专题 67 成对数据的统计分析	
知识梳理	考纲要求
	考点预测
	常用结论
	方法技巧
题型归类	题型一：相关关系与散点图
	题型二：回归方程与最小二乘法
	题型三：相关系数
	题型四：误差分析
	题型五：非线性回归
	题型六：列联表与等高条形图
	题型七：独立性检验
培优训练	训练一：
	训练二：
	训练三：
	训练四：
	训练五：
	训练六：
强化测试	单选题：共 8 题
	多选题：共 4 题
	填空题：共 4 题
	解答题：共 6 题

一、【知识梳理】

【考纲要求】

1.了解样本相关系数的统计含义.

2.了解一元线性回归模型和 2×2 列联表，会运用这些方法解决简单的实际问题.

3.会利用统计软件进行数据分析.

【考点预测】

1.变量的相关关系

(1)相关关系

两个变量有关系，但又没有确切到可由其中的一个去精确地决定另一个的程度，这种关系称为相关关系.

(2)相关关系的分类：正相关和负相关.

(3)线性相关

一般地，如果两个变量的取值呈现正相关或负相关，而且散点落在一条直线附近，我们就称这两个变量线性相关.

一般地，如果两个变量具有相关性，但不是线性相关，那么我们就称这两个变量非线性相关或曲线相关.

2.样本相关系数

(1)相关系数 r 的计算

变量 x 和变量 y 的样本相关系数 r 的计算公式如下：

$$r=\frac{\sum\limits_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\sum\limits_{i=1}^{n}(x_i-\overline{x})^2}\sqrt{\sum\limits_{i=1}^{n}(y_i-\overline{y})^2}}.$$

(2)相关系数 r 的性质

①当 $r>0$ 时，称成对样本数据正相关；当 $r<0$ 时，成对样本数据负相关；当 $r=0$ 时，成对样本数据间没有线性相关关系.

②样本相关系数 r 的取值范围为[－1，1].

当|r|越接近 1 时，成对样本数据的线性相关程度越强；

当|r|越接近 0 时，成对样本数据的线性相关程度越弱.

3.一元线性回归模型

(1)经验回归方程与最小二乘法

我们将$\hat{y}=\hat{b}x+\hat{a}$称为 Y 关于 x 的经验回归方程，也称经验回归函数或经验回归公式，其图形称为经验回归直线.这种求经验回归方程的方法叫做最小二乘法，求得的$\hat{b}$，$\hat{a}$叫做 b，a 的最小二乘估计，

其中

$$\begin{cases}\hat{b}=\dfrac{\sum\limits_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sum\limits_{i=1}^{n}(x_i-\overline{x})^2}=\dfrac{\sum\limits_{i=1}^{n}x_iy_i-n\,\overline{x}\,\overline{y}}{\sum\limits_{i=1}^{n}x_i^2-n\,\overline{x}^2},\\ \hat{a}=\overline{y}-\hat{b}\,\overline{x}.\end{cases}$$

(2)利用决定系数 R^2 刻画回归效果

$R^2=1-\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\overline{y})^2}$，$R^2$越大，即拟合效果越好，$R^2$越小，模型拟合效果越差.

4.列联表与独立性检验

(1)2×2 列联表

一般地，假设有两个分类变量 X 和 Y，它们的取值分别为$\{x_1，x_2\}$和$\{y_1，y_2\}$，其 2×2 列联表为

x	y		合计
	$y=y_1$	$y=y_2$	
$x=x_1$	a	b	$a+b$
$x=x_2$	c	d	$c+d$
合计	$a+c$	$b+d$	$n=a+b+c+d$

(2)临界值

$\chi^2=\dfrac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$.忽略 χ^2 的实际分布与该近似分布的误差后，对于任何小概率值 α，可以找到相应的正实数 x_α，使得 $P(\chi^2\geqslant x_\alpha)=\alpha$ 成立.我们称 x_α 为 α 的临界值，这个临界值就可作为判断 x^2 大小的标准.

(3)独立性检验

基于小概率值 α 的检验规则是：

当 $\chi^2\geqslant x_\alpha$ 时，我们就推断 H_0 不成立，即认为 X 和 Y 不独立，该推断犯错误的概率不超过 α；

当 $\chi^2<x_\alpha$ 时，我们没有充分证据推断 H_0 不成立 ，可以认为 X 和 Y 独立.

这种利用 χ^2 的取值推断分类变量 X 和 Y 是否独立的方法称为 χ^2 独立性检验，读作“卡方独立性检验”，简称独立性检验.

下表给出了 χ^2 独立性检验中几个常用的小概率值和相应的临界值

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

【常用结论】

1.求解经验回归方程的关键是确定回归系数$\hat{a}，\hat{b}$，应充分利用回归直线过样本点的中心$(\bar{x}，\bar{y})$.

2.根据经验回归方程计算的$\hat{y}$值，仅是一个预报值，不是真实发生的值.

3.根据 χ^2 的值可以判断两个分类变量有关的可信程度，若 χ^2 越大，则两分类变量有关的把握越大.

【方法技巧】

1.判断相关关系的两种方法：

(1)散点图法：如果样本点的分布从整体上看大致在某一曲线附近，变量之间就有相关关系；如果样本点的分布从整体上看大致在某一直线附近，变量之间就有线性相关关系.

(2)决定系数法：利用决定系数判定，R^2 越趋近 1，拟合效果越好，相关性越强.

2. 经验回归方程

(1)求经验回归方程：利用公式$\hat{b}=\dfrac{\sum\limits_{i=1}^{n}x_iy_i-n\bar{x}\ \bar{y}}{\sum\limits_{i=1}^{n}x_i^2-n\bar{x}^2}$求$\hat{b}$；利用$\hat{a}=\bar{y}-\hat{b}\bar{x}$求$\hat{a}$，写出经验回归方程.

(2)经验回归方程的拟合效果，可以利用相关系数$|r|$判断，当$|r|$越趋近于 1 时，两变量的线性相关性越强.或利用决定系数 R^2 判断，R^2 越大，拟合效果越好.

(3)非线性经验回归方程转化为线性经验回归方程的方法

①若$\hat{y}=\hat{a}+\hat{b}\sqrt{x}$，设 $t=\sqrt{x}$，则$\hat{y}=\hat{a}+\hat{b}t$；②若满足对数式：$\hat{y}=\hat{a}+\hat{b}\ln x$，设 $t=\ln x$，则$\hat{y}=\hat{a}+\hat{b}t$；③若满足指数式：$y=c_1e^{c_2x}$，两边取对数解 $\ln y=\ln c_1+c_2x$，设 $z=\ln y$，$a=\ln c_1$，$b=c_2$，则 $z=a+bx$.

3.在 2×2 列联表中，如果两个变量没有关系，则应满足 $ad-bc\approx0$.$|ad-bc|$越小，说明两个变量之间关系越弱；$|ad-bc|$越大，说明两个变量之间关系越强.

4.解决独立性检验的应用问题，一定要按照独立性检验的步骤得出结论.独立性检验的一般步骤：

(1)根据样本数据制成 2×2 列联表：

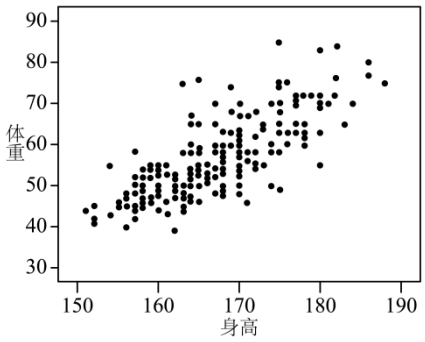
(2)根据公式 $\chi^2=\dfrac{n(ad-bc)^2}{(a+b)(a+c)(b+d)(c+d)}$计算 χ^2；

(3)通过比较 χ^2 与临界值的大小关系来作统计推断.

二、【题型归类】

【题型一】相关关系与散点图

【典例 1】（2023·上海·模拟预测）根据身高和体重散点图，下列说法正确的是（　）



A．身高越高，体重越重　　B．身高越高，体重越轻

C．身高与体重成正相关　　D．身高与体重成负相关

【解析】【答案】C

【分析】根据给定的散点图的特征，直接判断作答.

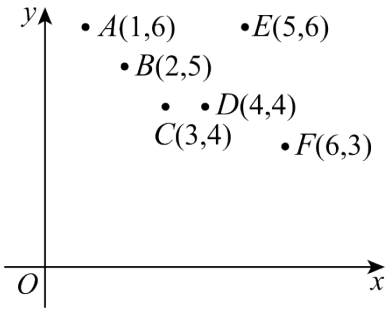
【详解】由于身高比较高的人，其体重可能大，也可能小，则选项 AB 不正确；

由散点图知，身高和体重有明显的相关性，且身高增加时，体重也呈现增加的趋势，

所以身高与体重呈正相关，C 正确，D 错误.

故选：C

【典例 2】（多选）（2023·湖南长沙·长沙市明德中学校考三模）6 个数据(x,y)构成的散点图，如图所示，采用一元线性回归模型建立经验回归方程，若在 6 个数据中去掉$E(5,6)$后，下列说法正确的是（　）



A．解释变量 x 与预报变量 y 的相关性变强B．样本相关系数 r 变大

C．残差平方和变小　　D．决定系数 R^2 变小

【解析】【答案】AC

【分析】根据散点图的性质可知去掉 E 后相关性变强判断 A 选项，相关系数为负判断 B 选项，残差平方和判断 C,D 选项.

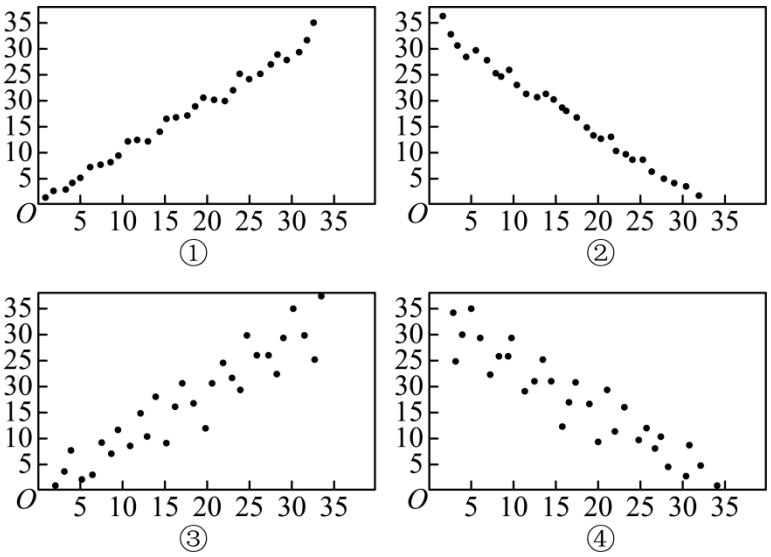
【详解】去掉$E(5,6)$后，变量 x 与预报变量 y 的相关性变强，故 A 正确；

但由于散点的分布是从左上到右下，故变量 x，y 负相关，所以相关系数 r 变小，

残差平方和变小， 决定系数 R^2 变大，C 正确，D 错误.

故选：AC.

【典例 3】（2023 下·四川成都·高二四川省成都市新都一中校联考期中）以下是标号分别为①、②、③、④的四幅散点图，它们的样本相关系数分别为 r_1,r_2,r_3,r_4，那么相关系数的大小关系为____（按由小到大的顺序排列）.



【解析】【答案】$r_2<r_4<r_3<r_1$

【分析】利用样本相关系数的性质即可判断出 r_1,r_2,r_3,r_4 的大小关系

【详解】根据散点图可知，图①③成正相关，图②④成负相关，

$\therefore r_1>0,r_2<0,r_3>0,r_4<0$，

又图①②的散点图近似在一条直线上，则图①②两变量的线性相关程度比较高，

图③④的散点图比较分散，故图③④两变量的线性相关程度比较低，

即 $|r_1|$ 与 $|r_2|$ 比较大，$|r_4|$ 与 $|r_3|$ 比较小，$\therefore r_2<r_4<r_3<r_1$，

故答案为：$r_2<r_4<r_3<r_1$.

【题型二】回归方程与最小二乘法

【典例 1】（2023·四川宜宾·统考一模）华为在过去几年面临了来自美国政府的封锁和限制，但华为并没有放弃，在自主研发和国内供应链的支持下，成功突破了封锁，实现了 5G 功能.某手机商城统计了最近 5 个月华为手机的实际销量，如下表所示：

时间x（月）	1	2	3	4	5
销售量y（万部）	0.5	0.8	1.0	1.2	1.5

若y与x线性相关，且线性回归方程为$\hat{y}=0.24x+\hat{a}$，则下列说法不正确的是（　）

A．样本中心点为$(3,1.0)$

B．由表中数据可知，变量y与x呈正相关

C．$\hat{a}=0.28$

D．预测$x=7$时华为手机销量约为1.86（万部）

【解析】【答案】D

【分析】根据表格中数据的变换趋势，平均数的计算公式，以及回归直线方程，逐项判定，即可求解.

【详解】由表格数据可以计算出$\overline{x}=\frac{1+2+3+4+5}{5}=3$，$\overline{y}=\frac{0.5+0.8+1.0+1.2+1.5}{5}=1.0$，

则样本中心点为$(3,1.0)$，即选项A说法正确；

从表格数据可得：y随着x的增加而增加，所以变量y与x正相关，即选项B说法正确；

将样本中心点$(3,1.0)$代入$\hat{y}=0.24x+\hat{a}$，可得$\hat{a}=0.28$，即选项C说法正确；

由C可知线性回归方程为$\hat{y}=0.24x+0.28$，

将$x=7$代入可得$\hat{y}=0.24\times7+0.28=1.96$，则选项D说法不正确.

故选：D.

【典例2】（多选）（2023·浙江绍兴·统考模拟预测）由变量x和变量y组成的10个成对样本数据$(x_1,y_1),(x_2,y_2),\cdots,(x_{10},y_{10})$得到的经验回归方程为$\hat{y}=2x-0.1$，设过点$(x_2,y_2),(x_9,y_9)$的直线方程为$y=mx+n$，记$\overline{x}=\frac{1}{10}\sum_{i=1}^{10}x_i,\overline{y}=\frac{1}{10}\sum_{i=1}^{10}y_i$，则（　）

A．变量x,y正相关

B．若$\overline{x}=1$，则$\overline{y}=1.9$

C．经验回归直线$\hat{y}=2x-0.1$至少经过$(x_i,y_i)(i=1,2,\cdots,10)$中的一个点

D．$\sum_{i=1}^{10}(y_i-2x_i+0.1)^2\le\sum_{i=1}^{10}(y_i-mx_i-n)^2$

【解析】【答案】ABD

【分析】根据回归直线的相关性质分别判断各个选项即可.

【详解】对于 A:回归方程一次项系数大于零是正相关，A 正确；

对于 B:$\overline{x}=1$ 代入回归直线可得 $\overline{y}=2\times1-0.1=1.9$，B 正确；

经验回归直线可以不经过任意一个点，C 错误；

根据回归直线的求法最小二乘法值，回归直线的残差平方和最小，D 正确.

故选：ABD.

【典例 3】（2023·陕西榆林·校考模拟预测）为帮助乡村脱贫，某勘探队计划了解当地矿脉某金属的分布情况，经勘测得到该金属含量 y（单位：$\mathrm{g/m^2}$）与样本对原点的距离 x（单位：m）的数据，并作了初步处理，得到下面的一些统计量的值.（表中 $u_i=\frac{1}{x_i},\overline{u}=\frac{1}{9}\sum_{i=1}^{9}u_i$）

$\overline{x}$	$\overline{y}$	$\overline{u}$	$\sum_{i=1}^{9}(x_i-\overline{x})^2$	$\sum_{i=1}^{9}(u_i-\overline{u})^2$	$\sum_{i=1}^{9}(y_i-\overline{y})^2$	$\sum_{i=1}^{9}(x_i-\overline{x})(y_i-\overline{y})$	$\sum_{i=1}^{9}(u_i-\overline{u})(y_i-\overline{y})$
6	97.90	0.21	60	0.14	14.12	26.13	-1.40

(1)利用样本相关系数的知识，判断 $y=a+bx$ 与 $y=c+\frac{d}{x}$ 哪一个更适宜作为该金属含量 y 关于样本对原点的距离 x 的回归方程类型？

(2)根据（1）的结果解决下列问题：

（i）建立 y 关于 x 的回归方程；

（ii）样本对原点的距离 $x=20$ 时，该金属含量的预报值是多少？

(3)已知该金属在距离原点 $x\mathrm{m}$ 时的平均开采成本 W（单位：元）与 x,y 的关系为 $W=100\left(y-\ln x\right)\left(1\leqslant x\leqslant 100\right)$，根据（2）的结论说明，$x$ 为何值时，开采成本最大？

附：线性回归方程 $\hat{y}=\hat{b}x+\hat{a}$ 的斜率和截距的最小二乘法公式分别为 $\hat{b}=\frac{\sum_{i=1}^{n}(x_i-\overline{x})(y-\overline{y})}{\sum_{i=1}^{n}(x_i-\overline{x})^2},\hat{a}=\overline{y}-\hat{b}\overline{x}$.

【解析】【答案】(1) $y=c+\frac{d}{x}$ 更适宜

(2)（i）$\hat{y}=100-\frac{10}{x}$；（ii）$99.5\mathrm{g/m^3}$；

(3) $x=10\mathrm{m}$ 时，开采成本最大.

【分析】（1）由题意，根据线性相关系数的计算公式计算出 r_1,r_2，即可下结论；

（2）利用最小二乘法求出回归方程，令 $x=20$，即可求解；

（3）由题意可得$W=1000\left(100-\dfrac{10}{x}-\ln x\right)$，结合导数讨论该函数的性质即可求解.

【详解】（1）$y=a+bx$的线性相关系数$r_1=\dfrac{\sum\limits_{i=1}^{9}(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\sum\limits_{i=1}^{9}(x_i-\overline{x})^2\sum\limits_{i=1}^{9}(y_i-\overline{y})^2}}=\dfrac{26.13}{\sqrt{60\times14.12}}\approx0.898$，

$y=c+\dfrac{d}{x}$的线性相关系数$r_2=\dfrac{\sum\limits_{i=1}^{9}(u_i-\overline{u})(y_i-\overline{y})}{\sqrt{\sum\limits_{i=1}^{9}(u_i-\overline{u})^2\sum\limits_{i=1}^{9}(y_i-\overline{y})^2}}=\dfrac{-1.40}{\sqrt{0.14\times14.12}}\approx-0.996$，

Q$|r_1|<|r_2|$

$\therefore y=c+\dfrac{d}{x}$更适宜作为该金属含量$y$关于样本对原点的距离$x$的回归方程类型.

（2）（i）依题意，可得$\hat{\beta}=\dfrac{\sum\limits_{i=1}^{9}(u_i-\overline{u})(y_i-\overline{y})}{\sum\limits_{i=1}^{9}(u_i-\overline{u})^2}=\dfrac{-1.40}{0.14}=-10$，

$\hat{\alpha}=\overline{y}-\hat{\beta}\overline{u}=97.9-(-10)\times0.21=100$，

$\therefore\hat{y}=100-10u=100-\dfrac{10}{x},\therefore y$关于$x$的回归方程为$\hat{y}=100-\dfrac{10}{x}$.

（ii）当$x=20$时，金属含量的预报值为$\hat{y}=100-\dfrac{10}{20}=99.5\text{g}/\text{m}^3$.

（3）Q$W=1000(y-\ln x)=1000\left(100-\dfrac{10}{x}-\ln x\right)$，

令$f(x)=100-\dfrac{10}{x}-\ln x$，则$f'(x)=\dfrac{10}{x^2}-\dfrac{1}{x}=\dfrac{10-x}{x^2}$，

当$1\le x<10$时，$f'(x)>0,f(x)$在$[1,10)$上单调递增；

当$10<x\le100$时，$f'(x)<0,f(x)$在$(10,100]$上单调递减，

\ $f(x)$在$x=10$处取得极大值，也是最大值，此时W取得最大值，

故$x=10\text{m}$时，开采成本最大.

【题型三】相关系数

【典例 1】（2023·全国·模拟预测）某厂近几年陆续购买了几台 A 型机床，该型机床已投入生产的时间 x（单位：年）与当年所需要支出的维修费用 y（单位：万元）有如下统计资料：

x	2	3	4	5	6
y	2.2	3.8	5.5	6.5	7.0

已知$\sum_{i=1}^{5}x_i^2=90$，$\sum_{i=1}^{5}y_i^2\approx140.8$，$\sum_{i=1}^{5}x_iy_i=112.3$，$\sqrt{79}\approx8.9$，$\sqrt{2}\approx1.4$.

(1)计算y与x的样本相关系数r（精确到0．001），并判断该型机床的使用年限与所支出的维修费用的相关性强弱（若$0.75\le|r|\le1$，则认为y与x相关性很强，否则不强）.

(2)该厂购入一台新的A型机床，工人们分别使用这台机床（记为X）和一台已经使用多年的A型机床（记为Y）各制造50个零件，统计得出的数据如下表：

机床	零件		合计
	合格	不合格	
X		4	
Y	40		
合计			

根据所给数据完成上表，试根据小概率值$\alpha=0.1$的独立性检验，分析零件合格情况是否与机床的使用情况有关.

附：相关系数$r=\dfrac{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)\left(y_i-\bar{y}\right)}{\sqrt{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2\sum_{i=1}^{n}\left(y_i-\bar{y}\right)^2}}=\dfrac{\sum_{i=1}^{n}x_iy_i-n\bar{x}\bar{y}}{\sqrt{\left(\sum_{i=1}^{n}x_i^2-n\bar{x}^2\right)\left(\sum_{i=1}^{n}y_i^2-n\bar{y}^2\right)}}$.

$\chi^2=\dfrac{n\left(ad-bc\right)^2}{\left(a+b\right)\left(c+d\right)\left(a+c\right)\left(b+d\right)}$，其中$a+b+c+d=n$.

α	0.10	0.05	0.010	0.005	0.001
x_α	2.706	3.841	6.635	7.879	10.828

【解析】【答案】(1)$r\approx0.987$，A型机床的使用年限与当年所支出的维修费用之间具有很强的相关性

(2)填表见解析；认为零件合格情况与机床的使用情况有关

【分析】（1）计算相关系数r，即可得到答案；

（2）根据题意完成表格，计算得到$\chi^2\approx2.990$，进而可判断结果.

【详解】（1）$\bar{x}=\dfrac{2+3+4+5+6}{5}=4,\bar{y}=\dfrac{2.2+3.8+5.5+6.5+7.0}{5}=5$，

$\sum_{i=1}^{5}x_iy_i-5\overline{xy}=12.3,\sum_{i=1}^{5}x_i^2-5\overline{x}^2=10,\sum_{i=1}^{5}y_i^2-5\overline{y}^2\approx15.8$，

所以$r=\dfrac{\sum_{i=1}^{5}x_iy_i-5\overline{xy}}{\sqrt{\left(\sum_{i=1}^{5}x_i^2-5\overline{x}^2\right)\left(\sum_{i=1}^{5}y_i^2-5\overline{y}^2\right)}}\approx\dfrac{12.3}{\sqrt{10\times15.8}}=\dfrac{12.3}{\sqrt{2}\times\sqrt{79}}\approx\dfrac{12.3}{1.4\times8.9}\approx0.987$，

r 接近 1，说明 A 型机床的使用年限与当年所支出的维修费用之间具有很强的相关性.

（2）补充 2×2 列联表如下：

机床	零件		合计
	合格	不合格	
X	46	4	50
Y	40	10	50
合计	86	14	100

零假设为 H_0：零件合格情况与机床的使用情况无关.

根据列联表中的数据，经计算得到 $\chi^2=\dfrac{100\times(46\times10-4\times40)^2}{50\times50\times86\times14}\approx2.990>2.706=x_{0.1}$，

所以根据小概率值 $\alpha=0.1$ 的独立性检验，

我们推断 H_0 不成立，即认为零件合格情况与机床的使用情况有关.

【典例 2】（2023·全国·模拟预测）直播带货是一种直播和电商相结合的销售手段，目前已被广大消费者所接受．针对这种现状，某公司决定逐月加大直播带货的投入，直播带货金额稳步提升，以下是该公司 2023 年前 5 个月的带货金额：

月份 x	1	2	3	4	5
带货金额 y /万元	350	440	580	700	880

(1)计算变量 x，y 的相关系数 r（结果精确到 0.01）.

(2)求变量 x，y 之间的线性回归方程，并据此预测 2023 年 7 月份该公司的直播带货金额.

(3)该公司随机抽取 55 人进行问卷调查，得到如下不完整的列联表：

	参加过直播带货	未参加过直播带货	总计
女性	25		30

男性		10	
总计			

请填写上表，并判断是否有 90%的把握认为参加直播带货与性别有关.

参考数据：$\overline{y}=590$，$\sum_{i=1}^{5}\left(x_i-\overline{x}\right)^2=10$，$\sum_{i=1}^{5}\left(y_i-\overline{y}\right)^2=176400$，

$\sum_{i=1}^{5}\left(x_i-\overline{x}\right)\left(y_i-\overline{y}\right)=1320$，$\sqrt{441000}\approx 664$.

参考公式：相关系数$r=\dfrac{\sum_{i=1}^{n}\left(x_i-\overline{x}\right)\left(y_i-\overline{y}\right)}{\sqrt{\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2}\sqrt{\sum_{i=1}^{n}\left(y_i-\overline{y}\right)^2}}$，线性回归方程的斜率$\hat{b}=\dfrac{\sum_{i=1}^{n}\left(x_i-\overline{x}\right)\left(y_i-\overline{y}\right)}{\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2}$，截距

$\hat{a}=\overline{y}-\hat{b}\overline{x}$.

附：$K^2=\dfrac{n\left(ad-bc\right)^2}{\left(a+b\right)\left(c+d\right)\left(a+c\right)\left(b+d\right)}$，其中$n=a+b+c+d$.

$P\left(K^2\geq k_0\right)$	0.15	0.10	0.05	0.025
k_0	2.072	2.706	3.841	5.024

【解析】【答案】(1)0.99

(2) $\hat{y}=132x+194$，1118 万元

(3)表格见解析，有

【分析】（1）直接代入求相关系数即可；

（2）根据线性回归方程求解回归方程即可；

（3）零假设之后计算K^2，再比较大小判断零假设是否成立即可.

【详解】（1）$r=\dfrac{\sum_{i=1}^{5}\left(x_i-\overline{x}\right)\left(y_i-\overline{y}\right)}{\sqrt{\sum_{i=1}^{5}\left(x_i-\overline{x}\right)^2}\sqrt{\sum_{i=1}^{5}\left(y_i-\overline{y}\right)^2}}=\dfrac{1320}{\sqrt{10}\times\sqrt{176400}}=\dfrac{1320}{2\times\sqrt{441000}}\approx 0.99$

（2）因为$\overline{x}=\dfrac{1}{5}\times\left(1+2+3+4+5\right)=3$，$\overline{y}=590$，$\sum_{i=1}^{5}\left(x_i-\overline{x}\right)^2=10$，$\sum_{i=1}^{5}\left(x_i-\overline{x}\right)\left(y_i-\overline{y}\right)=1320$，

所以$\hat{b}=\dfrac{\sum_{i=1}^{5}\left(x_i-\overline{x}\right)\left(y_i-\overline{y}\right)}{\sum_{i=1}^{5}\left(x_i-\overline{x}\right)^2}=\dfrac{1320}{10}=132$，$\hat{a}=590-132\times 3=194$，

所以变量x，y之间的线性回归方程为$\hat{y}=132x+194$，

当$x=7$时，$\hat{y}=132\times7+194=1118$（万元）.

所以预测2023年7月份该公司的直播带货金额为1118万元.

（3）补全完整的列联表如下.

	参加过直播带货	未参加过直播带货	总计
女性	25	5	30
男性	15	10	25
总计	40	15	55

零假设H_0：参加直播带货与性别无关，

根据以上数据，经计算得到$K^2=\dfrac{55\times(25\times10-5\times15)^2}{30\times25\times40\times15}\approx3.743>2.706=x_{0.1}$，

根据小概率值$\alpha=0.1$的独立性检验我们推断H_0不成立，即参加直播带货与性别有关，该判断犯错误的概率不超过10%.

【典例3】（2023·四川内江·统考一模）某企业为响应国家号召，汇聚科研力量，加强科技创新，准备加大研发资金投入，为了解年研发资金投入额x（单位：亿元）对年盈利额y（单位：亿元）的影响，通过对“十二五”和“十三五”规划发展10年期间年研发资金投入额x_i和年盈利额$y_i(i=1,2,\cdots,10)$数据进行分析，建立了两个函数模型：$y=\alpha+\beta x^2$；$y=\mathrm{e}^{\lambda x+t}$，其中$\alpha$、$\beta$、$\lambda$、$t$均为常数，$\mathrm{e}$为自然对数的底数，令$u_i=x_i^2$，$v_i=\ln y_i(i=1,2,\cdots,10)$，经计算得如下数据：

$\bar{x}=26$	$\bar{y}=215$	$\bar{u}=680$	$\bar{v}=5.36$
$\sum\limits_{i=1}^{10}(x_i-\bar{x})^2=100$	$\sum\limits_{i=1}^{10}(u_i-\bar{u})^2=22500$	$\sum\limits_{i=1}^{10}(u_i-\bar{u})(y_i-\bar{y})=260$	$\sum\limits_{i=1}^{10}(y_i-\bar{y})^2=4$
$\sum\limits_{i=1}^{10}(v_i-\bar{v})^2=4$	$\sum\limits_{i=1}^{10}(x_i-\bar{x})(v_i-\bar{v})=18$		

(1)请从相关系数的角度，分析哪一个模型拟合度更好？

(2)根据（1）的选择及表中数据，建立y关于x的回归方程.（系数精确到0.01）

附：相关系数$r=\dfrac{\sum\limits_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum\limits_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum\limits_{i=1}^{n}(y_i-\bar{y})^2}}$

回归直线 $\hat{y}=\hat{b}x+\hat{a}$ 中：$\hat{b}=\dfrac{\sum\limits_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sum\limits_{i=1}^{n}(x_i-\overline{x})^2}$，$\hat{a}=\overline{y}-\hat{b}\overline{x}$.

【解析】【答案】(1)模型 $y=\mathrm{e}^{\lambda x+t}$ 拟合度更好

(2) $y=\mathrm{e}^{0.18x+0.68}$

【分析】（1）计算出两个模型的相关系数，判断即可；

（2）根据最小二乘法计算即可.

【详解】（1）设模型 $y=\alpha+\beta x^2$ 的相关系数为 r_1，模型 $y=\mathrm{e}^{\lambda x+t}$ 的相关系数为 r_2，

对于模型 $y=\alpha+\beta x^2$，令 $u=x^2$，即 $y=\alpha+\beta u$，

所以 $r_1=\dfrac{\sum\limits_{10}^{i=1}(u_i-\overline{u})(y_i-\overline{y})}{\sqrt{\sum\limits_{10}^{i=1}(u_i-\overline{u})^2}\sqrt{\sum\limits_{10}^{i=1}(y_i-\overline{y})^2}}=\dfrac{260}{\sqrt{22500}\sqrt{4}}\approx 0.87$，

对于模型 $y=\mathrm{e}^{\lambda x+t}$，有 $\ln y=\ln \mathrm{e}^{\lambda x+t}=\lambda x+t$，令 $v=\ln y$，即 $v=\lambda x+t$，

所以 $r_2=\dfrac{\sum\limits_{10}^{i=1}(x_i-\overline{x})(v_i-\overline{v})}{\sqrt{\sum\limits_{10}^{i=1}(x_i-\overline{x})^2}\sqrt{\sum\limits_{10}^{i=1}(v_i-\overline{v})^2}}=\dfrac{18}{\sqrt{100}\sqrt{4}}=0.9$，

因为 $r_1<r_2$，所以模型 $y=\mathrm{e}^{\lambda x+t}$ 拟合度更好.

（2）因为 $\hat{\lambda}=\dfrac{\sum\limits_{10}^{i=1}(x_i-\overline{x})(v_i-\overline{v})}{\sum\limits_{10}^{i=1}(x_i-\overline{x})^2}=\dfrac{18}{100}=0.18$，$\hat{t}=\overline{v}-\hat{\lambda}\overline{x}=5.36-0.18\times 26=0.68$，

所以 y 关于 x 的回归方程为 $y=\mathrm{e}^{0.18x+0.68}$.

【点睛】本题考查回归方程的求解，其中第二问中，需要对 $y=\mathrm{e}^{\lambda x+t}$ 取对数得 $\ln y=\lambda x+t$，求得 v 关于 x 的线性回归方程，再转化为 y 关于 x 的回归方程，是处理本题的难点和关键点.

【题型四】误差分析

【典例 1】（2023·山东潍坊·统考模拟预测）某地区未成年男性的身高 x（单位：cm）与体重平均值 y（单位：kg）的关系如下表 1：

表 1 未成年男性的身高与体重平均值

身高/cm	60	70	80	90	100	110	120	130	140	150	160	170

体重平均值/kg	6.13	7.90	9.99	12.15	15.02	17.50	20.92	26.86	31.11	38.85	47.25	55.05

直观分析数据的变化规律，可选择指数函数模型、二次函数模型、幂函数模型近似地描述未成年男性的身高与体重平均值之间的关系．为使函数拟合度更好，引入拟合函数和实际数据之间的误差平方和、拟合优度判断系数 R^2（如表 2）．误差平方和越小、拟合优度判断系数 R^2 越接近 1，拟合度越高．

表 2 拟合函数对比

函数模型	函数解析式	误差平方和	R^2
指数函数	$y=2.004\mathrm{e}^{0.0197x}$	6.6764	0.9976
二次函数	$y=0.0037x^2-0.431x+19.6973$	8.2605	0.9971
幂函数	$y=0.001x^{2.1029}$	74.6846	0.9736

(1)问哪种模型是最优模型？并说明理由；

(2)若根据生物学知识，人体细胞是人体结构和生理功能的基本单位，是生长发育的基础．假设身高与骨细胞数量成正比，比例系数为 k_1；体重与肌肉细胞数量成正比，比例系数为 k_2．记时刻 t 的未成年时期骨细胞数量 $G(t)=G_0\mathrm{e}^{r_1t}$，其中 G_0 和 r_1 分别表示人体出生时骨细胞数量和增长率，记时刻 t 的未成年时期肌肉细胞数量 $J(t)=J_0\mathrm{e}^{r_2t}$，其中 J_0 和 r_2 分别表示人体出生时肌肉细胞数量和增长率．求体重 y 关于身高 x 的函数模型；

(3)在（2）的条件下，若 $k_2J_0\left(\dfrac{1}{k_1G_0}\right)^{\frac{r_2}{r_1}}=0.001$，$\dfrac{r_2}{r_1}=2.1029$．当刚出生的婴儿身高为 50cm 时，与（1）的模型相比较，哪种模型跟实际情况更符合，试说明理由．

注：$\mathrm{e}^{0.985}\approx2.67781$，$50^{2.1029}\approx3739.07$；婴儿体重 $y\in[2.5,4)$ 符合实际，婴儿体重 $y\in[4,5)$ 较符合实际，婴儿体重 $y\in[5,6)$ 不符合实际．

【解析】【答案】(1)指数函数模型是最优模型；理由见解析

(2) $y=k_2J_0\left(\dfrac{1}{k_1G_0}\right)^{\frac{r_2}{r_1}}x^{\frac{r_2}{r_1}}$

(3)（2）中幂函数模型更适合，理由见解析

【分析】（1）由表中数据比较指数函数模型误差平方和以及 R^2 的大小，即得结论；

（2）根据身高与骨细胞数量以及体重与肌肉细胞数量的关系，结合已知数据，即可求得答案；

（3）分别计算出两种模型函数下的婴儿体重，比较大小，即得结论.

【详解】（1）因为$6.6764<8.2605<74.6846$，所以指数函数模型误差平方和最小，

因为$0.9736<0.9971<0.9976$，所以指数函数模型R^2最大，

所以指数函数模型是最优模型；

（2）因为$x(t)=k_1G(t)=k_1G_0\mathrm{e}^{r_1t}$，所以$\mathrm{e}^{r_1t}=\dfrac{x}{k_1G_0}$，

因为$y(t)=k_2J(t)=k_2J_0\mathrm{e}^{r_2t}$，

所以$\mathrm{e}^{r_2t}=\dfrac{y}{k_2J_0}$，所以$\left(\dfrac{x}{k_1G_0}\right)^{r_2}=\mathrm{e}^{r_1r_2t}=\left(\dfrac{y}{k_2J_0}\right)^{r_1}$，

所以体重y关于身高x的函数模型为$y=k_2J_0\left(\dfrac{1}{k_1G_0}\right)^{\frac{r_2}{r_1}}x^{\frac{r_2}{r_1}}$；

（3）把$x=50\mathrm{cm}$代入$y=2.004\mathrm{e}^{0.0197x}$，得$y\approx2.004\times2.67781\approx5.3663(\mathrm{kg})\in[5,6)$不符合实际，

把$k_2J_0\left(\dfrac{1}{k_1G_0}\right)^{\frac{r_2}{r_1}}=0.001$，$\dfrac{r_2}{r_1}=2.1029$代入$y=k_2J_0\left(\dfrac{1}{k_1G_0}\right)^{\frac{r_2}{r_1}}x^{\frac{r_2}{r_1}}$得$y=0.001x^{2.1029}$，

把$x=50\mathrm{cm}$代入$y=0.001x^{2.1029}$，得$y\approx0.001\times3739.07\approx3.7391(\mathrm{kg})\in[2.5,4)$符合实际，

所以（2）中幂函数模型更适合.

【典例 2】（2023 上·山西朔州·高三校考开学考试）某校 20 名学生的数学成绩$x_i\left(i=1,2,\cdots,20\right)$和知识竞赛成绩$y_i\left(i=1,2,\cdots,20\right)$如下表：

学生编号 i	1	2	3	4	5	6	7	8	9	10
数学成绩 x_i	100	99	96	93	90	88	85	83	80	77
知识竞赛成绩 y_i	290	160	220	200	65	70	90	100	60	270
学生编号 i	11	12	13	14	15	16	17	18	19	20
数学成绩 x_i	75	74	72	70	68	66	60	50	39	35
知识竞赛成绩 y_i	45	35	40	50	25	30	20	15	10	5

计算可得数学成绩的平均值是$\overline{x}=75$，知识竞赛成绩的平均值是$\overline{y}=90$，并且$\sum_{i=1}^{20}\left(x_i-\overline{x}\right)^2=6464$，

$\sum_{i=1}^{20}\left(y_i-\overline{y}\right)^2=149450$，$\sum_{i=1}^{20}\left(x_i-\overline{x}\right)\left(y_i-\overline{y}\right)=21650$.

(1)求这组学生的数学成绩和知识竞赛成绩的样本相关系数（精确到0.01）；

(2)设$N\in\mathrm{N}^*$，变量x和变量y的一组样本数据为$\left\{\left(x_i,y_i\right)|i=1,2,\cdots,N\right\}$，其中$x_i\left(i=1,2,\cdots,N\right)$两两不相同，$y_i\left(i=1,2,\cdots,N\right)$两两不相同.记$x_i$在$\left\{x_n|n=1,2,\cdots,N\right\}$中的排名是第$R_i$位，$y_i$在$\left\{y_n|n=1,2,\cdots,N\right\}$中的排名是第$S_i$位，$i=1,2,\cdots,N$.定义变量$x$和变量$y$的"斯皮尔曼相关系数"（记为$\rho$）为变量$x$的排名和变量$y$的排名的样本相关系数.

（i）记$d_i=R_i-S_i$，$i=1,2,\cdots,N$.证明：$\rho=1-\frac{6}{N\left(N^2-1\right)}\sum_{i=1}^{N}d_i^2$；

（ii）用（i）的公式求得这组学生的数学成绩和知识竞赛成绩的"斯皮尔曼相关系数"约为0.91，简述"斯皮尔曼相关系数"在分析线性相关性时的优势.

注：参考公式与参考数据.

$r=\frac{\sum_{i=1}^{n}\left(x_i-\overline{x}\right)\left(y_i-\overline{y}\right)}{\sqrt{\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2\sum_{i=1}^{n}\left(y_i-\overline{y}\right)^2}}$；$\sum_{k=1}^{n}k^2=\frac{n(n+1)(2n+1)}{6}$；$\sqrt{6464\times149450}\approx31000$.

【解析】【答案】(1)证明见解析

(2)答案见解析

【分析】（1）利用相关系数的公式进行计算即可；

（2）（i）根据题意即相关系数的公式进行计算即可证明；（ii）只要能说出斯皮尔曼相关系数与一般的样本相关系数相比的优势即可.

【详解】（1）由题意，这组学生数学成绩和知识竞赛成绩的样本相关系数为

$r=\frac{\sum_{i}^{20}\left(x_i-\overline{x}\right)\left(y_i-\overline{y}\right)}{\sqrt{\sum_{i}^{20}\left(x_i-\overline{x}\right)^2\sum_{i}^{20}\left(y_i-\overline{y}\right)^2}}=\frac{21650}{\sqrt{6464\times149450}}\approx\frac{21650}{31000}\approx0.70$；

（2）（i）证明：因为$\left\{R_i\right\}$和$\left\{S_i\right\}$都是1，2，$\llcorner$，N的一个排列，所以

$\sum_{i=1}^{N}R_i=\sum_{i=1}^{N}S_i=\frac{N(N+1)}{2}$，

$\sum_{i=1}^{N}R_i^2=\sum_{i=1}^{N}S_i^2=\frac{N(N+1)(2N+1)}{6}$，

从而$\{R_i\}$和$\{S_i\}$的平均数都是$\overline{R}=\overline{S}=\frac{N+1}{2}$.

因此，$\sum_{i=1}^{N}\left(R_i-\overline{R}\right)^2=\sum_{i=1}^{N}R_i^2-2\overline{R}\sum_{i=1}^{N}R_i+\sum_{i=1}^{N}\overline{R}^2=\sum_{i=1}^{N}R_i^2-N\overline{R}^2=\frac{N(N+1)(2N+1)}{6}-\frac{N(N+1)^2}{4}=\frac{N(N+1)(N-1)}{12}$，

同理可得$\sum_{i=1}^{N}\left(S_i-\overline{S}\right)^2=\frac{N(N+1)(N-1)}{12}$，

由于$\sum_{i=1}^{N}d_i^2=\sum_{i=1}^{N}\left(R_i-S_i\right)^2=\sum_{i=1}^{N}\left[\left(R_i-\overline{R}\right)-\left(S_i-\overline{S}\right)\right]^2=\sum_{i=1}^{N}\left(R_i-\overline{R}\right)^2+\sum_{i=1}^{N}\left(S_i-\overline{S}\right)^2-2$

$\sum_{i=1}^{N}\left(R_i-\overline{R}\right)\left(S_i-\overline{S}\right)=2\cdot\frac{N(N+1)(N-1)}{12}-2\sum_{i=1}^{N}\left(R_i-\overline{R}\right)\left(S_i-\overline{S}\right)$，

所以$\rho=\frac{\sum_{i=1}^{N}\left(R_i-\overline{R}\right)\left(S_i-\overline{S}\right)}{\sqrt{\sum_{i=1}^{N}\left(R_i-\overline{R}\right)^2}\sqrt{\sum_{i=1}^{N}\left(S_i-\overline{S}\right)^2}}=\frac{\frac{N(N+1)(N-1)}{12}-\frac{1}{2}\sum_{i=1}^{N}d_i^2}{\frac{N(N+1)(N-1)}{12}}=1-\frac{6}{N\left(N^2-1\right)}\sum_{i=1}^{N}d_i^2$.

（ii）这组学生的数学成绩和知识竞赛成绩的斯皮尔曼相关系数是0.91，

答案①：斯皮尔曼相关系数对于异常值不太敏感，如果数据中有明显的异常值，那么用斯皮尔曼相关系数比用样本相关系数更能刻画某种线性关系；

答案②：斯皮尔曼相关系数刻画的是样本数据排名的样本相关系数，与具体的数值无关，只与排名有关．如果一组数据有异常值，但排名依然符合一定的线性关系，则可以采用斯皮尔曼相关系数刻画线性关系．

【点睛】方法点睛：新定义题型的特点是：通过给出一个新概念，或约定一种新运算，或给出几个新模型来创设全新的问题情景，要求考生在阅读理解的基础上，依据题目提供的信息，联系所学的知识和方法，实现信息的迁移，达到灵活解题的目的；遇到新定义问题，应耐心读题，分析新定义的特点，弄清新定义的性质，按新定义的要求，“照章办事”，逐条分析、验证、运算，使问题得以解决.

【典例3】（2023·山东泰安·统考模拟预测）近年来，我国新能源汽车发展进入新阶段．某品牌2018年到2022年新能源汽车年销量w（万）如下表：其中年对应的年份代码t为$1-5$．

年份代码t	1	2	3	4	5
销量w（万）	4	9	14	18	25

(1)判断两个变量是否线性相关，并计算样本相关系数（精确到0.01）；

(2)（i）假设变量x与变量Y的n对观测数据为$\left(x_1,y_1\right),\left(x_2,y_2\right)\cdots\left(x_n,y_n\right)$，两个变量满足一元线性回归模型

$\begin{cases}Y=bx+e\\E(e)=0,D(e)=\sigma^2\end{cases}$（随机误差$e_i=y_i-bx_i$），请写出参数$b$的最小二乘估计；

（ii）令变量 $x=t-\bar{t}, y=w-\bar{w}$，则变量 x 与变量 Y 满足一元线性回归模型 $\begin{cases}Y=bx+e\\E(e)=0, D(e)=\sigma^2\end{cases}$，利用（i）中结论求 y 关于 x 的经验回归方程，并预测2025年该品牌新能源汽车的销售量.

附：样本相关系数 $r=\dfrac{\sum\limits_{i=1}^{n}(t_i-\bar{t})(w_i-\bar{w})}{\sqrt{\sum\limits_{i=1}^{n}(t_i-\bar{t})^2}\sqrt{\sum\limits_{i=1}^{n}(w_i-\bar{w})^2}}$，$\sum\limits_{i=1}^{5}(t_i-\bar{t})(w_i-\bar{w})=47$，$\sum\limits_{i=1}^{5}(w_i-\bar{w})^2=262$，$\sum\limits_{i=1}^{5}(t_i-\bar{t})^2=10$，$\sqrt{655}\approx 25.6$

【解析】【答案】(1)线性相关且正相关，0.92

(2)（i）$\hat{b}=\dfrac{\sum\limits_{i=1}^{n}x_iy_i}{\sum\limits_{i=1}^{n}x_i^2}$；（ii）$y=4.7x$，37.5万辆.

【分析】（1）首先画出散点图，根据参考数据，计算 r；

（2）（ i ）首先写出残差平方和公式，表示为关于 b 的二次函数，列式求解；（ ii ）根据参考公式求回归直线方程，即可求解.

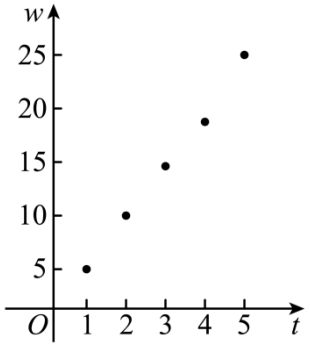
【详解】（1）通过做散点图发现，

样本点大致分布在一条直线附近，

因此是线性相关.

$$r=\frac{\sum\limits_{i=1}^{5}(t_i-\bar{t})(w_i-\bar{w})}{\sqrt{\sum\limits_{i=1}^{5}(t_i-\bar{t})^2}\sqrt{\sum\limits_{i=1}^{5}(w_i-\bar{w})^2}}=\frac{47}{\sqrt{262}\sqrt{10}}\approx\frac{47}{51.2}\approx 0.92$$

所以两变量有较强的正相关



（2）（i）$Q=\sum\limits_{i=1}^{n}e_i^2=\sum\limits_{i=1}^{n}(y_i-bx_i)^2=\sum\limits_{i=1}^{n}(y_i^2-2bx_iy_i+b^2x_i^2)=b^2\sum\limits_{i=1}^{n}x_i^2-2b\sum\limits_{i=1}^{n}x_iy_i+\sum\limits_{i=1}^{n}y_i^2$

要使残差平方和最小，当且仅当 $\hat{b}=\dfrac{\sum\limits_{i=1}^{n}x_iy_i}{\sum\limits_{i=1}^{n}x_i^2}$；

（ii）$\because x=t-\overline{t},y=w-\overline{w}$，

由（i）知 $\hat{b}=\dfrac{\sum\limits_{i=1}^{5}x_iy_i}{\sum\limits_{i=1}^{5}x_i^2}=\dfrac{\sum\limits_{i=1}^{5}(t_i-\overline{t})(w_i-\overline{w})}{\sum\limits_{i=1}^{5}(t_i-\overline{t})^2}=\dfrac{47}{10}\approx 4.7$，

$\therefore y$ 关于 x 的经验回归方程为 $y=4.7x$，

$\therefore w-\overline{w}=4.7(t-\overline{t})\because \overline{t}=3,\overline{w}=14,$

$\therefore w=4.7(t-\overline{t})+\overline{w}=4.7t-0.1,$

当 $t=8,w=4.7\times 8-0.1=37.5$（万），

因此，预计2025年该品牌新能源汽车的销售量将达到37.5万辆.

【题型五】非线性回归

【典例1】（2023·全国·模拟预测）一座城市的夜间经济不仅有助于拉动本地居民内需，还能延长外地游客、商务办公者等的留存时间，带动当地经济发展，是衡量一座城市生活质量、消费水平、投资环境及文化发展活力的重要指标．数据显示，近年来中国各地政府对夜间经济的扶持力度加大，夜间经济的市场发展规模保持稳定增长，下表为2017—2022年中国夜间经济的市场发展规模（单位：万亿元），其中2017—2022年对应的年份代码依次为1~6.

年份代码 x	1	2	3	4	5	6
中国夜间经济的市场发展规模 y / 万亿元	20.5	22.9	26.4	30.9	36.4	42.4

(1)已知可用函数模型 $y=a\cdot b^x$ 拟合 y 与 x 的关系，请建立 y 关于 x 的回归方程（a,b 的值精确到0.01）；

(2)某传媒公司预测2023年中国夜间经济的市场规模将达到48.1万亿元，现用（1）中求得的回归方程预测2023年中国夜间经济的市场规模，若两个预测规模误差不超过1万亿元，则认为（1）中求得的回归方程是理想的，否则是不理想的，判断（1）中求得的回归方程是否理想．参考数据：

$\overline{v}$	$\sum\limits_{i=1}^{6}x_iv_i$	$e^{2.848}$	$e^{0.148}$	1.16^7
3.366	73.282	17.25	1.16	2.83

其中 $v_i=\ln y_i$.

参考公式：对于一组数据 $(u_1,v_1),(u_2,v_2),\cdots,(u_n,v_n)$，其回归直线 $\hat{v}=\hat{\alpha}+\hat{\beta}u$

的斜率和截距的最小二乘估计分别为 $\hat{\beta}=\dfrac{\sum_{i=1}^{n}(u_i-\overline{u})(v_i-\overline{v})}{\sum_{i=1}^{n}(u_i-\overline{u})^2}$，$\hat{\alpha}=\overline{v}-\hat{\beta}\overline{u}$．

【解析】【答案】(1) $\hat{y}=17.25\times1.16^{x}$；

(2)是理想的

【分析】（1）通过对所给的的函数模型取对数，转换为求回归直线方程即可，再结合题中所给的直线方程与数据即可得解.

（2）利用（1）中求得的函数模型进行预测，结合回归方程理想的定义判断即可.

【详解】（1）将 $y=a\cdot b^{x}$ 的等号左右两边同时取自然对数得 $\ln y=\ln\left(a\cdot b^{x}\right)=\ln a+x\ln b$，

所以 $v=\ln a+x\ln b$．$\overline{x}=\dfrac{1+2+3+4+5+6}{6}=3.5$，

而 $\sum_{i=1}^{6}x_i^2=1^2+2^2+3^2+4^2+5^2+6^2=91$，

所以 $\hat{b}=\dfrac{\sum_{i=1}^{6}(x_i-\overline{x})(v_i-\overline{v})}{\sum_{i=1}^{6}(x_i-\overline{x})^2}=\dfrac{\sum_{i=1}^{6}x_iv_i-6\overline{x}\overline{v}}{\sum_{i=1}^{6}x_i^2-6\overline{x}^2}=\dfrac{73.282-6\times3.5\times3.366}{91-6\times3.5^2}=\dfrac{2.596}{17.5}\approx0.148$，

$\ln\hat{a}\approx3.366-0.148\times3.5=2.848$．

所以 $\hat{v}=2.848+0.148x$，即 $\ln\hat{y}=2.848+0.148x$，

所以 $\hat{y}=e^{2.848+0.148x}=17.25\times1.16^{x}$．

（2）2023 年对应的年份代码为 7，

当 $x=7$ 时，$\hat{y}=17.25\times1.16^{7}=17.25\times2.83\approx48.82$，$48.82-48.1=0.72<1$，

所以（1）中求得的回归方程 $\hat{y}=17.25\times1.16^{x}$ 是理想的.

【典例 2】（2023·全国·模拟预测）近三年的新冠肺炎疫情对我们的生活产生了很大的影响，当然也影响着我们的旅游习惯，乡村游、近郊游、周边游热闹了许多，甚至出现“微度假”的概念．在国家有条不紊的防疫政策下，旅游又重新回到了老百姓的日常生活中．某乡村抓住机遇，依托良好的生态环境、厚重的民族文化，开展乡村旅游．通过文旅度假项目考察，该村推出了多款套票文旅产品，得到消费者的积极回应．该村推出了六条乡村旅游经典线路，对应六款不同价位的旅游套票，相应的价格 x 与购买人数 y 的数据如下表.

旅游线路	奇山秀水游	古村落游	慢生活游	亲子游	采摘游	舌尖之旅
套票型号	A	B	C	D	E	F
价格 x/元	39	49	58	67	77	86

经数据分析、描点绘图，发现价格 x 与购买人数 y 近似满足关系式 $y=ax^b\left(a>0,b>0\right)$，即 $\ln y=b\ln x+\ln a\left(a>0,b>0\right)$，对上述数据进行初步处理，其中 $v_i=\ln x_i$，$w_i=\ln y_i$，$i=1$，2，…，6.

附：①可能用到的数据：$\sum\limits_{i=1}^{6}v_iw_i=75.3$，$\sum\limits_{i=1}^{6}v_i=24.6$，$\sum\limits_{i=1}^{6}w_i=18.3$，$\sum\limits_{i=1}^{6}v_i^2=101.4$.

②对于一组数据 $\left(v_1,w_2\right)$，$\left(v_2,w_2\right)$，…，$\left(v_n,w_n\right)$，其回归直线 $\hat{w}=\hat{b}v+\hat{a}$ 的斜率和截距的最小二乘估计值分别为 $\hat{b}=\dfrac{\sum\limits_{i=1}^{n}\left(v_i-\overline{v}\right)\left(w_i-\overline{w}\right)}{\sum\limits_{i=1}^{n}\left(v_i-\overline{v}\right)^2}=\dfrac{\sum\limits_{i=1}^{n}v_iw_i-n\overline{vw}}{\sum\limits_{i=1}^{n}v_i^2-n\overline{v}^2}$，$\hat{a}=\overline{w}-\hat{b}\overline{v}$.

(1)根据所给数据，求 y 关于 x 的回归方程.

(2)按照相关部门的指标测定，当套票价格 $x\in\left[49,81\right]$ 时，该套票受消费者的欢迎程度更高，可以被认定为“热门套票”．现有三位游客，每人从以上六款套票中购买一款旅游，购买任意一款的可能性相等．若三人买的套票各不相同，记三人中购买“热门套票”的人数为 X，求随机变量 X 的分布列和期望.

【解析】【答案】(1) $y=\mathrm{e}x^{\frac{1}{2}}$；

(2)分布列见解析，$E\left(X\right)=2$.

【分析】（1）将回归方程线性化处理，应用最小二乘法求线性方程，再由已知关系求回归方程；

（2）由题意确定 X 的可能取值，并求出对应概率，进而写出分布列，即可求期望.

【详解】（1）散点 $\left(v_i,w_i\right)=\left(i=1,2,\cdots,6\right)$ 集中在一条直线附近，

设回归直线方程为 $\hat{w}=\hat{b}v+\hat{a}$，$\overline{v}=\dfrac{1}{6}\sum\limits_{i=1}^{6}v_i=4.1$，$\overline{w}=\dfrac{1}{6}\sum\limits_{i=1}^{6}w_i=3.05$，

则 $\hat{b}=\dfrac{\sum\limits_{i=1}^{6}v_iw_i-6\overline{vw}}{\sum\limits_{i=1}^{6}v_i^2-6\overline{v}^2}=\dfrac{75.3-6\times4.1\times3.05}{101.4-6\times4.1^2}=\dfrac{1}{2}$，$\hat{a}=\overline{w}-\hat{b}\overline{v}=3.05-\dfrac{1}{2}\times4.1=1$，

所以回归直线方程为 $w=\dfrac{1}{2}v+1$.

因为 $v_i=\ln x_i$，$w_i=\ln y_i$，所以 $\ln y=\frac{1}{2}\ln x+1$，则 $b=\frac{1}{2}$，$\ln a=1$，所以 $y=\mathrm{e}x^{\frac{1}{2}}$.

综上，y 关于 x 的回归方程为 $y=\mathrm{e}x^{\frac{1}{2}}$.

（2）由题意知 B，C，D，E 为“热门套票”，则三人中购买“热门套票”的人数 X 服从超几何分布，

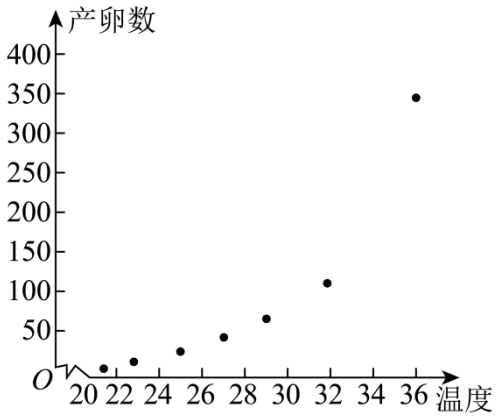
X 的可能取值为 1，2，3，且 $P(X=1)=\frac{C_4^1C_2^2}{C_6^3}=\frac{1}{5}$，$P(X=2)=\frac{C_4^2C_2^1}{C_6^3}=\frac{3}{5}$，$P(X=3)=\frac{C_4^3}{C_6^3}=\frac{1}{5}$.

X 的分布列如下.

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

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

【典例 3】（2022 上·广东深圳·高三校联考期中）红蜘蛛是柚子的主要害虫之一，能对柚子树造成严重伤害，每只红蜘蛛的平均产卵数 y（个）和平均温度 x（℃）有关，现收集了以往某地的 7 组数据，得到下面的散点图及一些统计量的值.



(1)根据散点图判断，$y=bx+a$ 与 $y=c\mathrm{e}^{dx}$（其中 $\mathrm{e}=2.718\ldots$ 为自然对数的底数）哪一个更适合作为平均产卵数 y（个）关于平均温度 x（℃）的回归方程类型？（给出判断即可，不必说明理由）

(2)由（1）的判断结果及表中数据，求出 y 关于 x 的回归方程.（计算结果精确到 0.1）

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

参考数据（$z=\ln y$）					
$\sum_{i=1}^{7}x_i^2$	$\sum_{i=1}^{7}x_iy_i$	$\sum_{i=1}^{7}x_iz_i$	$\bar{x}$	$\bar{y}$	$\bar{z}$

5215	17713	714	27	81.3	3.6

(3)根据以往每年平均气温以及对果园年产值的统计，得到以下数据：平均气温在22℃以下的年数占60%，对柚子产量影响不大，不需要采取防虫措施；平均气温在22℃至28℃的年数占30%，柚子产量会下降20%；平均气温在28℃以上的年数占10%，柚子产量会下降50%.为了更好的防治红蜘蛛虫害，农科所研发出各种防害措施供果农选择.

在每年价格不变，无虫害的情况下，某果园年产值为200万元，根据以上数据，以得到最高收益（收益＝产值－防害费用）为目标，请为果农从以下几个方案中推荐最佳防害方案，并说明理由.

方案1：选择防害措施A，可以防止各种气温的红蜘蛛虫害不减产，费用是18万；

方案2：选择防害措施B，可以防治22℃至28℃的蜘蛛虫害，但无法防治28℃以上的红蜘蛛虫害，费用是10万；

方案3：不采取防虫害措施.

【解析】【答案】(1) $y=ce^{dx}$更适宜

(2) $y=e^{0.3x-4.5}$

(3)选择方案1最佳，理由见解析

【分析】（1）根据散点图的形状，可判断$y=ce^{dx}$更适宜作为平均产卵数y关于平均温度x的回归方程类型;

（2）将$y=ce^{dx}$两边同时取自然对数，转化为线性回归方程，即可得到答案；

（3）求出三种方案的收益的均值，根据均值越大作为判断标准.

【详解】（1）由散点图可以判断，$y=ce^{dx}$更适宜作为平均产卵数y关于平均温度x的回归方程类型.

（2）将$y=ce^{dx}$两边同时取自然对数，可得$\ln y=\ln c+dx$，

由题中的数据可得，$\sum_{i=1}^{7}x_iz_i-7\overline{xz}=33.6$，$\sum_{i=1}^{7}\left(x_i-\overline{x}\right)^2=\sum_{i=1}^{7}x_i^2-7\overline{x}^2=112$，

所以$d=\dfrac{\sum_{i=1}^{7}x_iz_i-7\overline{x}\,\overline{z}}{\sum_{i=1}^{7}x_i^2-7\overline{x}^2}=\dfrac{33.6}{112}=0.3$，

则$\ln c=\overline{z}-d\overline{x}=3.6-0.3\times27=-4.5$，

所以z关于x的线性回归方程为$z=0.3x-4.5$，

故y关于x的回归方程为$y=e^{0.3x-4.5}$；

（3）用X_1，X_2和X_3分别表示选择三种方案的收益.

采用第 1 种方案，无论气温如何，产值不受影响，收益为$200-18=182$万，即$X_1=182$

采用第 2 种方案，不发生 28℃以上的红蜘蛛虫害，收益为$200-10=190$万，

如果发生，则收益为$100-10=90$万，即$X_2=\begin{cases}190,\text{不发生}28℃\text{以上的红蜘蛛虫害}\\90,\text{发生}28℃\text{以上的红蜘蛛虫害}\end{cases}$，

同样，采用第 3 种方案，有$X_3=\begin{cases}200,\text{不发生虫害}\\160,\text{只发生}22\text{-}28℃\text{虫害}\\100,\text{发生}28℃\text{以上虫害}\end{cases}$

所以，$E(X_1)=182$，

$E(X_2)=190\times P(X_2=190)+90\times P(X_2=90)=190\times0.9+90\times0.1=171+9=180$，

$E(X_3)=200\times P(X_3=200)+160\times P(X_3=160)+100\times P(X_3=100)$

$=200\times0.6+160\times0.3+100\times0.1=178$.

显然，$E(X_1)$最大，所以选择方案 1 最佳.

【题型六】列联表与等高条形图

【典例 1】（2023·四川自贡·统考一模）2025 年四川省将实行 3＋1＋2 的高考模式，其中，“3”为语文、数学，外语 3 门参加全国统一考试，选择性考试科目为政治、历史、地理、物理、化学，生物 6 门，由考生根据报考高校以及专业要求，结合自身实际，首先在物理，历史中 2 选 1，再从政治、地理、化学、生物中 4 选 2，形成自己的高考选考组合.

(1)若某小组共 6 名同学根据方案进行随机选科，求恰好选到“物化生”组合的人数的期望；

(2)由于物理和历史两科必须选择 1 科，某校想了解高一新生选科的需求.随机选取 100 名高一新生进行调查，得到如下统计数据，写出下列联表中 a，d 的值，并判断是否有 95％的把握认为“选科与性别有关”？

	选择物理	选择历史	合计
男生	a	10	
女生	30	d	
合计		30	

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

$P\left(K^2>k_0\right)$	0.10	0.05	0.025	0.01	0.005
k_0	2.706	3.841	5.024	6.635	7.879

【解析】【答案】(1)$\frac{1}{2}$

(2)40，20，有 95％的把握认为“选科与性别有关

【分析】（1）根据列举法求出一个学生恰好选到“物化生”组合的概率，确定 6 名同学根据方案进行随机选科，符合二项分布，即可求得答案；

（2）由题意确定 a,d 的值，计算 K^2 的值，与临界值表比较，即得结论.

【详解】（1）设物理、历史 2 门科目为 m,n，政治、地理、化学、生物科目为 e,b,c,f，

则根据高考选考组合要求共有组合为 $(m,e,b),(m,e,c),(m,e,f),(m,b,c)$，

$(m,b,f),(m,c,f),(n,e,b),(n,e,c),(n,e,f),(n,b,c),(n,b,f),(n,c,f)$，共 12 种，

所以一个学生恰好选到“物化生”组合的概率为 $P=\frac{1}{12}$，

则 6 名同学根据方案进行随机选科，符合二项分布 $B\left(6,\frac{1}{12}\right)$，

故恰好选到“物化生”组合的人数的期望为 $6\times\frac{1}{12}=\frac{1}{2}$；

（2）由题意可得 $a=40,d=20$；

则 $K^2=\frac{100\left(40\times20-10\times30\right)^2}{50\times50\times70\times30}\approx4.762>3.841$，

所以有 95％的把握认为“选科与性别有关”.

【典例 2】（2023·四川德阳·统考一模）2023 年 11 月，世界首届人工智能峰会在英国举行，我国因为在该领域取得的巨大成就受邀进行大会发言.为了研究不同性别的学生对人工智能的了解情况，我市某著名高中进行了一次抽样调查，分别抽取男、女生各 50 人作为样本.设事件 $A=$“了解人工智能”，$B=$“学生为男生”，据统计 $P(A|\overline{B})=\frac{3}{5},P(B|A)=\frac{4}{7}$.

(1)根据已知条件，填写下列 2×2 列联表，是否有 99% 把握推断该校学生对人工智能的了解情况与性别有关？

	了解人工智能	不了解人工智能	合计
男生			

女生			
合计			

(2)将样本的频率视为概率，现从全校的学生中随机抽取 30 名学生，设其中了解人工智能的学生的人数为 X，求使得 $P(X=k)$ 取得最大值时的 $k\left(k\in \mathrm{N}^{*}\right)$ 值.

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

$P\left(K^2\geq k\right)$	0.100	0.050	0.010
k	2.706	3.841	6.635

【解析】【答案】(1)答案见详解

(2) $k=21$

【分析】（1）根据条件概率求得人数填写列联表，代入公式后求出观测值，将其与临界值比较即可求解；

（2）根据二项分布求出概率，根据单调性列出不等式组求解即可.

【详解】（1）因为 $P(A\mid \overline{B})=\dfrac{3}{5}, P(B\mid A)=\dfrac{4}{7}$，

所以了解人工智能的女生为 $50\times\dfrac{3}{5}=30$，

了解人工智能人数为 $\dfrac{30}{1-\dfrac{4}{7}}=70$，

则了解人工智能的男生有 $70-30=40$ 人，

结合男生和女生各有 50 人，填写 2×2 列联表为：

	了解人工智能	不了解人工智能	合计
男生	40	10	50
女生	30	20	50
合计	70	30	100

则 $K^2=\dfrac{100(40\times 20-10\times 30)^2}{50\times 50\times 30\times 70}=\dfrac{100}{21}\approx 4.762<6.635$，

故没有 99% 把握推断该校学生对人工智能的了解情况与性别有关.

（2）由(1)知，了解人工智能的频率为$\frac{70}{100}=0.7$，

所以随机变量$X\sim B(30,0.7)$，

则$P\left(X=k\right)=\mathrm{C}_{30}^{k}(0.7)^{k}(1-0.7)^{30-k}$

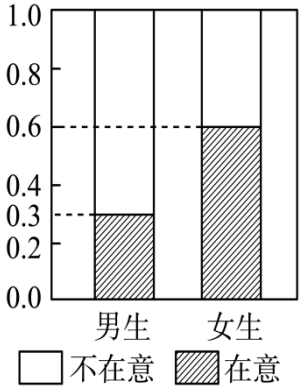
令$\begin{cases}\mathrm{C}_{30}^{k}(0.7)^{k}(1-0.7)^{30-k}\geq\mathrm{C}_{30}^{k-1}(0.7)^{k-1}(1-0.7)^{31-k}\\ \mathrm{C}_{30}^{k}(0.7)^{k}(1-0.7)^{30-k}\geq\mathrm{C}_{30}^{k+1}(0.7)^{k+1}(1-0.7)^{29-k}\end{cases}$，

解得$\frac{207}{10}\leq k\leq\frac{217}{10}$，又$k\in\mathrm{N}^{*}$，

所以$k=21$，

所以当$k=21$时，$P\left(X=k\right)$取得最大值.

【典例 3】（2023·山东烟台·统考二模）新修订的《中华人民共和国体育法》于 2023 年 1 月 1 日起施行，对于引领我国体育事业高质量发展，推进体育强国和健康中国建设具有十分重要的意义．某高校为调查学生性别与是否喜欢排球运动的关系，在全校范围内采用简单随机抽样的方法，分别抽取了男生和女生各 100 名作为样本，经统计，得到了如图所示的等高堆积条形图：



(1)根据等高堆积条形图，填写下列 2×2 列联表，并依据$\alpha=0.001$的独立性检验，是否可以认为该校学生的性别与是否喜欢排球运动有关联；

性别	是否喜欢排球运动	
	是	否
男生		
女生		

(2)将样本的频率视为概率，现从全校的学生中随机抽取 50 名学生，设其中喜欢排球运动的学生的人数为

X，求使得$P(X=k)$取得最大值时的k值.

附：$\chi^2=\dfrac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$，其中$n=a+b+c+d$，$x_{0.001}=10.828$.

【解析】【答案】(1)列联表见解析，有关联

(2)22

【分析】（1）结合条形等高图写出列联表，计算χ^2值即可判定；

（2）由题意知随机变量$X\sim B\left(50,\dfrac{9}{20}\right)$,结合二项分布的概率计算列不等式组求解即可.

【详解】（1）由等高堆积条形图知,2×2 列联表为：

性别	是否喜欢排球运动	
	是	否
男生	30	70
女生	60	40

零假设为H_0：性别与是否喜欢排球运动无关，根据列联表中的数据，

$\chi^2=\dfrac{200\times(40\times30-60\times70)^2}{100\times100\times110\times90}\approx18.182>10.828=x_{0.001}$，

依据$\alpha=0.001$的独立性检验，可以推断H_0不成立,即性别与是否喜欢排球运动有关.

（2）由(1)知,喜欢排球运动的频率为$\dfrac{90}{200}=\dfrac{9}{20}$,

所以,随机变量$X\sim B\left(50,\dfrac{9}{20}\right)$,

则$P(X=k)=C_{50}^k\left(\dfrac{9}{20}\right)^k\left(1-\dfrac{9}{20}\right)^{50-k}\ (0\le k\le50,k\in\mathbf{N})$，

令$\begin{cases}C_{50}^k\left(\dfrac{9}{20}\right)^k\left(1-\dfrac{9}{20}\right)^{50-k}\ge C_{50}^{k-1}\left(\dfrac{9}{20}\right)^{k-1}\left(1-\dfrac{9}{20}\right)^{51-k}\\ C_{50}^k\left(\dfrac{9}{20}\right)^k\left(1-\dfrac{9}{20}\right)^{50-k}\ge C_{50}^{k+1}\left(\dfrac{9}{20}\right)^{k+1}\left(1-\dfrac{9}{20}\right)^{49-k}\end{cases}$，解得$\dfrac{439}{20}\le k\le\dfrac{459}{20}$.

因为$k\in\mathbf{N}$,所以当$k=22$时,$P(X=k)$取得最大值.

【题型七】独立性检验

【典例 1】（2023·四川德阳·统考一模）2023 年 11 月，世界首届人工智能峰会在英国举行，我国因为在该领域取得的巨大成就受邀进行大会发言．为了研究不同性别的学生对人工智能的了解情况，我市某著名高中进行了一次抽样调查，分别抽取男、女生各 50 人作为样本．据统计女生中了解人工智能的占$\frac{3}{5}$，了解人工智能的学生中男生占$\frac{4}{7}$．

(1)根据已知条件，填写下列2×2列联表，是否有 99%把握推断该校学生对人工智能的了解情况与性别有关？

	了解人工智能	不了解人工智能	合计
男生			
女生			
合计			

(2)将样本的频率视为概率，现用分层抽样的方法从女生中抽取 5 人，再从 5 人中抽取 3 人了解恤况，求抽取的 3 人中至少有 2 人了解人工智能的概率．

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

$P(K^2\geq k)$	0.100	0.050	0.010
k	2.706	3.841	6.635

【解析】【答案】(1)填表见解析；没有

(2)$\frac{7}{10}$

【分析】（1）根据题意，得到2×2的列联表，利用公式求得K^2，结合附表即可得到结论；

（2）根据题意，利用分层抽样从中抽取人，结合古典概型的概率计算公式，即可求解.

【详解】（1）由女生中了解人工智能的占$\frac{3}{5}$，知 50 名女生中了解人工智能的有 30 人，又了解人工智能的学生中男生占$\frac{4}{7}$，

所以列联表为

	了解人工智能	不了解人工智能	合计

男生	40	10	50
女生	30	20	50
合计	70	30	100

所以 $K^2=\dfrac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}=\dfrac{100(40\times20-30\times10)^2}{70\times30\times50\times50}\approx4.76<6.635$

故没有99%把握推断该校学生对人工智能的了解情况与性别有关.

（2）分层抽样的方法从女生中抽取 5 人，所以 5 人中有 3 人了解人工智能，

另外 2 人不了解人工智能

设了解人工智能的 3 位女生为 a,b,c；不了解人工智能的 2 为女生为 1，2，

那么从 5 人中抽取 3 人的所有情况为 $abc,ab1,ab2,ac1,ac2,a12,bc1,bc2,b12,c12$ 共 10 种情况，

其中至少有 2 人了解人工智能的有 $abc,ab1,ab2,ac1,ac2,bc1,bc2$ 共 7 种情况.

故抽取的 3 人中至少有 2 人了解人工智能的概率为 $\dfrac{7}{10}$.

【典例 2】（2023·全国·模拟预测）为了纪念中国古代数学家祖冲之，2019 年 11 月 26 日，联合国教科文组织在第四十届大会宣布每年的 3 月 14 日为“国际数学日”．某高中为了让同学们感受数学魅力，传播数学文化，从 2020 年起，于每年的“国际数学日”开始举办为期一周的数学文化节，并且该校每年在数学文化节活动结束后，都会从全校学生中随机抽取 150 名学生了解他们参与活动的情况，经统计得到如下表格．

年份	2020	2021	2022	2023
年份代码 x	1	2	3	4
参与活动人数 y	95	100	105	120

(1)①已知可用线性回归模型拟合 y 与 x 之间的关系，求 y 关于 x 的回归方程 $\hat{y}=\hat{b}x+\hat{a}$；

②若该校共有 3600 名学生，据此预测 2024 年全校参与数学文化节活动的人数；

(2)2023 年，该校为了了解不同性别的学生对数学文化节是否满意，从参与数学文化节活动的学生中随机抽取 150 名，统计得到如下 2×2 列联表，判断是否有 90% 的把握认为该校学生对数学文化节活动是否满意与学生的性别有关.

	满意	不满意	合计