第八章 成对数据的统计分析

8.1 成对数据的统计相关性

8.1.1 变量的相关关系+8.1.2 样本相关系数

对点上分

1. C 【解析】考生考试的座位号只是确定考生坐的位置,与考试成绩无关,则①**错误**;
勤能补拙具有相关关系,水稻产量与气候具有相关关系,则②③**正确**;
正方形的边长与正方形的面积是函数关系,则④**错误**.
故选 C.

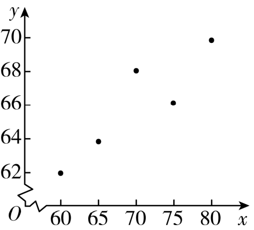
易错警示 区分函数关系与相关关系

函数关系是一种确定性的关系,意味着自变量取特定值时,因变量与之对应的数值是确定的,它可以由数学表达式精确地表示出来;相关关系则是指变量之间存在的一种非确定性的相互依存关系,在这种关系中,一个变量的取值变化时,另一个变量由于受随机因素影响,与其所对应的数值是非确定性的. 简单来说,函数关系是一种严格的一一对应关系,而相关关系则是一种松散的关联,变量之间的变化并不是唯一确定的.

2. **攻略上分** 根据散点图判断变量间的相关关系,先画出散点图,根据散点的分布情况判断是否有线性相关关系,再根据横坐标变化时纵坐标的变化确定是正相关还是负相关.

【解】涉及两个变量,数学成绩与物理成绩,可以以数学成绩的变化考查物理成绩的变化趋势.
用 x 轴表示数学成绩,y 轴表示物理成绩,建立直角坐标系,可得相应的散点

图，如图所示.



由散点图可知，两者之间具有线性相关关系. 且随着 x 增大，y 大体上也增大，故这两个变量正相关.

方法总结 如何判断两个变量间的相关性

(1) 若给出数据，可通过画散点图进行判断，散点呈现递增趋势则正相关，散点呈现递减趋势则负相关；

(2) 若给出样本相关系数，则可根据样本相关系数 r 的正负判断，r 是正数，则正相关，r 是负数，则负相关.

3. A 【解析】观察选项 A 的散点图，这些点紧密地聚集在一条斜向下的直线附近，其样本相关系数接近于−1；选项 B 的散点图中，线性负相关程度不及 A，比较分散，即样本相关系数要比选项 A 的大.

提示：负相关程度越强，样本相关系数越小

选项 C 的散点图中，散点呈现出一定的上升趋势，变量 x 和 y 之间具有较强的正线性相关关系，其样本相关系数为正数，比选项 A 的大.

提示：正相关程度越强，样本相关系数越大

选项 D 的散点图中，散点比较分散，大致是下降的趋势，线性相关程度比选项 A 要弱，样本相关系数比选项 A 的大.

综合比较四个选项，选项 A 线性负相关程度最强，所以样本相关系数最小.

故选 A.

4. C

攻略上分 给出样本相关系数 r 的值，解读 r 的符号，即解读型.

【解析】因为两种生物近期的数量的样

本相关系数为 0.75，所以两种生物的数量正相关，故两种生物的数量增减性有相同的趋势，C 正确.

因为相关关系是一种非确定关系，体现的是一种趋势，所以不能保证一种生物数量增长时，另一种生物数量也一定增长，B 错误. 故选 C.

5. 0.99

攻略上分 可结合题目中给出的数据计算样本相关系数 r.

【解析】由已知可得，$\bar{x}=\dfrac{7+8+9+10+11+12}{6}=9.5$，

$\bar{y}=\dfrac{11+12+14+15+18+20}{6}=15$，

则 $\sum\limits_{i=1}^{6}(x_i-\bar{x})^2=(7-9.5)^2+(8-9.5)^2+(9-9.5)^2+(10-9.5)^2+(11-9.5)^2+(12-9.5)^2=\dfrac{35}{2}$，

$\sum\limits_{i=1}^{6}(y_i-\bar{y})^2=(11-15)^2+(12-15)^2+(14-15)^2+(15-15)^2+(18-15)^2+(20-15)^2=60$，

所以 $r=\dfrac{\sum\limits_{i=1}^{6}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum\limits_{i=1}^{6}(x_i-\bar{x})^2}\sqrt{\sum\limits_{i=1}^{6}(y_i-\bar{y})^2}}=\dfrac{32}{\sqrt{\dfrac{35}{2}}\times\sqrt{60}}=\dfrac{32}{5\sqrt{42}}\approx\dfrac{32}{5\times6.48}\approx0.99$.

8.2 一元线性回归模型及其应用

8.2.1 一元线性回归模型+
8.2.2 一元线性回归模型参数的最小二乘估计

对点上分

1. ABC 【解析】由题表中数据可得 $\bar{x}=\dfrac{2+3+4+5+6}{5}=4$，$\bar{y}=\dfrac{2.2+3.8+5.5+6.5+7}{5}=5$，故样本点中心为 $(4,5)$，故 B 正确；

提示：经验回归直线恒过样本点中心 $(\bar{x},\bar{y})$

由于经验回归方程为 $\hat{y}=1.23x+\hat{a}$，斜率为正数，故样本相关系数 $r>0$，故 A 正确；

将$(4,5)$代入$\hat{y}=1.23x+\hat{a}$可得$\hat{a}=0.08$,**故C正确**;

当$x=10$时,$\hat{y}=1.23\times10+0.08=12.38$,故当该型机床已投入生产的时间为10年时,预计当年所需要支出的维修费用为12.38万元,**故D错误**.

提示:由经验回归方程所得$\hat{y}$为估计值

2. $\hat{y}=0.7x-2.3$　9　【解析】设y关于x的经验回归方程为$\hat{y}=\hat{b}x+\hat{a}$,直线过样本点中心$(\bar{x},\bar{y})$,

由表格数据得$\bar{x}=\dfrac{6+8+10+12}{4}=9$,$\bar{y}=\dfrac{2+3+5+6}{4}=4$,

$\sum\limits_{i=1}^{4}x_iy_i=6\times2+8\times3+10\times5+12\times6=158$,$\sum\limits_{i=1}^{4}x_i^2=6^2+8^2+10^2+12^2=344$,

故根据最小二乘法知$\hat{b}=\dfrac{\sum\limits_{i=1}^{4}x_iy_i-4\bar{x}\bar{y}}{\sum\limits_{i=1}^{4}x_i^2-4\bar{x}^2}=\dfrac{158-4\times9\times4}{344-4\times9^2}=0.7$,

所以$\hat{a}=\bar{y}-\hat{b}\bar{x}=4-0.7\times9=-2.3$,即经验回归方程为$\hat{y}=0.7x-2.3$.

将$y=4$代入方程,得$x=9$,即可预测判断力为4的学生的记忆力为9.

提示:利用经验回归方程进行预测时直接代入求值即可

3. (1)$\hat{y}=0.65x+56.5$　(2)95.5

【解析】(1)由题表数据得$\bar{x}=\dfrac{1}{5}\times(10+20+30+40+50)=30$,$\bar{y}=\dfrac{1}{5}\times(64+69+75+82+90)=76$,

因为$\sum\limits_{i=1}^{5}x_iy_i=12\ 050$,$\sum\limits_{i=1}^{5}x_i^2=5\ 500$,

所以$\hat{b}=\dfrac{\sum\limits_{i=1}^{5}x_iy_i-5\bar{x}\cdot\bar{y}}{\sum\limits_{i=1}^{5}x_i^2-5\bar{x}^2}=\dfrac{12\ 050-5\times30\times76}{5\ 500-5\times900}=0.65$,则$\hat{a}=\bar{y}-\hat{b}\bar{x}=76-0.65\times30=56.5$,

所以y关于x的经验回归方程为$\hat{y}=0.65x+56.5$.

(2)当$x=60$时,$\hat{y}=0.65\times60+56.5=95.5$,因此可以预测制作60个这种模型需要花费95.5分钟.

4. C　【解析】对于A选项,由样本数据

得到的经验回归直线 $\hat{y}=\hat{b}x+\hat{a}$ 必过样本点的中心 $(\bar{x},\bar{y})$，**A 正确**；

对于 B 选项，残差平方和越小的模型，拟合的效果越好，**B 正确**；

对于 C 选项，用决定系数 R^2 来刻画回归效果，R^2 的值越小，说明残差平方和越大，模型的拟合效果越差，**C 错误**；

对于 D 选项，若变量 y 和 x 之间的样本相关系数 $r=-0.9362$，$|r|$ 比较接近 1，则变量 y 与 x 之间具有线性相关关系，**D 正确. 故选 C.**

易错警示 **混淆决定系数与残差平方和对拟合效果的影响**

(1) 残差平方和越小，说明拟合效果越好；残差平方和越大，说明拟合效果越差.

(2) 决定系数 R^2 越大，越接近 1，表示模型的拟合效果越好，相反，决定系数 R^2 越小，表示拟合效果越差.

(3) 刻画样本回归效果的决定系数 $R^2=1-\dfrac{\sum\limits_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum\limits_{i=1}^{n}(y_i-\bar{y})^2}$，残差 $\hat{e}_i=y_i-\hat{y}_i$.

5. D 【解析】从题图中可以看出 $D(10,2)$ 相较其他点比较偏离，故去掉 $D(10,2)$ 后，拟合效果更好.

对于 A，样本相关系数 $|r|$ 越接近于 1，模型的拟合效果越好，又散点图整体是上升的趋势，所以 $r>0$，故去掉 $D(10,2)$ 后，样本相关系数 r 变大，**故 A 错误**；

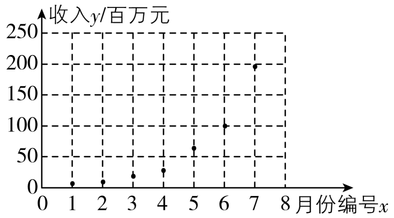
对于 B，决定系数 R^2 越大，模型的拟合效果越好，故去掉 $D(10,2)$ 后，决定系数 R^2 变大，**故 B 错误**；

对于 C，残差平方和越小，模型的拟合效果越好，故去掉 $D(10,2)$ 后，残差平方和变小，**故 C 错误**；

对于 D，若去掉 $D(10,2)$ 后，变量 y 与变量 x 的相关性变强，且是正相关，**故 D 正确. 故选 D.**

6. **攻略上分** 根据题表画出散点图，数据点呈现指数增长趋势，所以变量非线性相关，故 $y=c\cdot d^x$ 更适合作为回归模型.

【解】(1)散点图如图所示，



根据散点图判断，$y=c\cdot d^x$ 适合作为该团队经济收入 y 关于月份编号 x 的经验回归方程模型.

$\because y=c\cdot d^x$，两边同时取常用对数得，

$\lg y=\lg(c\cdot d^x)=\lg c+x\cdot\lg d$，

设 $\lg y=v$，$\therefore v=\lg c+x\cdot\lg d$，

提示：由于参考数据给出了 $v=\lg y$，提醒我们引出对数，于是两边同时取对数

$\because \bar{x}=\frac{1}{7}\times(1+2+3+4+5+6+7)=4$，

$\bar{v}=\frac{1}{7}\sum_{i=1}^{7}v_i\approx\frac{1}{7}\times 10.79\approx 1.54$，

$\sum_{i=1}^{7}x_i^2=1^2+2^2+3^2+4^2+5^2+6^2+7^2=140$，

$\therefore \lg\hat{d}=\frac{\sum_{i=1}^{7}x_iv_i-7\bar{x}\bar{v}}{\sum_{i=1}^{7}x_i^2-7\bar{x}^2}\approx\frac{50.12-7\times4\times1.54}{140-7\times4^2}=\frac{7}{28}=0.25$，

把样本点中心(4，1.54)代入 $\hat{v}=\lg\hat{c}+\hat{d}\cdot\lg x$，得 $1.54=\lg\hat{c}+0.25\times4$，

$\therefore \lg\hat{c}=0.54$，$\therefore \hat{v}=0.54+0.25x$，

$\therefore \lg\hat{y}=0.54+0.25x$，

$\therefore y$ 关于 x 的经验回归方程为 $\hat{y}=10^{0.54+0.25x}=3.47\times10^{0.25x}$.

(2)当 $x=8$ 时，$\hat{y}=3.47\times10^{0.25\times8}=347$，所以预测该团队下一个月的经济收入为 347 百万元.

(3)不合理，经验回归方程一般具有时效性，解释变量越接近样本数据，预测值越可信，否则会有显著误差.

7.【解】由题中散点图判断 $y=c\ln(x-2\,012)+d$ 更适合作为该机场飞往 A 地航班放行准点率 y 关于年份数 x 的经验回归方程模型.

令 $t=\ln(x-2\,012)$，先建立 y 关于 t 的线性回归方程.

由于 $\hat{c}=\frac{\sum_{i=1}^{10}t_iy_i-10\bar{t}\bar{y}}{\sum_{i=1}^{10}t_i^2-10\bar{t}^2}=$

$\frac{1\ 226.8-10\times1.5\times80.4}{27.7-10\times1.5^2}=4$,

则 $\hat{d}=\overline{y}-\hat{c}\overline{t}=80.4-4\times1.5=74.4$,

所以该机场飞往 A 地航班放行准点率 y 关于 t 的线性回归方程为 $\hat{y}=4t+74.4$,

因此 y 关于年份数 x 的经验回归方程为 $\hat{y}=4\ln(x-2\ 012)+74.4$.

所以当 $x=2\ 025$ 时,该机场飞往 A 地航班放行准点率 y 的预测值为 $\hat{y}=4\ln(2\ 025-2\ 012)+74.4=4\ln 13+74.4\approx4\times2.56+74.4=84.64$.

所以 2025 年该机场飞往 A 地航班放行准点率 y 的预测值为 84.64%.

8.1+8.2 节测上分

1. C 【解析】对于 A,某商品的销售价格与销售量呈负相关,**故 A 错误**;

对于 B,汽车匀速行驶时的路程与时间是函数关系,**故 B 错误**;

对于 C,气温与冷饮的销售量呈正相关,**故 C 正确**;

对于 D,人的年龄与视力呈负相关,**故 D 错误. 故选 C.**

2. A 【解析】设变量间的线性相关系数为 r,当 $|r|$ 越接近 1 时,相关程度越强,因为 $|-0.92|>|0.85|>|0.79|>|0.46|$,所以甲组数据变量间的线性相关程度最强. **故选 A.**

3. B 【解析】对于 A,各点分布没有明显相关性,不符;

对于 B,各点分布在一条直线附近,且有负相关性,符合;

对于 C,各点分布在一条抛物线附近,变量之间先呈正相关,后呈负相关,不符;

对于 D,各点分布在一条直线附近,且有正相关性,不符. **故选 B.**

4. C 【解析】由题图可知,曲线 2 更贴近散点的趋势,拟合效果更好,故 $0<R_1^2<R_2^2<1$. **故选 C.**

5. A 【解析】$\overline{x}=\frac{0+1+4+5+6+8}{6}=4$,因为样本点中心必在经验回归直线 $\hat{y}=1.03x+1.13$ 上,所以将 $\overline{x}=4$ 代入回归方程中,可得 $\overline{y}=5.25$,设污损数据的值为 m,由 $5.25=\frac{1.3+1.8+5.6+m+7.4+9.3}{6}$,解

得 $m=6.1$，即该数据为 6.1，**故选 A.**

6. ABD

思路导引 求得数据的样本点中心，即可判断 B；结合经验回归直线过点 $(\bar{x},\bar{y})$ 求出 $\hat{a}=1.5$，可判断 A；根据回归直线方程的实际意义可判断 C；将 $x=10$ 代入回归直线方程求得预测值，可判断 D.

【解析】对 A，B，由题意得 $\bar{x}=\frac{1}{5}\times(3+4+5+6+7)=5$，因为 $\sum_{i=1}^{5}y_i=105$，所以 $\bar{y}=21$，则该经验回归直线经过样本点的中心 $(5,21)$，**B 正确.**

由 $21=3.9\times5+\hat{a}$，得 $\hat{a}=1.5$，**A 正确.**

对 C，预测该商品宣传投入费用每增加 1 万元，对应利润相应增加 3.9 万元，**C 错误.**

对 D，当 $x=10$ 时，$\hat{y}=3.9\times10+1.5=40.5$，**D 正确. 故选 ABD.**

7. $(9,5)$ 【解析】由 $\bar{x}=8.2$ 可知 $\bar{y}=8.2-3=5.2$，即 $\sum_{i=1}^{10}x_i=82$，$\sum_{i=1}^{10}y_i=52$.

剔除 $(1,7)$ 后，$\bar{x}'=\frac{82-1}{9}=9$，$\bar{y}'=\frac{52-7}{9}=5$，

因为新样本点中心 $(\bar{x}',\bar{y}')$ 一定在新的经验回归直线上，故其必过点 $(9,5)$.

8. $\frac{83}{28}$ 【解析】$\bar{x}=\frac{3+4+5+6+6+7+8+9}{8}=6$，

$\sum_{i=1}^{8}(x_i-\bar{x})^2=(3-6)^2+(4-6)^2+(5-6)^2+2\times(6-6)^2+(7-6)^2+(8-6)^2+(9-6)^2=28$，

由 $r=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2\sum_{i=1}^{n}(y_i-\bar{y})^2}}=\frac{\sum_{i=1}^{8}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{28\times252}}=\frac{83}{84}$，

得 $\sum_{i=1}^{8}(x_i-\bar{x})(y_i-\bar{y})=83$，

所以 $\hat{b}=\frac{\sum_{i=1}^{8}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{8}(x_i-\bar{x})^2}=\frac{83}{28}$.

9. $\hat{y}=-\frac{5}{2}x+44$　【解析】由题干中的数据可得 $\bar{x}=\frac{1}{3}\sum_{i=1}^{3}x_i=8$，$\bar{y}=\frac{1}{3}\sum_{i=1}^{3}y_i=24$，

则 $\hat{b}=\frac{\sum_{i=1}^{3}x_iy_i-3\bar{x}\cdot\bar{y}}{\sum_{i=1}^{3}x_i^2-3\bar{x}^2}=\frac{556-3\times8\times24}{200-3\times8^2}=\frac{-20}{8}=-\frac{5}{2}$，则 $\hat{a}=\bar{y}-\hat{b}\bar{x}=24+\frac{5}{2}\times8=44$.

因此 y 关于 x 的经验线性回归方程是 $\hat{y}=-\frac{5}{2}x+44$.

10.【解】(1)由题表可得，$\bar{x}=\frac{1}{5}\times(1+2+3+4+5)=3$，$\bar{y}=\frac{1}{5}\times(59+61+64+68+73)=65$.

因为 $t=x^2$，则 $y=bt+a$，

所以 $\bar{t}=\frac{1}{5}\times(1^2+2^2+3^2+4^2+5^2)=11$，

则 $\hat{b}=\frac{\sum_{i=1}^{5}(t_i-\bar{t})(y_i-\bar{y})}{\sum_{i=1}^{5}(t_i-\bar{t})^2}=\frac{217}{374}\approx0.6$，$\hat{a}=\bar{y}-\hat{b}\bar{t}\approx65-0.6\times11=58.4$.

所以经验回归方程为 $\hat{y}=0.6x^2+58.4$.

(2)将 x 的值分别代入 $\hat{y}=0.6x^2+58.4$ 中可得对应 y 的估计值分别为 59，60.8，63.8，68，73.4，

则残差平方和为 $(59-59)^2+(61-60.8)^2+(64-63.8)^2+(68-68)^2+(73-73.4)^2=0.24$.

提示：残差 $\hat{e}_i=y_i-\hat{y}_i$，即用观测值减去预测值

因为 $0.24<0.5$，所以经验回归方程 $\hat{y}=0.6x^2+58.4$ 的拟合效果符合要求.

8.3　列联表与独立性检验

8.3.1　分类变量与列联表+

8.3.2　独立性检验

对点上分

1. BC　【解析】散点图适用于推断两个数值型变量之间是否有相关关系，故

A 错误；

等高堆积条形图和列联表适用于推断两个分类变量之间是否有关联，故 B，C 正确；

残差图用来体现观测值与预测值之间的差距，故 D 错误. 故选 BC.

2. C 【解析】根据题意，设样本中 35 岁以上的市民人数为 x，35 岁及以下的市民人数为 y，男性市民的人数为 a，女性市民的人数为 b，得到如下两个 2×2 列联表.

性别	年龄		总计
	35 岁以上	35 岁及以下	
男性	$0.56x$	$0.52y$	$0.56x+0.52y$
女性	$0.44x$	$0.48y$	$0.44x+0.48y$
总计	x	y	$x+y$

性别	年龄		总计
	35 岁以上	35 岁及以下	
男性	$0.57a$	$0.43a$	a
女性	$0.53b$	$0.47b$	b
总计	$0.57a+0.53b$	$0.43a+0.47b$	$a+b$

根据第一个列联表可知，样本中男性市民人数为 $0.56x+0.52y$，女性市民人数为 $0.44x+0.48y$，又 $0.56x+0.52y>0.44x+0.48y$，即样本中男性比女性多，故 A 正确；

根据第二个列联表可知，样本中 35 岁以上女性市民人数为 $0.53b$，35 岁及以下女性市民人数为 $0.47b$，又 $0.53b>0.47b$，即样本中多数女性是 35 岁以上，故 B 正确；

由题意，$\frac{0.52y}{0.56x+0.52y}=0.43$，则 $x>1.23y$，所以 $0.52y<0.44x$，故 C 不正确；

根据第二个列联表可知，样本中 35 岁以上的市民人数为 $0.57a+0.53b$，35 岁及以下的市民人数为 $0.43a+0.47b$，

又 $0.57a+0.53b>0.43a+0.47b$，即样本中 35 岁以上的市民比 35 岁及以下的市民多，**故 D 正确. 故选 C.**

3. C 【解析】根据题图①可知样本中选择物理学科的人数较多，**故 C 正确**；

根据题图②可知样本中男生人数多于女生人数，**故 D 错误**；

样本中选择物理学科的人数多于选择历史学科的人数，而选择物理学科的男生比例高，选择历史学科的女生比例低，所以样本中选择物理学科的男生人数多于选择历史学科的女生人数，**故 A 错误**；

根据题图②可知样本中女生选择历史学科的人数少于男生选择历史学科的人数，**故 B 错误. 故选 C.**

4. B 【解析】因为 $P(\chi^2\geqslant 6.635)=0.01$，$\chi^2\approx 4.444<6.635$，所以根据小概率值 $\alpha=0.01$ 的独立性检验，认为“进入不同颜色的蜂蜡罐与蜜蜂种类无关”，**A 错误**；因为 $P(\chi^2\geqslant 3.841)=0.05$，$\chi^2\approx 4.444>3.841$，所以根据小概率值 $\alpha=0.05$ 的独立性检验，认为“进入不同颜色的蜂蜡罐与蜜蜂种类有关”，**B 正确**；

因为 $P(\chi^2\geqslant 7.879)=0.005$，$\chi^2\approx 4.444<7.879$，所以根据小概率值 $\alpha=0.005$ 的独立性检验，认为“进入不同颜色的蜂蜡罐与蜜蜂种类无关”，**C 错误**；

因为 $P(\chi^2\geqslant 2.706)=0.1$，$\chi^2\approx 4.444>2.706$，所以根据小概率值 $\alpha=0.1$ 的独立性检验，认为“进入不同颜色的蜂蜡罐与蜜蜂种类有关”，**D 错误. 故选 B.**

5. A 【解析】根据题意，得 χ^2 的观测值在区间 $[3.841,6.635)$ 内，所以 χ^2 的观测值不可能为 3.622. **故选 A.**

6. D

思路导引 设男生的人数为 $5m(m\in\mathbf{N}^*)$，利用独立性检验得到关于 m 的不等式，从而得解.

【解析】由题意可设男生的人数为 $5m$（$m\in\mathbf{N}^*$），则女生的人数为 $10m$，根据题意可列出如下的 2×2 列联表：

喜好	性别		合计
	男生	女生	
喜欢吃甜食	$2m$	$8m$	$10m$
不喜欢吃甜食	$3m$	$2m$	$5m$
合计	$5m$	$10m$	$15m$

$$\chi^2=\frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}=\frac{15m\cdot(2m\cdot2m-8m\cdot3m)^2}{10m\cdot5m\cdot5m\cdot10m}=2.4m,$$

因为根据小概率值 $\alpha=0.01$ 的独立性检验，推断出“学生性别和是否喜欢吃甜食”有关，所以 $2.4m\geqslant6.635$，

提示：当认为两个变量有关联时 $\chi^2\geqslant x_\alpha$

解得 $m>2.7645$，所以 $5m>13.8225$，

因为 $m\in\mathbf{N}^*$，所以 $5m$ 的最小值为 15.

故选 D.

7. **攻略上分** 根据题目已知条件补充列联表，按照攻略中独立性检验的步骤即可求得结果.

【解】(1) 根据题意可得如下 2×2 列联表.

语文成绩	数学成绩		合计
	优秀	不优秀	
优秀	50	40	90
不优秀	75	35	110
合计	125	75	200

(2) 零假设为 H_0：高中学生的语文成绩与数学成绩无关.

提示：提出零假设

由列联表得 $\chi^2=$

$$\frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}=\frac{200\times(50\times35-40\times75)^2}{90\times110\times125\times75}\approx3.367>2.706=x_{0.1},$$

提示：计算 χ^2 的值并与临界值进行比较

所以根据小概率值 $\alpha=0.1$ 的独立性检验，推断 H_0 不成立，即认为高中学生的语文成绩与数学成绩有关，此推

断犯错误的概率不大于 0. 1.

提示: 得出结论

8.【解】(1) 根据题意,得到列联表如下.

平衡力	心脏病		合计
	未患心脏病	患心脏病	
平衡力好	900	100	1 000
平衡力差	850	150	1 000
合计	1 750	250	2 000

(2) 零假设为 H_0: 平衡力的好坏与心脏病风险没有关联.

根据列联表中的数据,经计算得到 $\chi^2=\frac{2\ 000\times(900\times150-100\times850)^2}{1\ 000\times1\ 000\times1\ 750\times250}=\frac{80}{7}\approx11.429>10.828=x_{0.001}$.

根据小概率值 $\alpha=0.001$ 的独立性检验,推断 H_0 不成立,即认为平衡力的好坏与心脏病风险有关联,此推断犯错误的概率不大于 0. 001.

8. 3 节测上分

1. C 【解析】由题意得 χ^2 的值应位于 6. 635 与 10. 828 之间,**故 C 正确. 故选 C.**

2. A 【解析】零假设为 H_0: 爱好跳绳与性别无关,因为 $\chi^2=\frac{110\times(40\times30-20\times20)^2}{60\times50\times60\times50}=\frac{352}{45}\approx7.822$,**故 A 正确;**

因为 $7.822<10.828$,所以没有充分证据推断 H_0 不成立,即根据小概率值 $\alpha=0.001$ 的独立性检验,认为爱好跳绳与性别无关,**故 B 错误;**

又因为 $7.822>6.635$,所以根据小概率值 $\alpha=0.01$ 的独立性检验,推断 H_0 不成立,即认为爱好跳绳与性别有关,**故 C 错误;**

所以在犯错误的概率不超过 0. 01 的前提下,认为爱好跳绳与性别有关,**故 D 错误. 故选 A.**

3. C 【解析】由题意,2×2 列联表如下:

	去年体检人数	去年未体检人数	合计
高中教师	70	30	100
初中教师	80	20	100
合计	150	50	200

$\chi^2=\dfrac{200\times(1\ 400-2\ 400)^2}{100\times100\times150\times50}=\dfrac{8}{3}$. 故选 C.

4. 30 【解析】设男生人数为 x,由题意得列联表如下:

性别	是否喜欢追星		合计
	喜欢追星	不喜欢追星	
男生	$\frac{1}{3}x$	$\frac{2}{3}x$	x
女生	$\frac{1}{3}x$	$\frac{1}{6}x$	$\frac{1}{2}x$
合计	$\frac{2}{3}x$	$\frac{5}{6}x$	$\frac{3}{2}x$

由列联表中的数据,计算得 $\chi^2=\dfrac{\frac{3}{2}x\cdot\left(\frac{1}{3}x\cdot\frac{1}{6}x-\frac{2}{3}x\cdot\frac{1}{3}x\right)^2}{x\cdot\frac{1}{2}x\cdot\frac{2}{3}x\cdot\frac{5}{6}x}=\dfrac{3}{20}x\geqslant3.841$,

提示: 认为中学生追星与学生性别有关,则 $\chi^2\geqslant x_\alpha=3.841$

解得 $x\geqslant\dfrac{20\times3.841}{3}\approx26$.

又 $\dfrac{1}{2}x,\dfrac{1}{3}x\in\mathbf{N}^*$,故设 $x=6k,k\in\mathbf{N}^*$,

所以 $x_{\min}=30$,

即根据 $\alpha=0.05$ 的独立性检验,认为中学生追星与学生性别有关,则男生至少有 30 人.

5.【解】(1)由题意,2×2 列联表如下:

整理错题	数学成绩		合计
	优秀	非优秀	
每天都整理数学错题	95	5	100
不是每天都整理数学错题	60	40	100
合计	155	45	200

(2)零假设为 H_0:学生的数学成绩优秀与每天都整理数学错题相互独立,即学生的数学成绩优秀与每天都整理数学错题无关.

根据列联表数据，计算可得 $\chi^2=\dfrac{200\times(95\times40-5\times60)^2}{100\times100\times155\times45}=\dfrac{9\ 800}{31\times9}\approx35.125>10.828=x_{0.001}$.

根据小概率值 $\alpha=0.001$ 的独立性检验，推断 H_0 不成立，即认为学生的数学成绩优秀与每天都整理数学错题有关，此推断犯错误的概率不大于 0.001.

6.【解】(1) 因为 $\hat{y}=\hat{b}x+0.95$，且 $\bar{x}=480$，$\bar{y}=0.35$，故 $0.35=\hat{b}\times480+0.95$，故 $\hat{b}=\dfrac{0.35-0.95}{480}=-0.001\ 25$. 当 $x=600$ 时，$\hat{y}=-0.001\ 25\times600+0.95=0.2$.

即此时"AI 生成概率"的得分为 0.2.

(2) 抽取的文本中，该平台文本 AI 生成的篇数约为 $200\times15\%=30$，人类撰写的篇数约为 $200-30=170$，

真实 AI 生成且识别准确的篇数约为 $30\times98\%=29.4\approx29$，

真实人类撰写且识别准确的篇数约为 $170\times96.5\%=164.05\approx164$，

故列联表为：

真实结果	检测结果		总计
	识别为 AI 生成	识别为人类撰写	
真实 AI 生成	29	1	30
真实人类撰写	6	164	170
总计	35	165	200

零假设为 H_0："检测结果"与"真实结果"无差异.

由列联表数据计算得，$\chi^2=\dfrac{200\times(29\times164-1\times6)^2}{30\times170\times35\times165}\approx153.213>6.635=x_{0.01}$，

所以依据小概率值 $\alpha=0.01$ 的独立性检验，可以认为"检测结果"与"真实结果"有差异，此推断犯错误的概率不大于 0.01.

真题上分

1. C 【解析】由题中散点图可知这些散点大致落在一条从左下角到右上角的直线附近，表明随花萼长度的增加，相应的花瓣长度呈增加的趋势，由成对样本数据的分布规律可知两者呈线性相关关系，且为正相关，**故 A，B 错误**，因为花瓣长度关于花萼长度的经验回归方程为$\hat{y}=0.750\ 1x+0.610\ 5$，所以当$x=7$时，$\hat{y}=0.750\ 1\times7+0.610\ 5=5.861\ 2$，C **正确**. 样本具有随机性，样本相关系数会随着样本成对数据的变化而变化，**故 D 错误**. **故选 C.**

2.【解】(1)由题意可知$\sum\limits_{i=1}^{10}x_i=0.6$，$\sum\limits_{i=1}^{10}y_i=3.9$，

所以这10棵树木的根部横截面积和材积量的平均值分别为$\bar{x}=\dfrac{1}{10}\sum\limits_{i=1}^{10}x_i=0.06$，$\bar{y}=\dfrac{1}{10}\sum\limits_{i=1}^{10}y_i=0.39$，所以估计该林区这种树木平均一棵的根部横截面积与平均一棵的材积量分别为$0.06\ m^2$和$0.39\ m^3$.

(2)因为$r=\dfrac{\sum\limits_{i=1}^{10}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum\limits_{i=1}^{10}(x_i-\bar{x})^2\sum\limits_{i=1}^{10}(y_i-\bar{y})^2}}$

$=\dfrac{\sum\limits_{i=1}^{10}x_iy_i-10\bar{x}\,\bar{y}}{\sqrt{\sum\limits_{i=1}^{10}x_i^2-10\bar{x}^2}\cdot\sqrt{\sum\limits_{i=1}^{10}y_i^2-10\bar{y}^2}}$

$=\dfrac{0.247\ 4-10\times0.06\times0.39}{\sqrt{0.038-10\times0.06^2}\times\sqrt{1.615\ 8-10\times0.39^2}}$

$=\dfrac{0.013\ 4}{\sqrt{0.002}\times\sqrt{0.094\ 8}}$

$=\dfrac{0.013\ 4}{\sqrt{1.896\times10^{-4}}}\approx\dfrac{134}{137.7}\approx0.97$，

所以该林区这种树木的根部横截面积与材积量的样本相关系数是0.97.

(3)因为树木的材积量与其根部横截面积近似成正比，所以设$y=kx$(k为常数).

由(1)知该林区这种树木平均一棵的根部横截面积与平均一棵的材积量的估计值分别为0.06和0.39，

所以$0.39=0.06k$，即$k=\dfrac{13}{2}$，所以$y=$

$\frac{13}{2}x$.

又因为 $\sum_{i=1}^{n}x_i=186$，所以 $\sum_{i=1}^{n}y_i=\frac{13}{2}\sum_{i=1}^{n}x_i=\frac{13}{2}\times186=1\ 209$，

故该林区这种树木的总材积量的估计值为 $1\ 209\ \mathrm{m}^3$.

3.【解】(1)根据题表数据可知，超声波检查结果不正常的有 200 人，其中患该疾病的有 180 人，因此估计超声波检查结果不正常者患该疾病的概率 $p=\frac{180}{200}=\frac{9}{10}$.

易错：注意该问所求概率中用到的数据，数据找错，计算也就出错

(2)零假设为 H_0：超声波检查结果与患该疾病无关.

$$\chi^2=\frac{1\ 000\times(20\times20-180\times780)^2}{200\times800\times800\times200}=765.625>10.828.$$

根据小概率值 $\alpha=0.001$ 的独立性检验，超声波检查结果与患该疾病有关，此推断犯错误的概率不大于 0.001.

4.【解】(1)补全列联表如下：

	优级品	非优级品
甲车间	26	24
乙车间	70	30

$K^2=\frac{150\times(26\times30-24\times70)^2}{50\times100\times96\times54}=\frac{75}{16}=4.687\ 5>3.841$，$4.687\ 5<6.635$，所以有 95% 的把握认为甲、乙两车间产品的优级品率存在差异，没有 99% 的把握认为甲、乙两车间产品的优级品率存在差异.

(2)由题中数据知，

$\bar{p}=\frac{96}{150}=0.64$，

$p+1.65\sqrt{\frac{p(1-p)}{n}}=0.5+1.65\cdot\sqrt{\frac{0.5\times0.5}{150}}=0.5+1.65\times\frac{0.5}{\sqrt{150}}\approx0.5+1.65\times\frac{0.5}{12.247}\approx0.57$.

又 $0.64>0.57$，

所以可以认为生产线智能化升级改造后，该工厂产品的优级品率提高了.

5.【解】(i)40 只小白鼠体重的增加量从小到大排序为

7.8	9.2	11.4	12.4	13.2
15.2	15.5	16.5	18.0	18.8
18.8	19.2	19.8	20.2	20.2
21.3	21.6	22.5	22.8	23.2
23.6	23.9	25.1	25.8	26.5
27.5	28.2	30.1	32.3	32.6
34.3	34.8	35.6	35.6	35.8
36.2	36.5	37.3	40.5	43.2

所以 40 只小白鼠体重的增加量的中位数 $m=\dfrac{23.2+23.6}{2}=23.4$.

所以列联表如下：

	$<m$	$\geqslant m$	合计
对照组	6	14	20
试验组	14	6	20
合计	20	20	40

(ii) 因为 $K^2=\dfrac{40\times(6\times6-14\times14)^2}{20\times20\times20\times20}=6.4>3.841$,

所以有 95% 的把握认为小白鼠在高浓度臭氧环境中与在正常环境中体重的增加量有差异.

6. (1)【解】

	不够良好	良好	合计
病例组	40	60	100
对照组	10	90	100
合计	50	150	200

$K^2=\dfrac{200\times(40\times90-60\times10)^2}{100\times100\times50\times150}=24>6.635$,

所以有 99% 的把握认为患该疾病群体与未患该疾病群体的卫生习惯有差异.

(2) (i)【证明】$R=\dfrac{P(B|A)}{P(\overline{B}|A)}\cdot$

$$\frac{P(\overline{B}|\overline{A})}{P(B|\overline{A})}=\frac{\dfrac{P(AB)}{P(A)}}{\dfrac{P(A\overline{B})}{P(A)}}\cdot\frac{\dfrac{P(\overline{A}\,\overline{B})}{P(\overline{A})}}{\dfrac{P(\overline{A}B)}{P(\overline{A})}}=\frac{P(AB)}{P(A\overline{B})}\cdot$$

$$\frac{P(\overline{A}\,\overline{B})}{P(\overline{A}B)}=\frac{\dfrac{P(AB)}{P(B)}}{\dfrac{P(\overline{A}B)}{P(B)}}\cdot\frac{\dfrac{P(\overline{A}\,\overline{B})}{P(\overline{B})}}{\dfrac{P(A\overline{B})}{P(\overline{B})}}=\frac{P(A|B)}{P(\overline{A}|B)}\cdot$$

$$\frac{P(\overline{A}|\overline{B})}{P(A|\overline{B})}$$,所以 $R=\dfrac{P(A|B)}{P(\overline{A}|B)}\cdot\dfrac{P(\overline{A}|\overline{B})}{P(A|\overline{B})}$.

提示:条件概率公式的应用

(ii)【解】由调查数据可得 $P(A|B)=\dfrac{40}{100}=\dfrac{2}{5}$,$P(A|\overline{B})=\dfrac{10}{100}=\dfrac{1}{10}$,

则 $P(\overline{A}|B)=1-P(A|B)=\dfrac{3}{5}$,

$P(\overline{A}|\overline{B})=1-P(A|\overline{B})=\dfrac{9}{10}$,

所以 $R=\dfrac{P(A|B)}{P(\overline{A}|B)}\cdot\dfrac{P(\overline{A}|\overline{B})}{P(A|\overline{B})}=\dfrac{\frac{2}{5}}{\frac{3}{5}}\times\dfrac{\frac{9}{10}}{\frac{1}{10}}=6$.

素养上分

1. AC

思路导引 方案①:根据 $y_i=ax_i+b$,将 $b=9$ 代入判断 A;利用对比度公式可得 $C_{\{y_i\}}=a^2C_{\{x_i\}}$,即可判断 C;方案②:当 $z_i=9\lg(x_i+1)$ 时代入特殊值 $x_i=9$ 判断 B;根据条件得出 $z_i-z_{i+1}=t\ln\left(1-\dfrac{9}{9i+n}\right)$,代入特殊值 $n=1$ 判断 D.

【解析】使用方案①调整,当 $b=9$ 时,$y_i=ax_i+9$ 且 $a>0$,又 $x_i\in[0,9]$,则 $y_i>x_i$,A 正确;

由题意得,$C_{\{x_i\}}=\dfrac{1}{n}\sum\limits_{i=1}^{n}(x_i-x_{i+1})^2$,

则 $C_{\{y_i\}}=\dfrac{a^2}{n}\sum\limits_{i=1}^{n}(x_i-x_{i+1})^2$,

当 $C_{\{x_i\}}<C_{\{y_i\}}$ 时,即 $\dfrac{a^2}{n}>\dfrac{1}{n}$ 且 $n\in\mathbf{N}^*$,

又 $a>0$，所以 $a>1$，C 正确；

使用方案②调整，当 $c=9$ 时，$z_i=9\lg(x_i+1)$，显然若 $x_i=9$，则 $z_i=9$，B 错误；

$z_i=c\lg(x_i+1)=c\cdot\dfrac{\ln(x_i+1)}{\ln 10}$，而 $0<c\leqslant\ln 10$，令 $t=\dfrac{c}{\ln 10}$，则 $t\in(0,1]$，故 $z_i=t\ln(x_i+1)$，

又 $x_i=\dfrac{9(i-1)}{n}(i=1,2,\cdots,n+1)$，则

$z_i=t\ln\left(\dfrac{9i-9+n}{n}\right)$，$z_{i+1}=t\ln\left(\dfrac{9i+n}{n}\right)$，

所以 $z_i-z_{i+1}=t\left[\ln\left(\dfrac{9i-9+n}{n}\right)-\ln\left(\dfrac{9i+n}{n}\right)\right]=t\ln\left(1-\dfrac{9}{9i+n}\right)$，

当 $n=1$ 时，$(z_i-z_{i+1})^2=t^2\ln^2 10$，则 $C_{\{z_i\}}=t^2\ln^2 10$，

此时 $C_{\{x_i\}}=\left(\dfrac{9}{n}\right)^2=81$，显然存在 $C_{\{x_i\}}>C_{\{z_i\}}$，D 错误.

故选 AC.

2.【解】(1) 2×2 列联表如下：

	关注	不关注	合计
男生	55	5	60
女生	20	10	30
合计	75	15	90

$\chi^2=\dfrac{90\times(55\times10-5\times20)^2}{60\times30\times75\times15}=9>6.635$，

所以依据小概率值 $\alpha=0.01$ 的独立性检验，认为该校学生对探月工程的关注与性别有关.

(2) 记这 4 个问题分别为 a,b,c,d，振华答对 a,b,c,d 的事件分别记为 A,B,C,D，记振华按方案一、二晋级的概率分别为 P_1,P_2，

则 $P_1=P(ABC\overline{D})+P(AB\overline{C}D)+$

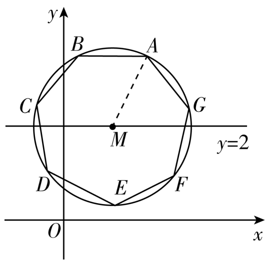
$P(A\bar{B}CD)+P(\bar{A}BCD)+P(ABCD)=$

$\left(\frac{2}{3}\right)^3\times\frac{1}{2}+\left(\frac{2}{3}\right)^2\times\frac{1}{3}\times\frac{1}{2}+\frac{2}{3}\times\frac{1}{3}\times\frac{2}{3}\times\frac{1}{2}+\frac{1}{3}\times\left(\frac{2}{3}\right)^2\times\frac{1}{2}+\left(\frac{2}{3}\right)^3\times\frac{1}{2}=\frac{14}{27}$,

$P_2=\frac{1}{6}P(AB)+\frac{1}{6}P(AC)+\frac{1}{6}\cdot P(AD)+\frac{1}{6}P(BC)+\frac{1}{6}P(BD)+\frac{1}{6}\cdot P(CD)=\frac{1}{6}\left[\left(\frac{2}{3}\right)^2\times3+\frac{2}{3}\times\frac{1}{2}\times3\right]=\frac{7}{18}$.

因为$\frac{14}{27}>\frac{7}{18}$,所以振华选择方案一晋级的可能性更大.

3. $y=2$ 【解析】设圆心为$M(1,2)$,连接MA,设α为$\overrightarrow{MA}$与x轴正方向的夹角,



取直线$y=2$,则七个点到该直线距离的平方和为

$$3\sum_{k=0}^{6}\sin^2\left(\alpha+\frac{2k\pi}{7}\right)$$

$$=3\sum_{k=0}^{6}\frac{1-\cos\left(2\alpha+\frac{4k\pi}{7}\right)}{2}$$

$$=3\times\frac{7}{2}-\frac{3}{2}\sum_{k=0}^{6}\cos\left(2\alpha+\frac{4k\pi}{7}\right)$$

$=\frac{21}{2}$,是定值,

任取一条过M的直线l,由对称性可知,七个点到直线l距离的平方和为$\frac{21}{2}$,

由垂线段最短和最小二乘法知回归直线方程为$y=2$.

第八章 全章上分

1. A 【解析】从题中散点图来看，这些点大致落在一条从左上角到右下角的直线附近，呈递减的趋势，所以据此可以推断变量 x 与 y 之间很可能存在负相关. **故选** A.

2. A 【解析】由题可得原数据 $\overline{x}=\frac{1}{5}\times(1+2+3+4+5)=3$，又 $\hat{y}=0.28x+0.16$ 过点 $(\overline{x},\overline{y})$，

提示：经验回归直线过样本点的中心

则 $\overline{y}=0.28\times3+0.16=1$，从而 $\overline{y}=\frac{1}{5}\times(0.5+0.6+m+1.4+1.5)=1$，所以 $m=1$.

设去掉数据 $(3,1)$ 后，新数据为 (k_i,t_i) $(i=1,2,3,4)$，则 $\overline{k}=3=\overline{x}$，$\overline{t}=1=\overline{y}$，

$$r_2=\frac{\sum_{i=1}^{4}(k_i-\overline{k})(t_i-\overline{t})}{\sqrt{\sum_{i=1}^{4}(k_i-\overline{k})^2}\sqrt{\sum_{i=1}^{4}(t_i-\overline{t})^2}},$$

又因为 $r_1=\frac{\sum_{i=1}^{5}(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\sum_{i=1}^{5}(x_i-\overline{x})^2}\sqrt{\sum_{i=1}^{5}(y_i-\overline{y})^2}}$，$x_3-\overline{x}=y_3-\overline{y}=(x_3-\overline{x})(y_3-\overline{y})=0$，

则 $\sum_{i=1}^{5}(x_i-\overline{x})(y_i-\overline{y})=\sum_{i=1}^{4}(k_i-\overline{k})(t_i-\overline{t})$，$\sqrt{\sum_{i=1}^{5}(x_i-\overline{x})^2}\cdot\sqrt{\sum_{i=1}^{5}(y_i-\overline{y})^2}=\sqrt{\sum_{i=1}^{4}(k_i-\overline{k})^2}\cdot\sqrt{\sum_{i=1}^{4}(t_i-\overline{t})^2}$，从而 $r_1=r_2$. **故选** A.

3. A 【解析】设男生人数为 $6n(n\in\mathbf{N}^*)$，因为被调查的男、女生人数相同，

所以女生人数也为 $6n(n\in\mathbf{N}^*)$，根据题意列出列联表：

喜好	性别		合计
	男生	女生	
喜欢冰雪运动	$5n$	$4n$	$9n$
不喜欢冰雪运动	n	$2n$	$3n$
合计	$6n$	$6n$	$12n$

则 $\chi^2=\dfrac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}=$

$\dfrac{12n(5n\cdot 2n-n\cdot 4n)^2}{6n\cdot 6n\cdot 9n\cdot 3n}=\dfrac{432n^5}{972n^4}=\dfrac{4n}{9}$,

因为依据 $\alpha=0.05$ 的独立性检验,认为是否喜欢冰雪运动与学生性别有关联,

所以 $\chi^2\geqslant 3.841$,即 $\dfrac{4n}{9}\geqslant 3.841$,解得

$6n\geqslant 51.8535$,又 $n\in\mathbf{N}^*$,

故 B,C,D 正确,A 错误. 故选 A.

4. B 【解析】$\because \sum\limits_{i=1}^{8}x_i=\dfrac{9}{8}\times 8=9$,

$\therefore$ 增加两个样本数据后 x 的平均数 $\overline{x_1}=\dfrac{9-1+2}{10}=1$.

$\because \overline{y}=2\times\dfrac{9}{8}-\dfrac{1}{4}=2$,$\therefore \sum\limits_{i=1}^{8}y_i=2\times 8=16$,

$\therefore$ 增加两个样本数据后 y 的平均数 $\overline{y_1}=\dfrac{16+5+9}{10}=3$,

设新的经验回归方程为 $\hat{y}_1=3x+\hat{a}$,

$\therefore 3=3\times 1+\hat{a}$,解得 $\hat{a}=0$,

$\therefore$ 新的经验回归方程为 $\hat{y}_1=3x$,则当 $x=4$ 时,$\hat{y}_1=12$,

$\therefore$ 样本数据 $(4,10)$ 所对应的残差为 $10-12=-2$. **故选 B.**

5. A 【解析】因为 $y=c_1e^{c_2x}$,所以 $\ln y=z=c_2x+\ln c_1$,即 $c_2=0.2$.

由题可得 $\overline{x}=\dfrac{20+23+25+27+30}{5}=25$,

$\overline{z}=\dfrac{2+2.4+3+3+4.6}{5}=3$,因为经验回归直线过点 $(\overline{x},\overline{z})$,

所以 $3=0.2\times 25+\ln c_1$,则 $\ln c_1=-2$,

即 $c_1=e^{-2}$. **故选 A.**

6. B 【解析】由题中散点图可知，去掉点 $E(3,0.5)$ 后，y 与 x 的线性相关性变强，且为负相关，**故 B 正确，A 错误；**

由于 y 与 x 的线性相关性变强，所以残差平方和变小，**故 C 错误；**

由于 y 与 x 的线性相关性变强，且为负相关，所以样本相关系数的绝对值变大，而样本相关系数为负值，所以样本相关系数 r 变小，**故 D 错误. 故选 B.**

7. ABD 【解析】对于 A，经验回归方程为 $\hat{y}=5-3x$，经验回归直线的斜率为负，则变量 x 与 y 呈负相关，**故 A 正确；**

对于 B，经验回归直线一定经过样本的中心点$(\bar{x},\bar{y})$，**故 B 正确；**

对于 C，若散点图中所有点都在直线 $y=0.92x-4.21$ 上，则样本相关系数 $r=1$，**故 C 错误；**

对于 D，决定系数 R^2 的值越接近于 1，表示回归模型的拟合效果越好，**故 D 正确. 故选 ABD.**

8. ABD 【解析】由题表可知，$\bar{x}=4$，$\bar{y}=\frac{25.3+m}{7}$，

根据经验回归直线的性质，样本中心点必在经验回归直线上，

$\therefore \frac{25.3+m}{7}=0.5\times4+2.3$，

解得 $m=4.8$，**故 A 正确；**

经验回归直线的斜率为 $0.5>0$，即为正相关，$r>0$，**故 B 正确；**

将 $x=8$ 代入经验回归方程，得 $\hat{y}=0.5\times8+2.3=6.3$，**故 C 错误；**

将 $x=6$ 代入经验回归方程，得 $\hat{y}=0.5\times6+2.3=5.3$，由题表可知，实际值为 5.2，所以实际值比预测值小 0.1，**故 D 正确. 故选 ABD.**

9. 14.2 【解析】由题意可知，$\bar{x}=\frac{21+23+25+27}{4}=24$，$\bar{y}=\frac{15+18+19+20}{4}=18$，

将$(24,18)$代入 $\hat{y}=0.8x+\hat{a}$，得 $18=$

$0.8\times24+\hat{a}$，解得 $\hat{a}=-1.2$，

所以 $\hat{y}=0.8x-1.2$.

当 $x=20$ 时，预测值 $\hat{y}=20\times0.8-1.2=14.8$，则 $n=14.8-0.6=14.2$.

10. 6.818 【解析】已知选取的总人数为 200，不喜欢的人数为 80，那么喜欢的人数 $t=200-80=120$.

因为男性喜欢的人数为 45，所以女性喜欢的人数为 $120-45=75$.

又因为选取的女性人数为 110，所以 $s=110-75=35$. 此时完整的列联表为

	喜欢	不喜欢	合计
男性	45	45	90
女性	75	35	110
合计	120	80	200

在 $\chi^2=\dfrac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$ 中，$n=a+b+c+d=200$，$a=45$（男性喜欢的人数），$b=45$（男性不喜欢的人数），$c=75$（女性喜欢的人数），$d=35$（女性不喜欢的人数）.

将这些值代入公式中可得，$\chi^2=\dfrac{200\times(45\times35-45\times75)^2}{90\times110\times120\times80}=\dfrac{200\times(1\ 575-3\ 375)^2}{90\times110\times120\times80}\approx6.818$.

因为 $6.818>6.635$，所以在犯错误概率不超过 0.01 的前提下，认为性别因素与喜欢情况有关联.

11. 【解】(1) 因为在甲、乙两班共 105 人中，随机抽取 1 人为优秀的概率为 $\dfrac{2}{7}$，

所以优秀的人数为 $105\times\dfrac{2}{7}=30$，则非优秀的人数为 $105-30=75$，

所以列联表为

	优秀	非优秀	合计
甲班	10	45	55
乙班	20	30	50
合计	30	75	105

(2) 根据列联表中的数据可得 $\chi^2=\dfrac{105\times(10\times30-20\times45)^2}{55\times50\times30\times75}\approx6.109>$

$3.841=x_{0.05}$,

所以能在犯错误的概率不超过 0.05 的前提下认为"成绩是否优秀与班级有关联".

12.【解】(1)由题表,$\bar{t}=\frac{1}{5}\times(1+2+3+4+5)=3$,$\bar{y}=\frac{1}{5}\times(2.6+3.1+4.5+6.8+8.0)=5$.

因为 $\sum\limits_{i=1}^{5}t_iy_i=89.5$,$\sqrt{\sum\limits_{i=1}^{5}(t_i-\bar{t})^2}=\sqrt{10}$,$\sqrt{\sum\limits_{i=1}^{5}(y_i-\bar{y})^2}=\sqrt{21.86}$,

所以 $r=\frac{\sum\limits_{i=1}^{5}t_iy_i-5\bar{t}\bar{y}}{\sqrt{\sum\limits_{i=1}^{5}(t_i-\bar{t})^2}\sqrt{\sum\limits_{i=1}^{5}(y_i-\bar{y})^2}}=\frac{14.5}{\sqrt{218.6}}\approx\frac{14.5}{14.785}\approx0.98>0.75$.

故 y 与 t 的线性相关程度很高,可以用线性回归模型拟合.

(2) $\hat{b}=\frac{\sum\limits_{i=1}^{5}t_iy_i-5\bar{t}\bar{y}}{\sum\limits_{i=1}^{5}t_i^2-5\bar{t}^2}=\frac{14.5}{10}=1.45$,

$\hat{a}=\bar{y}-\hat{b}\bar{t}=5-1.45\times3=0.65$,

所以 $\hat{y}=1.45t+0.65$. 当 $t=7$ 时,$\hat{y}=1.45\times7+0.65=10.8$. 预测该专营店在 $t=7$ 时的利润额为 10.8 万元.

13.【解】(1)模型①更合适. 模型①残差点比较均匀地落在水平的带状区域中,且带状区域的宽度比模型②带状区域宽度窄,所以模型①的拟合效果更好,经验回归方程的预测精度相应就会更高,故选模型①比较合适.

(2)令 $z=\ln y$,z 与温度 x 可以用经验回归方程来拟合,设 $\hat{z}=\hat{a}+\hat{b}x$.

于是 $\hat{b}=\frac{\sum\limits_{i=1}^{8}(x_i-\bar{x})(z_i-\bar{z})}{\sum\limits_{i=1}^{8}(x_i-\bar{x})^2}=\frac{50.4}{168}=0.3$,$\hat{a}=\bar{z}-\hat{b}\bar{x}=2.9-0.3\times25=-4.6$,

因此 z 关于 x 的经验回归方程为 $\hat{z}=0.3x-4.6$,即 $\ln\hat{y}=0.3x-4.6$,

所以产卵数 y 关于温度 x 的经验回归方程为 $\hat{y}=e^{0.3x-4.6}$.