了检验某种血清对预防感冒的作用,把500名使用该血清的人与另外500名未使用血清的人一年中的感冒记录作比较,提出假设H_0:这种血清不能起到预防感冒的作用.利用2×2列联表计算得$K^2\approx3.918$,经查临界值表知$P(K^2\geqslant3.841)\approx0.05$.现给出下列三种说法:

①有95%的把握认为"这种血清能起到预防感冒的作用";

②如果某人未使用该血清,那么他在一年中有95%的可能性得感冒;

③这种血清预防感冒的有效率为5%.

则上述说法中,正确说法的序号是_____.

8.表5是关于出生男婴与女婴调查的列联表:

表5

	晚上	白天	总计
男婴	45	A	B
女婴	E	35	C
总计	98	D	180

那么$A=$_____,$B=$_____,$C=$_____,$D=$_____,$E=$_____.

三、解答题(本大题共4小题,共46分,解答应写出必要的文字说明、证明过程或演算步骤)

9.(10分)甲、乙两个班进行同一门课程的考试,按照学生成绩优秀和不优秀统计后得到表6:

表6

	优秀	不优秀	总计
甲班	10	35	45
乙班	7	38	45
总计	17	73	90

请判定学生成绩与班级是否有关?

10.(12分)运动员参加比赛前往往做热身运动,表7是一体育运动的研究机构对160位专业运动员追踪而得的数据.根据此表,判断运动员受伤与不做热身运动有关吗?

表7

	受伤	不受伤	总计
做热身	19	76	95
不做热身	45	20	65
总计	64	96	160

11.(12分)表8是某地区的一种传染病与饮用水的调查表:

表8

	得病	不得病	总计
干净水	52	466	518
不干净水	94	218	312
总计	146	684	830

(1)这种传染病是否与饮用水的卫生程度有关,请说明理由;

(2)若饮用干净水得病5人,不得病50人,饮用不干净水得病9人,不得病22人.按此样本数据分析这种病症是否与饮用水有关,并比较两种样本在反映总体时的差异.

12.(12分)(2018·河北高二期末(文))微信支付诞生于微信红包,早期只是作为社交的一部分"发红包"而诞生的,在发红包之余才发现,原来微信支付不仅可以用来发红包,还可以用来支付.现在微信支付被越来越多的人们所接受,现从某市市民中随机抽取300为对是否使用微信支付进行调查,得到如表9的2×2的列联表:

表9

	年轻人	非年轻人	总计
经常使用微信支付	165		225
不常使用微信支付			
合计		90	300

根据表中数据,我们得到的统计学的结论是:有多少的把握认为"使用微信支付与年龄有关"?

(下转第4版)