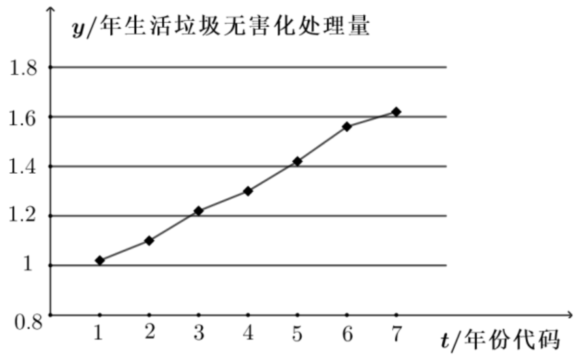
# 湛江一中 2023 届高三卓越班　NLXF2023—17

# 高三数学一轮复习——统计学——回归分析

1.如图是某地 2014 年至 2020 年生活垃圾无害化处理量（单位：万吨）的折线图.




注：年份代码 1～7 分别对应年份 2014～2020.

(1)由折线图看出，可用线性回归模型拟合 $y$ 与 $t$ 的关系，请用相关系数加以证明；

(2)建立 $y$ 关于 $t$ 的回归方程（系数精确到 0.01），预测 2022 年某地生活垃圾无害化处理量.

附注：若相关系数 $\mathrm{r}>0.75$，则可认为 $y$ 与 $t$ 之间存在较强的正相关关系.

参考数据：$\sum_{i=1}^{7}y_i=9.32$，$\sum_{i=1}^{7}t_iy_i=40.17$，$\sqrt{\sum_{i=1}^{7}\left(y_i-\bar{y}\right)^2}=0.55$，$\sqrt{7}\approx 2.646$.

参考公式：相关系数 $r=\dfrac{\sum_{i=1}^{n}\left(t_i-\bar{t}\right)\left(y_i-\bar{y}\right)}{\sqrt{\sum_{i=1}^{n}\left(t_i-\bar{t}\right)^2\sum_{i=1}^{n}\left(y_i-\bar{y}\right)^2}}$，回归方程 $\hat{y}=\hat{a}+\hat{b}t$ 中斜率和截距的最小二乘法估计公式分别为

$\hat{b}=\dfrac{\sum_{i=1}^{n}\left(t_i-\bar{t}\right)\left(y_i-\bar{y}\right)}{\sum_{i=1}^{n}\left(t_i-\bar{t}\right)^2}$，$\hat{a}=\bar{y}-\hat{b}\bar{t}$ .

（1）由折线图看出，$y$ 与 $t$ 之间存在较强的正相关关系，理由如下：

$\because \sum_{i=1}^{7}y_i=9.32$，$\sum_{i=1}^{7}t_iy_i=40.17$，$\sqrt{\sum_{i=1}^{7}\left(y_i-\bar{y}\right)^2}=0.55$，$\bar{t}=\dfrac{1+2+3+4+5+6+7}{7}=4$，

$$\therefore r=\frac{\sum_{i=1}^{7}\left(t_i-\bar{t}\right)\left(y_i-\bar{y}\right)}{\sqrt{\sum_{i=1}^{7}\left(t_i-\bar{t}\right)^2\sum_{i=1}^{7}\left(y_i-\bar{y}\right)^2}}=\frac{\sum_{i=1}^{7}t_iy_i-7\bar{t}\bar{y}}{\sqrt{\sum_{i=1}^{7}\left(t_i-\bar{t}\right)^2\sum_{i=1}^{7}\left(y_i-\bar{y}\right)^2}}\approx\frac{40.17-4\times 9.32}{2\sqrt{7}\times 0.55}\approx 0.993.$$

$\because 0.993>0.75$，故 $y$ 与 $t$ 之间存在较强的正相关关系.

(2)由（1）结合题中数据可得 $\hat{b}=\dfrac{\sum_{i=1}^{7}\left(t_i-\bar{t}\right)\left(y_i-\bar{y}\right)}{\sum_{i=1}^{7}\left(t_i-\bar{t}\right)^2}=\dfrac{\sum_{i=1}^{7}t_iy_i-7\bar{t}\bar{y}}{\sum_{i=1}^{7}t_i^2-7\bar{t}^2}\approx\dfrac{2.89}{28}\approx 0.103$，

$\hat{a}=\bar{y}-\hat{b}\bar{t}\approx 1.331-0.103\times 4\approx 0.92$，

$\therefore y$ 关于 $t$ 的回归方程 $\hat{y}=0.10t+0.92$，2022 年对应的 $t$ 值为 9，故 $\hat{y}=0.10\times 9+0.92=1.82$，

预测 2022 年该地生活垃圾无害化处理量为 1.82 万吨.

2.现代物流成为继劳动力、自然资源外影响企业生产成本及利润的重要因素.某企业去年前八个月的物流成本(单位：万元)和企业利润的数据(单位：万元)如下表所示：

| 月份 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| 物流成本 $x$ | 83 | 83.5 | 80 | 86.5 | 89 | 84.5 | 79 | 86.5 |
| 利润 $y$ | 114 | 116 | 106 | 122 | 132 | 114 | $m$ | 132 |
| 残差 $\hat{e}_i=y_i-\hat{y}_i$ | 0.2 | 0.6 | 1.8 | －3 | －1 | －4.6 | －1 | |

根据最小二乘法公式求得经验回归方程为 $\hat{y}=3.2x-151.8$.

(1)求 $m$ 的值，并利用已知的经验回归方程求出 8 月份对应的残差值 $\hat{e}_8$；

(2)请先求出线性回归模型 $\hat{y}=3.2x-151.8$ 的决定系数 $R^2$ (精确到 0.0001)，若根据非线性模型 $y=267.76\ln x-1069.2$ 求得解释变量(物流成本)对于响应变量(利润)的决定系数 $R_0^2=0.9057$，请说明以上两种模型哪种模型拟合效果更好.

参考公式及数据：$R^2=1-\dfrac{\sum\limits_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum\limits_{i=1}^{n}(y_i-\overline{y})^2}$，$\overline{x}=84$，$\sum\limits_{i=1}^{8}\left(y_i-\overline{y}\right)^2=904$.

(1)∵ $\hat{y}=3.2x-151.8$，$\overline{x}=84$，∴ $\overline{y}=3.2\times84-151.8=117$.则 $114+116+106+122+132+114+m+132=117\times8$，解得 $m=100$；8 月份对应的残差值 $\hat{e}_8=132-\left(3.2\times86.5-151.8\right)=7$.

(2) $\sum\limits_{i=1}^{8}\left(y_i-\hat{y}_i\right)^2=0.2^2+0.6^2+1.8^2+\left(-3\right)^2+\left(-1\right)^2+\left(-4.6\right)^2+\left(-1\right)^2+7^2=84.8$，

则 $R^2=1-\dfrac{\sum\limits_{i=1}^{8}\left(y_i-\hat{y}_i\right)^2}{\sum\limits_{i=1}^{8}\left(y_i-\overline{y}\right)^2}=1-\dfrac{84.8}{904}=0.9062>R_0^2$，∴线性回归模型 $\hat{y}=3.2x-151.8$ 拟合程度更好.

3.为了监控某种零件的一条生产线的生产过程，检验员每天从该生产线上随机抽取 16 个零件，并测量其尺寸（单位：cm）.根据长期生产经验，可以认为这条生产线正常状态下生产的零件的尺寸服从正态分布 $N\left(\mu,\sigma^2\right)$.

（1）假设生产状态正常，记 $X$ 表示一天内抽取的 16 个零件中其尺寸在 $\left(\mu-3\sigma,\mu+3\sigma\right)$ 之外的零件数，求 $P\left(X\geq1\right)$ 及 $X$ 的数学期望；

（2）一天内抽检的零件中，如果出现了尺寸在 $\left(\mu-3\sigma,\mu+3\sigma\right)$ 之外的零件，就认为这条生产线在这一天的生产过程可能出现了异常情况，需对当天的生产过程进行检查.

（i）试说明上述监控生产过程方法的合理性；

（ii）下面是检验员在一天内抽取的 16 个零件的尺寸：

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 9.95 | 10.12 | 9.96 | 9.96 | 10.01 | 9.92 | 9.98 | 10.04 |
| 10.26 | 9.91 | 10.13 | 10.02 | 9.22 | 10.04 | 10.05 | 9.95 |

经计算得 $\bar{x}=\frac{1}{16}\sum_{i=1}^{16}x_i=9.97$，$s=\sqrt{\frac{1}{16}\sum_{i=1}^{16}\left(x_i-\bar{x}\right)^2}=\sqrt{\frac{1}{16}\left(\sum_{i=1}^{16}x_i^2-16\bar{x}^2\right)}\approx0.212$，其中 $x_i$ 为抽取的第 $i$ 个零件的尺寸，$i=1,2,\cdots,16$.

用样本平均数 $\bar{x}$ 作为 $\mu$ 的估计值 $\hat{u}$，用样本标准差 $s$ 作为 $\sigma$ 的估计值 $\hat{\sigma}$，利用估计值判断是否需对当天的生产过程进行检查?剔除 $\left(\hat{\mu}-3\hat{\sigma},\hat{\mu}+3\hat{\sigma}\right)$ 之外的数据，用剩下的数据估计 $\mu$ 和 $\sigma$（精确到 0.01）.

附：若随机变量 $Z\sim N\left(\mu,\sigma^2\right)$，则 $P\left(\mu-3\sigma<Z<\mu+3\sigma\right)=0.9974$，$0.9974^{16}\approx0.9592$，$\sqrt{0.008}\approx0.09$.

【解析】（1）尺寸落在 $\left(\mu-3\sigma,\mu+3\sigma\right)$ 的概率为 0.9974，所以尺寸落在 $\left(\mu-3\sigma,\mu+3\sigma\right)$ 之外的概率为 0.0026，故 $X\sim B(16,0.0026)$，所以 $E\left(X\right)=16\times0.0026=0.0416$，且 $P\left(X\geq1\right)=1-P\left(X=0\right)=1-0.9974^{16}\approx0.0408$.

（2）(i) 如果生产状态正常，一个零件尺寸在 $\left(\mu-3\sigma,\mu+3\sigma\right)$ 之外的概率只有 0.0026，一天内抽取的 16 个零件中，出现尺寸在 $\left(\mu-3\sigma,\mu+3\sigma\right)$ 之外的概率为 0.0408，发生的概率很小，因此一旦发生这种情况，就有理由认为可能出现了异常情况，需对当天的生产过程进行检查，故监控生产过程的方法合理.

（ii）解法 1：由所给数据知 $\bar{x}=9.97$，$s\approx0.212$，所以 $\mu$ 和 $\sigma$ 的估计值分别为 $\hat{u}=9.97$，$\hat{\sigma}=0.212$，从而 $\hat{\mu}-3\hat{\sigma}=9.97-3\times0.212=9.334$，$\hat{\mu}+3\hat{\sigma}=9.97+3\times0.212=10.606$

抽检的 16 个尺寸中 $9.22\notin\left(9.334,10.606\right)$，故需对当天的生产过程进行检查

剔除 9.22 这个数据之后，$\mu$ 的估计值为 $\frac{9.97\times16-9.22}{15}=10.02$

由 $s=\sqrt{\frac{1}{16}\left(\sum_{i=1}^{16}x_i^2-16\bar{x}^2\right)}\approx0.212$ 可得 $\sum_{i=1}^{16}x_i^2=0.212^2\times16+16\times9.97^2\approx1591.134$

剔除 9.22 这个数据后，剩下的数据的样本方差约为 $\frac{1}{15}\times\left(1591.134-9.22^2-15\times10.02^2\right)\approx0.008$

所以 $\sigma$ 的估计值为 $\sqrt{0.008}\approx0.09$.

解法 2：由所给数据知 $\bar{x}=9.9'$，$s\approx0.212$，所以 $\mu$ 和 $\sigma$ 的估计值分别为 $\hat{u}=9.9'$，$\hat{\sigma}=0.212$，从而 $\hat{\mu}-3\hat{\sigma}=9.97\ \times3\ \ 0.212$ ，$\hat{\mu}+3\hat{\sigma}=9.97+3\times0.212=10.606$

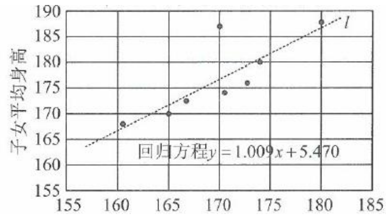
抽检的 16 个尺寸中 $9.22\notin\left(9.334,10.606\right)$，故需对当天的生产过程进行检查

剔除 9.22 这个数据之后，剩下的数据的平均数为 $\frac{9.97\times16-9.22}{15}=10.02$

$$\sigma=\sqrt{\frac{\left(9.95-10.02\right)^2+\left(10.12-10.02\right)^2+\left(9.96-10.02\right)^2+\cdots+\left(9.95-10.02\right)^2}{15}}=\sqrt{0.008}\approx0.09$$

4.研究表明，子女的平均身高 $y_i$（cm）与父母的平均身高 $x_i$（cm）有较强的线性相关性.某数学小组收集到 8 个家庭的相关数据，下面是小组制作的统计图（散点图、回归直线 $l$ 及回归方程）与原始数据表（局部缺失）：

父母与子女平均身高相关关系



父母平均身高

| 家庭编号 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| 父母平均身高（cm） | 160.5 | 165 | 167 | 170 | 170.5 | 173 | 174 | 180 |
| 子女平均身高（cm） | 168 | 170 | 172.5 | 187 | 174.5 | 176 | 180 | * |

（1）表中 8 号家庭的子女平均身高数据缺失，试根据统计学知识找回该数据；

（2）由图中观察到 4 号家庭的数据点明显偏离回归直线 $l$，试计算其残差（残差=观测值-预报值）；若剔除 4 号家庭数据点后，用余下的 7 个散点作线性回归分析，得到新的回归直线 $l'$，判断并证明 $l$ 与 $l'$ 的位置关系.

附:对于一组数据 $\left(x_1,y_1\right)$，$\left(x_2,y_2\right)$，…，$\left(x_n,y_n\right)$，其回归直线 $\hat{y}=\hat{b}x+\hat{a}$ 的斜率和截距的最小二乘估计分别为

$$\hat{b}=\frac{\sum_{i=1}^{n}\left(x_i-\overline{x}\right)\left(y_i-\overline{y}\right)}{\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2}=\frac{\sum_{i=1}^{n}x_iy_i-n\overline{xy}}{\sum_{i=1}^{n}x_i^2-n\overline{x}^2}，\quad \hat{a}=\overline{y}-\hat{b}\overline{x}.$$

【解析】(1)设 8 号家庭的子女平均身高为 $a$，由表中数据可得 $\overline{x}=\frac{160.5+165+167+170+170.5+173+174+180}{8}=170$

$\overline{y}=\frac{168+170+172.5+187+174.5+176+180+a}{8}=\frac{1228+a}{8}$ 因为样本中心点 $\left(\overline{x},\overline{y}\right)$ 在回归直线 $l$ 上，所以 $\frac{1228+a}{8}=1.009\times170+5.470$，解得：$a=188$.

（2）因为 $x_4=170=\overline{x}$，所以 4 号家庭子女平均身高的预报值 $\hat{y}_4=\overline{y}=\frac{1228+188}{8}=177$，

故残差为 $y_4-\hat{y}_4=187-177=10$，由题意，回归直线 $l$ 的斜率 $\hat{b}=\frac{\sum_{i=1}^{8}x_iy_i-8\overline{xy}}{\sum_{i=1}^{8}x_i^2-8\overline{x}^2}=1.009$，

剔除 4 号家庭的数据点后，$\overline{x}'=\frac{8\overline{x}-170}{7}=170$，$\overline{y}'=\frac{8\overline{y}-187}{7}=\frac{(1228+188)-187}{7}=\frac{1229}{7}$，

所以新的回归直线 $l'$ 的斜率 $\hat{b}'=\frac{\left(\sum_{i=1}^{8}x_iy_i-8\overline{xy}\right)-x_4y_4+8\overline{xy}-7\overline{x}'\overline{y}'}{\left(\sum_{i=1}^{8}x_i^2-8\overline{x}^2\right)-x_4^2+8\overline{x}^2-7\overline{x}'^2}$

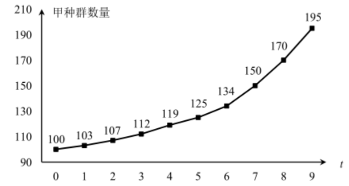
注意到 $-x_4y_4+8\overline{xy}-7\overline{x}'\overline{y}'=-170\times187+8\times170\times177-7\times170\times\frac{1229}{7}=0$，

$-x_4^2-8\overline{x}^2-7\overline{x}'^2=-170^2+8\times170^2-7\times170^2=0$，所以 $\hat{b}'=\frac{\left(\sum_{i=1}^{8}x_iy_i-8\overline{xy}\right)-x_4y_4+8\overline{xy}-7\overline{x}'\overline{y}'}{\left(\sum_{i=1}^{8}x_i^2-8\overline{x}^2\right)-x_4^2+8\overline{x}^2-7\overline{x}'^2}=\frac{\sum_{i=1}^{8}x_iy_i-8\overline{xy}}{\sum_{i=1}^{8}x_i^2-8\overline{x}^2}=\hat{b}$，

显然新的样本中心点 $\left(\overline{x}',\overline{y}'\right)$ 不在直线 $l$ 上，从而直线 $l'$ 与直线 $l$ 平行.

5.在某生态系统中，有甲、乙两个种群，两种群之间为竞争关系．设 $t$ 时刻甲、乙种群的数量分别为 $f\left(t\right)$，$g\left(t\right)$（起始时刻为 $t=0$）．由数学家 *Lotka* 和 *Volterra* 提出的模型是函数 $f\left(t\right)$，$g\left(t\right)$ 满足方程 $f'\left(t\right)=af\left(t\right)-bf\left(t\right)g\left(t\right)$，$g'\left(t\right)=cg\left(t\right)-df\left(t\right)g\left(t\right)$，其中 $a$，$b$，$c$，$d$ 均为非负实数.

(1)下图为没有乙种群时，一段时间内甲种群数量与时间的关系折线图．为预测甲种群的数量变化趋势，研究人员提出了两种可能的数学模型：① $f\left(t\right)=m\sqrt{t}+n$；② $f\left(t\right)=m\cdot n^t$，其中 $m$，$n$ 均为大于 1 的正数．根据折线图判断，应选用哪种模型进行预测，并说明理由.




(2)设 $a=c=0.08$，$d=2b=0.008$.

①函数 $F(t)=\mathrm{e}^{-0.08t}\left[2f(t)-g(t)\right]$ 的单调性；

②根据①中的结论说明：在绝大多数情况下，经过充分长的时间后，或者甲种群灭绝，或者乙种群灭绝.

注：在题设条件下，各种群数量均有上限值.

【答案】(1)应选用模型②预测甲种群数量的变化趋势；理由见解析

(2)① $F(t)$ 为常函数；②答案见解析

（1）由折线图知，甲种群数量的增长速度随着时间的推移而加快．而增长速度大致对应种群数量对时间的导数．如选用模型①，$f'(t)=\dfrac{m}{2\sqrt{t}}$，$f'(t)$ 是关于时间的减函数，不符合折线图；如选用模型②，$f'(t)=mn^t\ln n$，$f'(t)$ 是关于时间的增函数，符合折线图．所以应选用模型②预测甲种群数量的变化趋势

（2）由题设知 $f'(t)=0.08f(t)-0.004f(t)g(t)$，$g'(t)=0.08g(t)-0.008f(t)g(t)$．（i）$F(t)=\mathrm{e}^{-0.08t}[2f(t)-g(t)]$，$F'(t)=\mathrm{e}^{-0.08t}\left[-0.16f(t)+0.08g(t)+2f'(t)-g'(t)\right]$．消去条件中的 $f(t)g(t)$ 得 $g'(t)-0.08g(t)=2\left[f'(t)-0.08f(t)\right]$，所以 $F'(t)=0$．所以 $F(t)$ 为常函数．（ii）由（i），$F(t)=F(0)=2f(0)-g(0)$，$2f(t)-g(t)=[2f(0)-g(0)]\mathrm{e}^{0.08t}$．由于各种群数量均有上限值，不妨设甲乙种群数量的上限值分别为 $M_1$，$M_2$．①若 $g(0)>2f(0)$，$g(t)>2f(t)$．则当 $t>\dfrac{25}{2}\ln\dfrac{M_2-2}{g(0)-2f(0)}$ 时，$f(t)=\dfrac{1}{2}\left[g(t)-(2f(0)-g(0))\mathrm{e}^{0.08t}\right]<1$，此时可以近似认为甲种群灭绝；②若 $g(0)<2f(0)$，$g(t)<2f(t)$．则当 $t>\dfrac{25}{2}\ln\dfrac{2M_1}{2f(0)-g(0)}$ 时，$g(t)=2f(t)-(2f(0)-g(0))\mathrm{e}^{0.08t}<1$，此时可以近似认为乙种群灭绝；③若 $g(0)=2f(0)$，$g(t)=2f(t)$，甲乙种群数量之比保持恒定，可能不出现灭绝的情况．综上所述，对所有 $g(0)\neq 2f(0)$ 的情况，经过充分长的时间后，或者甲种群灭绝，或者乙种群灭绝

6.规定抽球试验规则如下：盒子中初始装有白球和红球各一个，每次有放回的任取一个，连续取两次，将以上过程记为一轮．如果每一轮取到的两个球都是白球，则记该轮为成功，否则记为失败．在抽取过程中，如果某一轮成功，则停止；否则，在盒子中再放入一个红球，然后接着进行下一轮抽球，如此不断继续下去，直至成功．

(1)某人进行该抽球试验时，最多进行三轮，即使第三轮不成功，也停止抽球，记其进行抽球试验的轮次数为随机变量 $X$，求 $X$ 的分布列和数学期望；

(2)为验证抽球试验成功的概率不超过 $\dfrac{1}{2}$，有 1000 名数学爱好者独立的进行该抽球试验，记 $t$ 表示成功时抽球试验的轮次数，$y$ 表示对应的人数，部分统计数据如下：

| $t$ | 1 | 2 | 3 | 4 | 5 |
| --- | --- | --- | --- | --- | --- |
| $y$ | 232 | 98 | 60 | 40 | 20 |

求 $y$ 关于 $t$ 的回归方程 $y=\dfrac{\hat{b}}{t}+a$，并预测成功的总人数（精确到 1）；

(3)证明：$\dfrac{1}{2^2}+\left(1-\dfrac{1}{2^2}\right)\dfrac{1}{3^2}+\left(1-\dfrac{1}{2^2}\right)\left(1-\dfrac{1}{3^2}\right)\dfrac{1}{4^2}+\cdots+\left(1-\dfrac{1}{2^2}\right)\left(1-\dfrac{1}{3^2}\right)\cdots\left(1-\dfrac{1}{n^2}\right)\dfrac{1}{(n+1)^2}<\dfrac{1}{2}$.

附：经验回归方程系数：$\hat{b}=\dfrac{\sum\limits_{i=1}^{n}x_iy_i-n\overline{x}\cdot\overline{y}}{\sum\limits_{i=1}^{n}x_i^2-n\overline{x}^2}$，$\hat{a}=\overline{y}-\hat{b}\overline{x}$；

参考数据：$\sum\limits_{i=1}^{5}x_i^2=1.46$，$\overline{x}=0.46$，$\overline{x}^2=0.212$（其中 $x_i=\dfrac{1}{t_i}$，$\overline{x}=\dfrac{1}{5}\sum\limits_{i=1}^{5}x_i$）.

（1）由题知，$X$ 的取值可能为 1，2，3 所以 $P(X=1)=\left(\dfrac{1}{C_2^1}\right)^2=\dfrac{1}{4}$；

$P(X=2)=\left[1-\left(\dfrac{1}{C_2^1}\right)^2\right]\left(\dfrac{1}{C_3^1}\right)^2=\dfrac{1}{12}$；$P(X=3)=\left[1-\left(\dfrac{1}{C_2^1}\right)^2\right]\left[1-\left(\dfrac{1}{C_3^1}\right)^2\right]=\dfrac{2}{3}$；

所以 $X$ 的分布列为：

| $X$ | 1 | 2 | 3 |
| --- | --- | --- | --- |
| $P$ | $\frac{1}{4}$ | $\frac{1}{12}$ | $\frac{2}{3}$ |

所以数学期望为$E\left(X\right)=1\times\frac{1}{4}+2\times\frac{1}{12}+3\times\frac{2}{3}=\frac{3+2+24}{12}=\frac{29}{12}$.

（2）令 $x_i=\frac{1}{t_i}$，则 $\hat{y}=\hat{b}x+\hat{a}$，由题知：$\sum\limits_{i=1}^{5}x_iy_i=315$，$\overline{y}=90$，

所以 $\hat{b}=\frac{\sum\limits_{i=1}^{5}x_iy_i-5\overline{x}\cdot\overline{y}}{\sum\limits_{i=1}^{5}x_i^2-5\overline{x}^2}=\frac{315-5\times0.46\times90}{1.46-5\times0.212}=\frac{108}{0.4}=270$，所以 $a=90-270\times0.46=-34.2$，$y=270x-34.2$，

故所求的回归方程为：$y=\frac{270}{t}-34.2$，所以，估计 $t=6$ 时，$y\approx11$；估计 $t=7$ 时，$y\approx4$；估计 $t\geq8$ 时，$y<0$；

预测成功的总人数为$450+11+4=465$.

（3）由题知，在前 $n$ 轮就成功的概率为

$$P=\frac{1}{2^2}+\left(1-\frac{1}{2^2}\right)\frac{1}{3^2}+\left(1-\frac{1}{2^2}\right)\left(1-\frac{1}{3^2}\right)\frac{1}{4^2}+\cdots+\left(1-\frac{1}{2^2}\right)\left(1-\frac{1}{3^2}\right)\cdots\cdots\left(1-\frac{1}{n^2}\right)\frac{1}{\left(n+1\right)^2}$$

又因为在前 $n$ 轮没有成功的概率为$1-P=\left(1-\frac{1}{2^2}\right)\times\left(1-\frac{1}{3^2}\right)\times\cdots\times\left[1-\frac{1}{\left(n+1\right)^2}\right]$

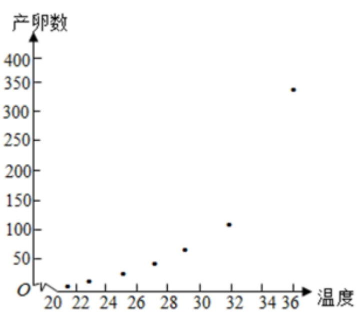
$$=\left(1-\frac{1}{2}\right)\left(1+\frac{1}{2}\right)\times\left(1-\frac{1}{3}\right)\times\left(1+\frac{1}{3}\right)\times\cdots\times\left(1-\frac{1}{n}\right)\times\left(1+\frac{1}{n}\right)\times\left(1-\frac{1}{n+1}\right)\times\left(1+\frac{1}{n+1}\right)$$

$$=\left(\frac{1}{2}\right)\times\left(\frac{3}{2}\right)\times\left(\frac{2}{3}\right)\times\left(\frac{4}{3}\right)\times\cdots\times\left(\frac{n-1}{n}\right)\times\left(\frac{n+1}{n}\right)\times\left(\frac{n}{n+1}\right)\times\left(\frac{n+2}{n+1}\right)=\frac{n+2}{2n+2}=\frac{\frac{1}{2}\left(2n+2\right)+1}{2n+2}=\frac{1}{2}+\frac{1}{2n+2}>\frac{1}{2}，$$

故$\frac{1}{2^2}+\left(1-\frac{1}{2^2}\right)\frac{1}{3^2}+\left(1-\frac{1}{2^2}\right)\left(1-\frac{1}{3^2}\right)\frac{1}{4^2}+\cdots+\left(1-\frac{1}{2^2}\right)\left(1-\frac{1}{3^2}\right)\cdots\cdots\left(1-\frac{1}{n^2}\right)\frac{1}{\left(n+1\right)^2}<\frac{1}{2}$.

7.红铃虫是棉花的主要害虫之一，能对农作物造成严重伤害，每只红铃虫的平均产卵数 $y$ 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.

| | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 平均温度 $x/℃$ | 21 | 23 | 25 | 27 | 29 | 31 | 33 |
| 平均产卵数 $y$/个 | 7 | 11 | 21 | 24 | 66 | 115 | 325 |
| $z=\ln y$ | 1.9 | 2.4 | 3.0 | 3.2 | 4.2 | 4.7 | 5.8 |




（1）根据散点图判断，$y=bx+a$ 与 $y=ce^{dx}$（其中 $e=2.718\cdots$ 为自然对数的底数）哪一个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？（给出判断即可，不必说明理由）并由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 0.01）

（2）根据以往统计，该地每年平均温度达到 28℃ 以上时红铃虫会造成严重伤害，需要人工防治，其他情况均不需要人工防治，记该地每年平均温度达到 28℃ 以上的概率为 $p$.记该地今后 5 年中，恰好需要 3 次人工防治的概率为 $f\left(p\right)$，求 $f\left(p\right)$ 的最大值，并求出相应的概率 $p_0$.

附：回归方程 $\hat{y}=\hat{b}x+\hat{a}$ 中，$\hat{b}=\dfrac{\sum\limits_{i=1}^{n}\left(x_i-\overline{x}\right)\left(y_i-\overline{y}\right)}{\sum\limits_{i=1}^{n}\left(x_i-\overline{x}\right)^2}=\dfrac{\sum\limits_{i=1}^{n}x_iy_i-n\overline{x}\overline{y}}{\sum\limits_{i=1}^{n}x_i^2-n\overline{x}^2}$，$\hat{a}=\overline{y}-\hat{b}\overline{x}$.

| 参考数据 | | | | |
|---|---|---|---|---|
| $\sum\limits_{i=1}^{7}x_i^2$ | $\sum\limits_{i=1}^{7}x_iy_i$ | $\sum\limits_{i=1}^{7}x_iz_i$ | $\overline{y}$ | $\overline{z}$ |
| 5215 | 17713 | 717 | 81.3 | 3.6 |

【详解】解：（1）由散点图可以判断，$y=ce^{dx}$ 适宜作为卵数 $y$ 关于温度 $x$ 的回归方程类型.

对 $y=ce^{dx}$ 两边取自然对数，得 $\ln y=\ln c+dx$，

令 $\hat{z}=\ln y$，$a=\ln c$，$\hat{b}=d$，则 $\hat{z}=\hat{b}x+a$，

由数据得 $\overline{x}=\dfrac{21+23+25+27+29+31+33}{7}=27$，$\sum\limits_{i=1}^{7}\left(x_iz_i-7\overline{x}\overline{z}\right)=36.6$，$\sum\limits_{i=1}^{7}\left(x_i-\overline{x}\right)^2=\sum\limits_{i=1}^{7}x_i^2-7\overline{x}^2=112$，

所以 $\hat{b}=\dfrac{\sum\limits_{i=1}^{7}\left(x_iz_i-7\overline{x}\overline{z}\right)}{\sum\limits_{i=1}^{7}x_i^2-7\overline{x}^2}=\dfrac{36.6}{112}\approx 0.33$，$a=\overline{z}-\hat{b}\overline{x}=3.6-0.33\times 27=-5.31$，

所以 $z$ 关于 $x$ 的线性回归方程为 $\hat{z}=0.33x-5.31$，则 $y$ 关于 $x$ 的回归方程为 $y=e^{0.33x-5.31}$；

（2）由 $f\left(p\right)=C_5^3\cdot p^3\left(1-p\right)^2$ 得 $f'\left(p\right)=C_5^3\cdot p^2\left(1-p\right)\left(3-5p\right)$，

因为 $0<p<1$，令 $f'\left(p\right)>0$ 得 $3-5p>0$，解得 $0<p<\dfrac{3}{5}$；

所以 $f\left(p\right)$ 在 $\left(0,\dfrac{3}{5}\right)$ 上单调递增，在 $\left(\dfrac{3}{5},1\right)$ 上单调递减，所以 $f\left(p\right)$ 有唯一的极大值为 $f\left(\dfrac{3}{5}\right)$，也是最大值；

所以当 $p_0=\dfrac{3}{5}$ 时，$f\left(p\right)_{\max}=f\left(\dfrac{3}{5}\right)=\dfrac{216}{625}$.

8.在疫情防控常态化的背景下，山东省政府各部门在保安全，保稳定的前提下有序恢复生产，生活和工作秩序，五一期间，文旅部门在落实防控举措的同时，推出了多款套票文旅产品，得到消费者的积极回应．下面是文旅部门在某地区推出六款不同价位的旅游套票，每款的套票价格 $x$（单位：元）与购买人数 $y$（单位：万人）的数据如下表：

| 旅游类别 | 城市展馆科技游 | 乡村特色游 | 齐鲁红色游 | 登山套票 | 游园套票 | 观海套票 |
|---|---|---|---|---|---|---|
| 套票价格 $x$（元） | 39 | 49 | 58 | 67 | 77 | 86 |
| 购买数量 $y$（万人） | 16.7 | 18.7 | 20.6 | 22.5 | 24.1 | 25.6 |

在分析数据、描点绘图中，发现散点 $\left(v_i，\omega_i\right)\left(1\le i\le 6\right)$ 集中在一条直线附近，其中 $v_i=\ln x_i$，$\omega_i=\ln y_i$

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

②对于一组数据 $\left(v_1，\omega_1\right),\left(v_2，\omega_2\right),\cdots,\left(v_n，\omega_n\right)$，其回归直线 $\hat{\omega}=\hat{b}v+\hat{a}$ 的斜率和截距的最小二乘估计值分别为

$\hat{b}=\dfrac{\sum\limits_{i=1}^{n}v_i\omega_i-n\overline{v}\overline{\omega}}{\sum\limits_{i=1}^{n}v_i^2-n\overline{v}^2}$，$\hat{a}=\overline{\omega}-\hat{b}\overline{v}$

(1)根据所给数据，求 $y$ 关于 $x$ 的回归方程；

(2)按照文旅部门的指标测定，当购买数量$y$与套票价格$x$的比在区间$\left[\frac{e}{9},\frac{e}{7}\right]$上时，该套票受消费者的欢迎程度更高，可以被认定为"热门套票"，现有三位同学从以上六款旅游套票中，购买不同的三款各自旅游．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．

【答案】(1) $y=ex^{\frac{1}{2}}$

(2)分布列见解析，数学期望为2.

（1）

解：$\because$散点$(v_i,\ \omega_i)(1\le i\le 6)$集中在一条直线附近，设回归直线方程为$\hat{\omega}=\hat{b}v+\hat{a}$

由$\bar{v}=\frac{1}{6}\sum_{i=1}^{6}v_i=4.1$，$\bar{\omega}=\frac{1}{6}\sum_{i=1}^{6}\omega_i=3.05$，则$\hat{b}=\frac{\sum_{i=1}^{n}v_i\omega_i-n\bar{v}\bar{\omega}}{\sum_{i=1}^{n}v_i^2-n\bar{v}^2}=\frac{75.3-6\times4.1\times3.05}{101.4-6\times4.1^2}=\frac{1}{2}$，$\hat{a}=\bar{\omega}-\hat{b}\bar{v}=3.05-\frac{1}{2}\times4.1=1$，

$\therefore$变量$\omega$关于$v$的回归方程为$\omega=\frac{1}{2}v+1$，$\because v_i=\ln x_i$，$\omega_i=\ln y_i$，$\therefore \ln y=\frac{1}{2}\ln x+1$，$\therefore y=ex^{\frac{1}{2}}$，

综上，$y$关于$x$的回归方程为$y=ex^{\frac{1}{2}}$；

（2）解：由$\frac{y}{x}=\frac{ex^{\frac{1}{2}}}{x}=\frac{e}{x^{\frac{1}{2}}}\in\left[\frac{e}{9},\frac{e}{7}\right]$，解得$49\leqslant x\leqslant 81$，$\therefore x=49,58,67,77$，

$\therefore$乡村特色游，齐鲁红色游，登山套票，游园套票为"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，

$P(X=1)=\frac{C_4^1C_2^2}{C_6^3}=\frac{1}{5}$，$P(X=2)=\frac{C_4^2C_2^1}{C_6^3}=\frac{3}{5}$，$P(X=3)=\frac{C_4^3}{C_6^3}=\frac{1}{5}$

$\therefore$ $X$的分布列为：

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

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

9.数据显示，中国在线直播用户规模及在线直播购物规模近几年都保持高速增长态势，下表为2017－2021年中国在线直播用户规模（单位：亿人），其中2017年－2021年对应的代码依次为1－5.

| 年份代码$x$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 市场规模$y$ | 3.98 | 4.56 | 5.04 | 5.86 | 6.36 |

参考数据：$\bar{y}=5.16$，$\bar{v}=1.68$，$\sum_{i=1}^{5}v_iy_i=45.10$，其中$v_i=\sqrt{x_i}$．

参考公式：对于一组数据$(v_1,\ y_1)$，$(v_2,\ y_2)$，…，$(v_n,\ y_n)$，其回归直线$\hat{y}=\hat{b}v+\hat{a}$的斜率和截距的最小二乘估计公式分别为$\hat{b}=\frac{\sum_{i=1}^{n}v_iy_i-n\bar{v}\bar{y}}{\sum_{i=1}^{n}v_i^2-n\bar{v}^2}$，$\hat{a}=\bar{y}-\hat{b}\bar{v}$．

(1)由上表数据可知，可用函数模型$\hat{y}=\hat{b}\sqrt{x}+\hat{a}$拟合$y$与$x$的关系，请建立$y$关于$x$的回归方程（$\hat{a}$，$\hat{b}$的值精确到0.01）；

(2)已知中国在线直播购物用户选择在品牌官方直播间购物的概率为$p$，现从中国在线直播购物用户中随机抽取4人，记这4人中选择在品牌官方直播间购物的人数为$X$，若$P(X=3)=P(X=4)$，求$X$的分布列与期望．

（1）设$v=\sqrt{x}$，则$\hat{y}=\hat{b}v+\hat{a}$，因为$\bar{y}=5.16$，$\bar{v}=1.68$，$\sum_{i=1}^{5}v_i^2=\sum_{i=1}^{5}x_i=15$，

所以 $\hat{b}=\frac{\sum_{i=1}^{5}v_iy_i-5\overline{v}\overline{y}}{\sum_{i=1}^{5}v_i^2-5\overline{v}^2}=\frac{45.10-5\times1.68\times5.16}{15-5\times1.68^2}=\frac{1.756}{0.888}\approx1.98$．

把 $(1.68,5.16)$ 代入 $\hat{y}=\hat{b}v+\hat{a}$，得 $\hat{a}=5.16-1.98\times1.68\approx1.83$．

即 $y$ 关于 $x$ 的回归方程为 $\hat{y}=1.98\sqrt{x}+1.83$；

（2）

由题意知 $X\sim B(4,\ p)$，$P(X=3)=C_4^3p^3(1-p)=4p^3(1-p)$，$P(X=4)=C_4^4p^4=p^4$，由 $4p^3(1-p)=p^4$ 得 $p=\frac{4}{5}$，

所以，$X$ 的取值依次为 0，1，2，3，4，$P(X=0)=C_4^0\left(1-\frac{4}{5}\right)^4=\frac{1}{625}$，$P(X=1)=C_4^1\cdot\frac{4}{5}\cdot\left(1-\frac{4}{5}\right)^3=\frac{16}{625}$，

$P(X=2)=C_4^2\left(\frac{4}{5}\right)^2\left(1-\frac{4}{5}\right)^2=\frac{96}{625}$，$P(X=3)=C_4^3\left(\frac{4}{5}\right)^3\left(1-\frac{4}{5}\right)=\frac{256}{625}$，$P(X=4)=C_4^4\left(\frac{4}{5}\right)^4=\frac{256}{625}$，所以 $X$ 的分布列为

| $X$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| $P$ | $\frac{1}{625}$ | $\frac{16}{625}$ | $\frac{96}{625}$ | $\frac{256}{625}$ | $\frac{256}{625}$ |

$E(X)=4\times\frac{4}{5}=\frac{16}{5}$．

**10.在国家大力发展新能源汽车产业的政策下，我国新能源汽车的产销量高速增长．已知某地区 2014 年底到 2021 年底新能源汽车保有量的数据统计表如下：**

| **年份（年）** | **2014** | **2015** | **2016** | **2017** | **2018** | **2019** | **2020** | **2021** |
|---|---|---|---|---|---|---|---|---|
| **年份代码 $x$** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** |
| **保有量 $y$/千辆** | **1.95** | **2.92** | **4.38** | **6.58** | **9.87** | **15.00** | **22.50** | **33.70** |

**(1)根据统计表中的数据判断，$\hat{y}=\hat{b}x+\hat{a}$ 与 $\hat{y}=e^{\hat{c}x+\hat{d}}$ 哪一个更适合作为 $y$ 关于 $x$ 的经验回归方程（给出判断即可，不必说明理由），并根据你的判断结果建立 $y$ 关于 $x$ 的经验回归方程；**

**(2)假设每年新能源汽车保有量按（1）中求得的函数模型增长，且传统能源汽车保有量每年下降的百分比相同．若 2021 年底该地区传统能源汽车保有量为 500 千辆，预计到 2026 年底传统能源汽车保有量将下降 10%．试估计到哪一年底新能源汽车保有量将超过传统能源汽车保有量．**

**参考数据：**

$\overline{y}=12.1$，$\overline{t}=2.1$，$\sum_{i=1}^{8}x_i^2=204,\sum_{i=1}^{8}x_iy_i=613.7,\sum_{i=1}^{8}x_it_i=92.4$，**其中** $t_i=\ln y_i$，$\lg2\approx0.30$，$\lg3\approx0.48$，$\lg e\approx0.43$．

**参考公式：**

**对于一组数据（$u_1$，$v_1$），（$u_2$，$v_2$），…，（$u_n$，$v_n$），其经验回归直线 $\hat{v}=\hat{\beta}u+\hat{\alpha}$ 的斜率和截距的最小二乘估计公式分别为** $\hat{\beta}=\frac{\sum_{i=1}^{n}(u_i-\overline{u})(v_i-\overline{v})}{\sum_{i=1}^{n}(u_i-\overline{u})^2}=\frac{\sum_{i=1}^{n}u_iv_i-n\overline{u}\cdot\overline{v}}{\sum_{i=1}^{n}u_i^2-n\overline{u}^2},\hat{a}=\overline{v}-\beta\overline{u}$；

(1)根据该地区新能源汽车保有量的增长趋势知，应选择的函数模型是 $y=e^{\hat{c}x+\hat{d}}$，令 $t=\ln y$，则 $t=\hat{c}x+\hat{d}$，

因为 $\overline{x}=4.5$，$\sum_{i=1}^{8}x_i^2=204,\sum_{i=1}^{8}x_it_i=92.4,$，所以 $\hat{c}=\frac{\sum_{i=1}^{8}x_it_t-8\overline{x}\cdot\overline{t}}{\sum_{i=1}^{n}x_i^2-8\overline{x}^2}=\frac{92.4-8\times4.5\times2.1}{204-8\times4.5^2}=\frac{16.8}{42}=0.4$，

$\hat{d}=\overline{t}-\hat{c}\overline{x}=2.1-0.4\times4.5=0.3$． 所以 $y=e^{0.4x+0.3}$．

(2)设传统能源汽车保有量每年下降的百分比为 $r$，

依题意得，$500(1-r)^5=500(1-10\%)$，解得 $1-r=0.9^{\frac{1}{5}}$，

设从 2021 年底起经过 $x$ 年后的传统能源汽车保有量为 $y$ 千辆，

则有 $y=500(1-r)^x=500\left(0.9^{\frac{1}{5}}\right)^x$，

设从 2021 年底起经过 $x$ 年后新能源汽车的数量将超过传统能源汽车，则有 $e^{0.4(x+8)+0.3}>500\left(0.9^{\frac{1}{5}}\right)^x$.

所以 $(0.4x+3.5)\lg e>3-\lg 2+0.2x(2\lg 3-1)$，解得 $x>\dfrac{3-\lg 2-3.5\lg e}{0.2+0.4\lg e-0.4\lg 3}\approx 6.64$

故从 2021 年底起经过 7 年后，即 2028 年底新能源汽车的数量将超过传统能源汽车

**11.目前，新冠病毒引起的疫情仍在全球肆虐，在党中央的正确领导下，全国人民团结一心，使我国疫情得到了有效的控制．为了应对最新型的奥密克戎病毒，各大药物企业积极投身到新疫苗的研发中．某药企为评估一款新药的药效和安全性，组织一批志愿者进行临床用药实验，结果显示临床疗效评价指标 $A$ 的数量 $y$ 与连续用药天数 $x$ 具有相关关系．刚开始用药时，指标 $A$ 的数量 $y$ 变化明显，随着天数增加，$y$ 的变化趋缓．根据志愿者的临床试验情况，得到了一组数据 $(x_i,y_i)$，$i=1,2,3,4,5,\cdots,10$，$x_i$ 表示连续用药 $i$ 天，$y_i$ 表示相应的临床疗效评价指标 $A$ 的数值．该药企为了进一步研究药物的临床效果，建立了 $y$ 关于 $x$ 的两个回归模型：**

**模型①：由最小二乘公式可求得 $y$ 与 $x$ 的线性回归方程：$y=2.50x-2.50$；**

**模型②：由样本点的分布，可以认为样本点集中在曲线：$y=\hat{b}\ln x+a$ 的附近，令 $t=\ln x$，则有 $\sum_{i=1}^{10}t_i=22.00$，$\sum_{i=1}^{10}y_i=230$，$\sum_{i=1}^{10}t_iy_i=569.00$，$\sum_{i=1}^{10}t_i^2=50.92$.**

**(1)根据所给的统计量，求模型②中 $y$ 关于 $x$ 的回归方程；**
**(2)根据下列表格中的数据，说明哪个模型的预测值精度更高、更可靠；**
**(3)根据（2）中精确度更高的模型，预测用药一个月后，临床疗效评价指标 $A$ 相对于用药半个月的变化情况（一个月以 30 天计，结果保留两位小数）.**

| 回归模型 | 模型① | 模型② |
|---|---|---|
| 残差平方和 $\sum_{i=1}^{10}\left(y_i-y_i\right)^2$ | **102.28** | **36.19** |

**参考数据：$\ln 2\approx 0.6931$.**

(1)解：由题意，知 $\sum_{i=1}^{10}t_i=22.00$，$\sum_{i=1}^{10}y_i=230$，所以 $\bar{t}=2.20$，$\bar{y}=23$，

又由 $\hat{b}=\dfrac{\sum_{i=1}^{n}(t_i-\bar{t})(y-\bar{y})}{\sum_{t=1}^{n}(t_i-\bar{t})^2}=\dfrac{\sum_{i=1}^{n}t_iy_i-10\bar{t}\cdot\bar{y}}{\sum_{i=1}^{n}t_i^2-10\bar{t}^2}$，$=\dfrac{569.00-10\times 2.20\times 23}{50.92-10\times 2.20\times 2.20}=25$，则 $a=\bar{y}-\hat{b}\bar{t}=23-25\times 2.20=-32$，

所以，模型②中 $y$ 关于 $x$ 的回归方程 $y=25\ln x-32$；

(2)由表格中的数据，可得 $102.28>36.19$，即 $\dfrac{102.28}{\sum_{i=1}^{10}(y_i-\bar{y})^2}>\dfrac{36.19}{\sum_{i=1}^{10}(y_i-\bar{y})^2}$，

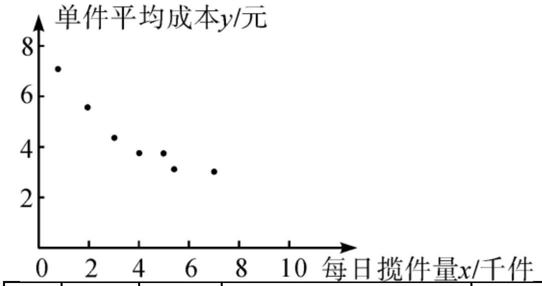
所以模型①的 $R^2$ 小于模型②，说明回归模型②刻画的拟合效果更好；

(3)根据模型②，当连续用药 30 天后，$y_{30}=25\ln 30-32$，连续用药 15 天后，$y_{15}=25\ln 15-32$，

∵ $y_{30}-y_{15}=25\ln 2=17.3275\approx 17.33$，

∴用药一个月后，疗效评价指标相对于用药半个月提高 17.33.

**12.**快递业的迅速发展导致行业内竞争日趋激烈．某快递网点需了解一天中收发一件快递的平均成本 $y$（单位：元）与当天揽收的快递件数即揽件量 $x$（单位：千件）之间的关系，对该网点近 7 天的每日揽件量 $x_i$（单位：千件）与当日收发一件快递的平均成本 $y_i$（单位：元）（$i=1,2,3,4,5,6,7$）的数据进行了初步处理，得到散点图及一些统计量的值．




| $\overline{x}$ | $\overline{y}$ | $\overline{w}$ | $\sum_{i=1}^{7}(x_i-\overline{x})(y_i-\overline{y})$ | $\sum_{i=1}^{7}(w_i-\overline{w})(y_i-\overline{y})$ | $\sum_{i=1}^{7}(x_i-\overline{x})^2$ | $\sum_{i=1}^{7}(w_i-\overline{w})^2$ |
|---|---|---|---|---|---|---|
| 4 | 4.6 | 0.37 | –18 | 2.75 | 25.5 | 0.55 |

表中 $w_i=\frac{1}{x_i}$， $\overline{w}=\frac{1}{7}\sum_{i=1}^{7}w_i$ .

(1)根据散点图判断 $y=ax+b$ 与 $y=c+\frac{d}{x}$ 哪一个更适宜作为 $y$ 关于 $x$ 的经验回归方程类型？并根据判断结果及表中数据求出 $y$ 关于 $x$ 的经验回归方程；

(2)已知该网点每天的揽件量 $x$（单位：千件）与单件快递的平均价格 $t$（单位：元）之间的关系是 $x=\sqrt{59-4t}\,(5.75\le t\le 14.5)$，收发一件快递的利润等于单件的平均价格减去平均成本，根据（1）中建立的经验回归方程解决以下问题：

①预测该网点某天揽件量为 2 千件时可获得的总利润；

②单件快递的平均价格 $t$ 为何值时，该网点一天内收发快递所获利润的预报值最大？

附：对于一组具有线性相关关系的数据 $(\mu_i,v_i)(i=1,2,\cdots,n)$，其经验回归直线 $\hat{v}=\hat{\beta}\mu+\hat{\alpha}$ 的斜率和截距的最小二乘估计分别为 $\hat{\beta}=\frac{\sum_{i=1}^{n}(\mu_i-\overline{\mu})(v_i-\overline{v})}{\sum_{i=1}^{n}(\mu_i-\overline{\mu})^2}$， $\hat{\alpha}=\overline{v}-\hat{\beta}\overline{\mu}$ .

【答案】(1) $y=c+\frac{d}{x}$ 更适宜作为 $y$ 关于 $x$ 的经验回归方程类型， $y=2.75+\frac{5}{x}$

(2)①17000 元；②单件快递的平均价格 $t=10.75$ 元时，该网点一天内收发快递所获利润的预报值最大.

(1)由散点图可知： $y=c+\frac{d}{x}$ 更适宜作为 $y$ 关于 $x$ 的经验回归方程类型；

令 $w=\frac{1}{x}$， 则 $d=\frac{\sum_{i=1}^{7}(w_i-\overline{w})(y_i-\overline{y})}{\sum_{i=1}^{7}(w_i-\overline{w})^2}=\frac{2.75}{0.55}=5$， $c=\overline{y}-d\overline{w}=4.6-5\times0.37=2.75$，

$\therefore$ $y$ 关于 $x$ 的经验回归方程为： $y=2.75+\frac{5}{x}$.

(2)设收发 $x$ 千件快递获利 $z$ 千元，则 $z=(t-y)x=\left(\frac{59-x^2}{4}-\frac{5}{x}-2.75\right)x=-\frac{x^3}{4}+12x-5\,(1\le x\le 6)$；

①当 $x=2$ 时， $z=17$，即该网点某天揽收 2000 件快递可获得的总利润约为17000 元.

②$\because z'=-\frac{3}{4}x^2+12$，令 $z'=0$，解得： $x=4$，

$\therefore$ 当 $x\in[1,4)$ 时， $z'>0$；当 $x\in(4,6]$ 时， $z'<0$；

$\therefore z$ 在 $[1,4)$ 上单调递增，在 $(4,6]$ 上单调递减，

$\therefore$ 当 $x=4$ 时， $z_{\max}=27$，此时 $t=10.75$；

$\therefore$ 单件快递的平均价格 $t=10.75$ 元时，该网点一天内收发快递所获利润的预报值最大.

13.某制造企业从生产的产品中随机抽查了 1000 件，经检验，其中一等品有 800 件，二等品有 150 件，次品有 50 件.若销售 1 件该产品，一等品的利润为 200 元，二等品的利润为 100 元，次品直接销毁，亏损 200 元.

(1)用样本估计总体，估计该制造企业随机销售 1 件产品的利润的期望值.

(2)根据统计，该制造企业在 2021 年 12 月至 2022 年 5 月的产量 $y$（万件）与月份编号（记 2021 年 12 月，2022 年 1 月，$\cdots$ 编号分别为 $1,2,\cdots$）近似满足关系式 $y=b\cdot x^a(a>0,b>0)$，相关统计量的值如下：

$\sum_{6}^{i=1}\ln x_i=6.60,\sum_{6}^{i=1}\ln y_i=-2.70,\sum_{6}^{i=1}\left(\ln x_i\right)^2=9.46,\sum_{6}^{i=1}\left(\ln x_i\cdot\ln y_i\right)=-1.87,e\approx 2.7$.根据所给的统计量，求$y$关于$x$的回归方程，并估计该制造企业2022年8月份的利润为多少万元.（结果精确到$0.01$）

附：对于一组数据$\left(u_i,v_i\right)\left(i=1,2,3,\cdots,n\right)$，其回归直线$\hat{v}=\beta u+\alpha$的斜率和截距的最小二乘估计分别为

$$\hat{\beta}=\frac{\sum_{i=1}^{n}u_iv_i-n\overline{u}\overline{v}}{\sum_{i=1}^{n}u_i^2-n\overline{u}^2},\alpha=\overline{v}-\hat{\beta}\overline{u}$$

【答案】(1)165元

(2) $y=\frac{\sqrt{x}}{e}$，估计该制造企业2022年8月份的利润为$183.15$万元

(1)因为该制造企业生产的产品中一等品、二等品和次品的频率分别为$\frac{4}{5},\frac{3}{20},\frac{1}{20}$，

所以该制造企业随机销售1件产品的利润$X$的分布列为

| $X$ | 200 | 100 | $-200$ |
|---|---|---|---|
| $P$ | $\frac{4}{5}$ | $\frac{3}{20}$ | $\frac{1}{20}$ |

所以$E\left(X\right)=200\times\frac{4}{5}+100\times\frac{3}{20}-200\times\frac{1}{20}=165$，

即估计该制造企业随机销售1件产品的利润的期望值为165元.

(2)因为$y=b\cdot x^a(a>0,b>0)$，所以$\ln y=\ln\left(b\cdot x^a\right)=\ln b+a\ln x$.

令$v=\ln y,u=\ln x$，则$v=\ln b+au$.

因为$\sum_{i=1}^{6}\left(\ln x_i\right)=\sum_{i=1}^{6}u_i=6.60,\sum_{i=1}^{6}\left(\ln y_i\right)=\sum_{i=1}^{6}v_i=-2.70,\sum_{i=1}^{6}\left(\ln x_i\right)^2=\sum_{i=1}^{6}u_i^2=9.46$，

$\sum_{i=1}^{6}\left(\ln x_i\cdot\ln y_i\right)=\sum_{i=1}^{6}\left(u_i\cdot v_i\right)=-1.87=-1.87,$

所以$\overline{u}=1.10,\overline{v}=-0.45$，

$$\hat{a}=\frac{\sum_{i=1}^{6}u_iv_i-6\overline{u}\overline{v}}{\sum_{i=1}^{6}u_i^2-6\overline{u}^2}=\frac{-1.87-6\times1.10\times(-0.45)}{9.46-6\times1.10^2}=\frac{1}{2}$$

因为$\ln b=\overline{v}-\hat{a}\overline{u}=-0.45-\frac{1}{2}\times1.10=-1$，所以$b=\frac{1}{e}$，所以回归方程为$y=\frac{\sqrt{x}}{e}$.

当$x=9$时，$\hat{y}=\frac{3}{e}\approx1.11$，

故估计该制造企业2022年8月份的利润为$1.11\times165=183.15$万元.