回归分析专项参考答案

1.【详解】由题中数据可知：（1）（3）为正相关，（2）（4）为负相关，

故$r_1>0,r_3>0,r_2<0,r_4<0$；

又因为（1）与（2）中散点图更接近于一条直线，故$r_1>r_3,r_2<r_4$，

所以$r_2<r_4<0<r_3<r_1$.

故选：C.

2.【详解】根据题意，依次分析 4 个命题：

对于①，当$r=0$时，成对样本数据之间不存在任何关系，而不是较弱，错误；

对于②，当$r>0$时，表明成对样本数据正相关，正确；

对于③，若线性回归方程中的回归系数$\hat{b}<0$，说明成对样本数据呈负相关，则相关系数$r<0$，错误；

对于④，$|r|$越接近 1，线性相关程度越强，$|r|$越接近 0，线性相关程度越弱，正确.

故答案为：①③.

3.【详解】$\bar{x}=\dfrac{1+2+3+4+5}{5}=3$，$\bar{y}=\dfrac{1.1+1.6+2+2.5+m}{5}=\dfrac{7.2+m}{5}$，

$\because \hat{y}=0.43x+0.71$，

$\therefore \dfrac{7.2+m}{5}=0.43\times 3+0.71$，

解得$m=2.8$.

故答案为：2.8.

4.【详解】对于 A 选项：线性回归方程$\hat{y}=\hat{b}x+0.28$必过点$\left(\bar{x},\bar{y}\right)$，$\bar{x}=3$，$\bar{y}=1$，解得$\hat{b}=0.24$，所以选项 A 正确；

对于 B 选项：当$x=8$时，$\hat{y}=0.24x+0.28$可以的出y的预测值为 2.2,所以 B 选项正确；

对于 C 选项：从小到大排列共有 5 个数据，则$i=5\times 40\%=2$是整数，则第 40 百分位数为从小到大排列的第 3 个数据，

即第 40 百分位数为 3，所以 C 选项错误；

对于 D 选项：因为相关系数为$r=\dfrac{\sum\limits_{i=1}^{n}\left(x_i-\bar{x}\right)\left(y_i-\bar{y}\right)}{\sqrt{\sum\limits_{i=1}^{n}\left(x_i-\bar{x}\right)^2\sum\limits_{i=1}^{n}\left(y_i-\bar{y}\right)^2}}$，

组样本数据的相关系数为：

$$r=\frac{\sum_{i=1}^{5}\left(x_i-\overline{x}\right)\left(y_i-\overline{y}\right)}{\sqrt{\sum_{i=1}^{5}\left(x_i-\overline{x}\right)^2\sum_{i=1}^{5}\left(y_i-\overline{y}\right)^2}}=\frac{(-2)\times(-0.5)+(-1)\times(-0.2)+0\times0+1\times0.2+2\times0.5}{\sqrt{4+1+0+1+4}\times\sqrt{(0.5)^2+(0.2)^2+0^2+(0.2)^2+(0.5)^2}},$$

去 掉 样 本 中 心 点 $(3,1)$ 后 相 关 系 数 为

$$r=\frac{\sum_{i=1}^{4}\left(x_i-\overline{x}\right)\left(y_i-\overline{y}\right)}{\sqrt{\sum_{i=1}^{4}\left(x_i-\overline{x}\right)^2\sum_{i=1}^{4}\left(y_i-\overline{y}\right)^2}}=\frac{(-2)\times(-0.5)+(-1)\times(-0.2)+1\times0.2+2\times0.5}{\sqrt{4+1+1+4}\times\sqrt{(0.5)^2+(0.2)^2+(0.2)^2+(0.5)^2}},$$

所以相关系数 r 不变,所以 D 选项正确；

故选：ABD.

5【详解】$\overline{x}=\frac{6+7+8+9+10}{5}=8,\overline{y}=\frac{3.5+4+5+5.5+7}{5}=5$，

所以 $\hat{a}=\overline{y}-0.85\overline{x}=5-0.85\times8=-1.8$，

所以 $x=10$ 时，$\widehat{y_5}=0.85\times10-1.8=8.5-1.8=6.7$，

所以残差为 $7-6.7=0.3$.

故答案为：0.3.

6.【详解】$\because$ 残差 $\widehat{e_i}=y_i-\widehat{y_i}$，当 $x=2$ 时，$\hat{y}=5$，当 $x=3$ 时，$\widehat{y=7}$，当 $x=4$ 时，$\hat{y}=9$，

$\therefore$ 残差平方和为 $(5.1-5)^2+(6.9-7)^2+(8.9-9)^2=0.03$

故答案为：0.03

7【详解】对于 A，若两变量 x，y 具有线性相关关系，则所有样本点都可能不在回归直线上，A 错误；

对于 B，若两变量 x，y 具有线性相关关系，则回归直线一定经过样本点中心 $(\overline{x},\overline{y})$，B 正确；

对于 C，因为 $y=ae^{hx}(a>0)$，所以 $\ln y=hx+\ln a$，即 $z=hx+\ln a$，又 $z=6x+\ln3$，所以 a，h 的估计值分别是 3 和 6，C 正确；

对于 D，残差平方和越小，拟合效果越好，D 正确；

故选：BCD.

8.【详解】解：由散点图可得二氧化碳排放量 y 与时间 x 正相关，故 A 正确；

因为 $R_2{}^2>R_1{}^2$，所以线性回归模型的拟合程度更好，故 B 正确；

当 $x=4$ 时，$\hat{y}=1.58\times4+91.44=97.76$，

而 $98.06-97.76=0.30$，故 C 错误；

当 $x=10$ 时，$\hat{y}=1.58\times10+91.44=107.24$，

即利用线性回归方程预计 2025 年中国二氧化碳排放量为 107.24 亿吨，故 D 正确.

故选：ABD.

9.【详解】由表格数据知：$\overline{x}=\frac{1}{5}\times\left(20+23+25+27+30\right)=25$，$\overline{z}=\frac{1}{5}\times\left(2+2.4+3+3+4.6\right)=3$，

因为数对$(\overline{x},\overline{z})$满足$z=0.2x+a$，得$a=3-0.2\times25=-2$，

∴$z=0.2x-2$，即$\ln y=0.2x-2$，∴$y=e^{0.2x-2}$，∴x＝35 时，$y=e^5$.

故当x＝35 时，蝗虫的产卵量y的估计值为e^5.

10.【详解】(1) ∵$0.92<0.99$，根据统计学知识可知：R^2越大，模型拟合效果越好，

∴应选择模型$\hat{y}=\hat{m}x^2+\hat{n}$.

(2) 令$u_i=x_i^2$，

$$\because\overline{u}=\frac{1+4+9+16+25}{5}=11,\quad \overline{y}=\frac{0.8+1.1+1.5+2.4+3.7}{5}=1.9,$$

$$\therefore\sum_{i=1}^{5}\left(u_i-\overline{u}\right)^2=\left(1-11\right)^2+\left(4-11\right)^2+\left(9-11\right)^2+\left(16-11\right)^2+\left(25-11\right)^2=374,$$

又$\sum_{i=1}^{5}\left(u_i-\overline{u}_i\right)\left(y_i-\overline{y}\right)=45.1$，

$$\therefore\hat{m}=\frac{\sum_{i=1}^{5}\left(u_i-\overline{u}\right)\left(y_i-\overline{y}\right)}{\sum_{i=1}^{5}\left(u_i-\overline{u}\right)^2}=\frac{45.1}{374}\approx0.121\approx0.12,\quad \hat{n}=\overline{y}-\hat{m}\overline{u}=1.9-0.121\times11=0.569\approx0.57,$$

∴y关于x的回归方程为$\hat{y}=0.12x^2+0.57$.

11.【详解】(1) $\overline{x}=\frac{1}{5}\times\sum_{i=1}^{5}x_i=\frac{1}{5}\times880=176$，$\overline{y}=\frac{1}{5}\times\sum_{i=1}^{5}y_i=\frac{1}{5}\times885=177$，

$$\hat{b}=\frac{\sum_{i=1}^{5}x_iy_i-5\overline{x}\cdot\overline{y}}{\sum_{i=1}^{5}x_i^2-5\overline{x}^2}=\frac{156045-5\times176\times177}{155450-5\times176^2}=0.5,\quad \hat{a}=\overline{y}-\hat{b}\overline{x}=177-0.5\times176=89,$$

故回归方程为：$\hat{y}=0.5x+89$，

取$\hat{y}=0.5x+89>x$，解得$x<178$，即$x<178$时，儿子比父亲高；

取$\hat{y}=0.5x+89<x$，解得$x>178$，即$x>178$时，儿子比父亲矮；

父亲较高时，儿子平均身高要矮于父亲，父亲较矮时，儿子平均身高要高于父亲，

即儿子身高有一个回归，回归到全种群平均高度的趋势.

(2) $\widehat{y_1}=0.5\times160+89=169$，$\widehat{e_1}=170-169=1$；

$\widehat{y_2}=0.5\times170+89=174$，$\widehat{e_2}=174-174=0$；

$\widehat{y_3}=0.5\times175+89=176.5$，$\widehat{e_3}=175-176.5=-1.5$；

$\widehat{y_4}=0.5\times185+89=181.5$，$\widehat{e_4}=180-181.5=-1.5$；

$\widehat{y_5}=0.5\times190+89=184$，$\widehat{e_5}=186-184=2$；

故残差的和为$1+0-1.5-1.5+2=0$.

对任意具有线性相关关系的变量$\sum_{i=1}^{n}\hat{e}_i=0$.

证明如下：

$$\sum_{i=1}^{n}\hat{e}_i=\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)=\sum_{i=1}^{n}\left(y_i-\hat{b}x_i-\hat{a}\right)=\sum_{i=1}^{n}y_i-\hat{b}\sum_{i=1}^{n}x_i-n\hat{a}=n\overline{y}-n\hat{b}\overline{x}-n\left(\overline{y}-\hat{b}\overline{x}\right)=0.$$

12.【详解】（1）解：

$$r=\frac{\sum_{i=1}^{5}\left(x_i-\overline{x}\right)\left(y_i-\overline{y}\right)}{\sqrt{\sum_{i=1}^{5}\left(x_i-\overline{x}\right)^2}\sqrt{\sum_{i=1}^{5}\left(y_i-\overline{y}\right)^2}}=\frac{\sum_{i=1}^{5}x_iy_i-5\overline{x}\overline{y}}{\sqrt{\sum_{i=1}^{5}x_i^2-5\overline{x}^2}\cdot\sqrt{\sum_{i=1}^{5}y_i^2-5\overline{y}^2}}=\frac{3765-5\times7\times113}{\sqrt{10\times3630}}\approx-0.9972,$$

所以，这两个变量负相关，且具有较强的线性相关性.

（2）解：①$\hat{b}=\frac{\sum_{i=1}^{5}x_iy_i-5\overline{x}\overline{y}}{\sum_{i=1}^{5}x_i^2-5\overline{x}^2}=\frac{-190}{10}=-19$，则$\hat{a}=\overline{y}-\hat{b}\overline{x}=113+19\times7=246$，

所以，Y关于x的经验回归方程为$\hat{y}=-19x+246$，

当$x=10$时，则$\hat{y}=-19\times10+246=56$，

所以，当$x=10$时，Y的预测值为56；

②由$\hat{e}_i=y_i-\hat{y}_i=y_i+19x_i-246$，计算得该回归模型的残差如下表所示：

x_i	9	8	7	6	5
$\hat{e}_i$	0	1	-3	3	-1

所以，残差的方差为$\hat{\sigma}^2=\frac{0^2+1^2+(-3)^2+3^2+(-1)^2}{5}=4$.

13.【详解】(1)解：依题意可$\overline{x}=\frac{165+170+175+170+170}{5}=170$，$\overline{y}=\frac{58+67+67+65+63}{5}=64$，

$$\sum_{i=1}^{n}(x_i-\overline{x})^2=(165-170)^2+(170-170)^2+(175-170)^2+(170-170)^2+(170-170)^2=50,$$

$$\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})=(165-170)\times(58-64)+(175-170)\times(67-64)=45$$

所以$\hat{b}=\frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sum_{i=1}^{n}(x_i-\overline{x})^2}=\frac{45}{50}=0.9$，$\hat{a}=\overline{y}-\hat{b}\overline{x}=64-0.9\times170=-89$

所以回归直线方程为$\hat{y}=0.9x-89$，

当$x=180$时$\hat{y}=0.9\times180-89=73$，所以身高为180cm的同学的体重大约为73kg；

（2）由（1）回归方程可得，各组数据的残差，如表所示：

学生编号	1	2	3	4	5
身高 x/cm	165	170	175	170	170
体重 y/kg	58	67	67	65	63
残差 $\hat{e}$	-1.5	3	-1.5	1	-1

所以 $\sum_{i=1}^{5}(y_i-\widehat{y_i})^2=(-1.5)^2+3^2+(-1.5)^2+1^2+(-1)^2=15.5$，

$\sum_{i=1}^{5}(y_i-\overline{y})^2=(-6)^2+3^2+3^2+1^2+(-1)^2=56$，

则 $R^2=1-\frac{15.5}{56}\approx 0.72$，

故学生的体重差异约有72% 是由身高引起的．

14【详解】(1) 根据图 2 可知，模型①的残差波动性很大，说明拟合关系较差；
模型②的残差波动性很小，基本分布在 0 的附近，说明拟合关系很好，所以选择模型②更适宜.

(2) (i) 设 $t=\sqrt{x}$，所以 $y=c+dt$，

所以 $\hat{d}=\frac{\sum_{i=1}^{10}\left(y_i-\overline{y}\right)\left(t_i-\overline{t}\right)}{\sum_{i=1}^{10}\left(t_i-\overline{t}\right)^2}=\frac{28.35}{4.5}=6.3$， $\hat{c}=\overline{y}-\hat{d}\overline{t}=75-6.3\times 2.25=60.825$，

所以 y 关于 x 的经验回归方程为 $y=60.825+6.3\sqrt{x}$

(ii) 由题设可得 $L=\left(111.225-y\right)\sqrt{x}=\left(111.225-6.3\sqrt{x}-60.825\right)\sqrt{x}=-6.3x+50.4\sqrt{x}$，

当取对称轴即 $\sqrt{x}=\frac{50.4}{2\times 6.3}=4$，即 $x=16$ 时，年利润 L 有最大值，

故该公司 2028 年的年利润最大.

15.【详解】(1) 解：由 $X\sim N(0.54,0.02^2)$，根据正态分布曲线的对称性，

可得 $P(X>0.56)=\frac{1-P(0.54-0.02<X<0.54+0.02)}{2}=0.15865.$

(2) 解：(i) 由散点图可知 y 与 x 的关系不是线性关系，所以 $y=cx^d$ 适宜作为粮食亩产量 y 关于每亩化肥施用量 x 的回归方程；

(ii) 因为 $y=cx^d$，所以 $\ln y=d\ln x+\ln c$，令 $t=\ln x, z=\ln y$，则 $\hat{z}=\widehat{d}t+\ln c$，

由表可得 $\bar{t}=1.5, \bar{z}=1.5$，所以 $d=\frac{\sum_{i=1}^{10}t_iz_i-10\bar{t}\cdot\bar{z}}{\sum_{i=1}^{10}t_i^2-10\bar{t}^2}=\frac{30.5-10\times 1.5\times 1.5}{46.6-10\times 1.5^2}\approx\frac{1}{3}$，

所以 $\ln c=\bar{z}-d\bar{t}=1.5-\frac{1}{3}\times 1.5=1$，所以 $c=e$，所以 $y=e\cdot x^{\frac{1}{3}}$，

当 $x=27$ 时，$y=3e\approx 3\times 2.7=8.1$（百公斤

独立性检验专项：

1.【详解】(1)

	$20^\circ C$	$30^\circ C$	合计
菌种甲	28	20	48
菌种乙	22	30	52
合计	50	50	100

零假设 H_0：菌种甲、乙的产量与温度没有关系，根据表中数据，计算得 $\chi^2=\dfrac{100\times(28\times30-20\times22)^2}{50\times50\times48\times52}\approx2.564<2.706$，

根据小概率值 $\alpha=0.1$ 的独立性检验，我们没有充分的证据推断 H_0 不成立，

因此可以认为 H_0 成立，即认为菌种甲、乙的产量与温度无关.

(2) 由题意可知，$X\sim B\left(3,\dfrac{5}{6}\right)$，$X$ 的可能取值有 $0,1,2,3$，

由公式可得 $P(X=0)=C_3^0\left(\dfrac{5}{6}\right)^0\left(\dfrac{1}{6}\right)^3=\dfrac{1}{216}$，$P(X=1)=C_3^1\left(\dfrac{5}{6}\right)^1\left(\dfrac{1}{6}\right)^2=\dfrac{15}{216}=\dfrac{5}{72}$，

$P(X=2)=C_3^2\left(\dfrac{5}{6}\right)^2\left(\dfrac{1}{6}\right)^1=\dfrac{75}{216}=\dfrac{25}{72}$，$P(X=3)=C_3^3\left(\dfrac{5}{6}\right)^3\left(\dfrac{1}{6}\right)^0=\dfrac{125}{216}$，

所以 X 的分布列为

X	0	1	2	3
P	$\dfrac{1}{216}$	$\dfrac{5}{72}$	$\dfrac{25}{72}$	$\dfrac{125}{216}$

所以 $E(X)=3\times\dfrac{5}{6}=\dfrac{5}{2}$.

2.【详解】(1) 2×2 列联表补充如下：

	选物理类	选历史类	合计
男生	35	15	50
女生	25	25	50
合计	60	40	100

零假设为 H_0：选科分类与性别无关联，

因为 $\chi^2=\dfrac{100\times(35\times25-25\times15)^2}{60\times40\times50\times50}=\dfrac{25}{6}\approx4.167>3.841=x_{0.05}$，

根据小概率值 $\alpha=0.05$ 的独立性检验，推断 H_0 不成立，

即认为选科分类与性别有关联，此推断犯错误的概率不大于 0.05.

(2) 由已知，50 名女学生中选择物理类和选择历史类的比例为 $1:1$，

因此抽取 6 名女生中，选择物理类和选择历史类的人数均为 3 名.

所以随机变量 X 的取值为 $1,2,3$.

$P(X=1)=\frac{C_3^1C_3^3}{C_6^4}=\frac{3}{15}=\frac{1}{5}$, $P(X=2)=\frac{C_3^2C_3^2}{C_6^4}=\frac{3}{5}$, $P(X=3)=\frac{C_3^3C_3^1}{C_6^4}=\frac{1}{5}$,

所以随机变量X的分布列如下表：

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

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