(2023 江苏南通二模) 我国风云系列卫星可以检测气象和国土资源情况。某地区水文研究人员为了了解汛期人工测雨量 $x$(单位：dm) 与遥测雨量 $y$(单位：dm) 的关系，统计得到该地区 10 组雨量数据如下：

| 样本号 i | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 人工测雨量 $x_i$ | 5.38 | 7.99 | 6.37 | 6.71 | 7.53 | 5.53 | 4.18 | 4.04 | 6.02 | 4.23 |
| 遥测雨量 $y_i$ | 5.43 | 8.07 | 6.57 | 6.14 | 7.95 | 5.56 | 4.27 | 4.15 | 6.04 | 4.49 |
| $\lvert x_i - y_i\rvert$ | 0.05 | 0.08 | 0.2 | 0.57 | 0.42 | 0.03 | 0.09 | 0.11 | 0.02 | 0.26 |

并计算得 $\sum\limits_{i=1}^{10} x_i^2 = 353.6$，$\sum\limits_{i=1}^{10} y_i^2 = 361.7$，$\sum\limits_{i=1}^{10} x_i y_i = 357.3$，$\bar{x}^2 \approx 33.62$，$\bar{y}^2 \approx 34.42$，$\bar{x}\bar{y} \approx 34.02$

(1) 求该地区汛期遥测雨量 $y$ 与人工测雨量 $x$ 的样本相关系数（精确到 0.01），并判断它们是否具有线性相关关系；

(2) 规定：数组 $(x_i, y_i)$ 满足 $|x_i - y_i| < 0.1$ 为“I 类误差”；满足 $0.1 \le |x_i - y_i| < 0.3$ 为“II 类误差”；满足 $|x_i - y_i| \ge 0.3$ 为“III 类误差”. 为进一步研究，该地区水文研究人员从“I 类误差”、“II 类误差”中随机抽取 3 组数据与“III 类误差”数据进行对比，记抽到“I 类误差”的数据的组数为 $X$，求 $X$ 的概率分布与数学期望。

附：相关系数 $r = \dfrac{\sum\limits_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2 \sum\limits_{i=1}^{n}(y_i - \bar{y})^2}}$，$\sqrt{304.5} \approx 17.4$.

(2020 年全国卷 II) 某沙漠地区经过治理，生态系统得到很大改善，野生动物数量有所增加。为调查该地区某种野生动物的数量，将其分成面积相近的 200 个地块，从这些地块中用简单随机抽样的方法抽取 20 个作为样区，调查得到样本数据 $(x_i,\ y_i)(i=1,2,\cdots,20)$，其中 $x_i$ 和 $y_i$ 分别表示第 $i$ 个样区的植物覆盖面积 (单位：公顷) 和这种野生动物的数量，并计算得 $\sum\limits_{i=1}^{20}x_i=60$，$\sum\limits_{i=1}^{20}y_i=1200$，$\sum\limits_{i=1}^{20}(x_i-\bar{x})^2=80$，$\sum\limits_{i=1}^{20}(y_i-\bar{y})^2=9000$，$\sum\limits_{i=1}^{20}(x_i-\bar{x})(y_i-\bar{y})=800$

(1) 求该地区这种野生动物数量的估计值（这种野生动物数量的估计值等于样区这种野生动物数量的平均数乘以地块数）；

(2) 求样本 $(x_i,\ y_i)(i=1,2,\cdots,20)$ 的相关系数 (精确到 0.01)；

(3) 根据现有统计资料，各地块间植物覆盖面积差异很大。为提高样本的代表性以获得该地区这种野生动物数量更准确的估计，请给出一种你认为更合理的抽样方法，并说明理由.

(2017 全国卷 I) 为了监控某种零件的一条生产线的生产过程，检验员每隔 30 min 从该生产线上随机抽取一个零件，并测量其尺寸（单位：cm）. 下面是检验员在一天内依次抽取的 16 个零件的尺寸：

| 抽取次序 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| 零件尺寸 | 9.95 | 10.12 | 9.96 | 9.96 | 10.01 | 9.92 | 9.98 | 10.04 |
| 抽取次序 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
| 零件尺寸 | 10.26 | 9.91 | 10.13 | 10.02 | 9.22 | 10.04 | 10.05 | 9.95 |

经计算得 $\bar{x}=\frac{1}{16}\sum_{i=1}^{16}x_i=9.97$，$s=\sqrt{\frac{1}{16}\left(\sum_{i=1}^{16}x_i^2-16\bar{x}^2\right)}\approx 0.212$，$\sqrt{\sum_{i=1}^{16}(i-8.5)^2}\approx 18.439$，$\sum_{i=1}^{16}(x_i-\bar{x})(i-8.5)=-2.78$，，其中 $x_i$ 为抽取的第个零件的尺寸，$i=1,2,\cdots,16$.

(1) 求 $(x_i,i)(i=1,2,\cdots,16)$ 的相关系数 $r$，并回答是否可以认为这一天生产的零件尺寸不随生产过程的进行而系统地变大或变小（若 $|r|<0.25$，则可以认为零件的尺寸不随生产过程的进行而系统地变大或变小）.

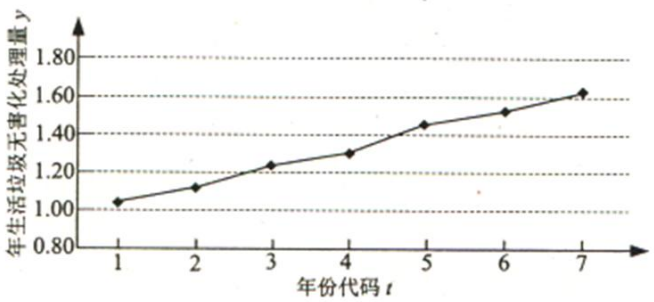
(2) 一天内抽检零件中，如果出现了尺寸在 $(\bar{x}-3s,\bar{x}+3s)$ 之外的零件，就认为这条生产线在这一天的生产过程可能出现了异常情况，需对当天的生产过程进行检查.

(i) 从这一天抽检的结果看，是否需对当天的生产过程进行检查？

(ii) 在 $(\bar{x}-3s,\bar{x}+3s)$ 之外的数据称为离群值，试剔除离群值，估计这条生产线当天生产的零件尺寸的均值与标准差.（精确到 0.01）

附：样本 $(x_i,i)(i=1,2,\cdots,16)$ 的相关系数 $r=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2\sum_{i=1}^{n}(y_i-\bar{y})^2}}$，$\sqrt{0.008}\approx 0.09$.

(2016 年全国高考卷 I) 下图是我国 2008 年至 2014 年生活垃圾无害化处理量（单位：亿吨）的折线图




注：年份代码 1～7 分别对应年份 2008～2014.

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

(2) 建立 y 关于 t 的回归方程 (系数精确到 0.01)，预测 2016 年我国生活垃圾无害化处理量.

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

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

- 1. 计算公式

$$r=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2\sum_{i=1}^{n}(y_i-\bar{y})^2}}=\frac{\sum_{i=1}^{n}(x_iy_i)-n\bar{x}\bar{y}}{\sqrt{\left(\sum_{i=1}^{n}x_i^2-n(\bar{x})^2\right)\left(\sum_{i=1}^{n}y_i^2-n(\bar{y})^2\right)}}$$

- 2. 相关系数的性质
- (1) $|r|\leq 1$
- (2) $|r|$ 越接近于 1，相关程度越强；$|r|$ 越接近于 0，相关程度越弱

(1) 解：由题可知，$(x_i, i)$ 的相关系数 $r = \frac{-2.78}{0.848 \times 18.439} \approx -0.18$，因为 $|r| < 0.25$，所以可以认为这一天生产的零件尺寸不随生产过程的进行而系统地变大或变小.

(2) 解：3s=0.636，$\bar{x} - 3s = 9.334$，$\bar{x} + 3s = 10.606$，而 9.22 出现在了 $(9.334, 10.606)$ 之外，所以需要对当天的生产过程进行检查.

(3) 剔除离群值之后，均值为 $\bar{x}' = \frac{9.97 \times 16 - 9.22}{15} = 10.02$，$\sum_{i=1}^{16} x_i^2 = 16 \times 0.212^2 + 16 \times 9.97^2 \approx 1591.134$，剔除离群值后，剩下数据的样本方差为 $\frac{1}{15}(1591.134 - 9.22^2 - 15 \times 10.02^2) \approx 0.008$, 这条生产线当天生产的零件尺寸的标准差的估计值为 $\sqrt{0.008} \approx 0.09$.

(1) 解：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：

$$\because r=\frac{\sum\limits_{i=1}^{7}(t_i-\bar{t})(y_i-\bar{y})}{\sqrt{\sum\limits_{i=1}^{7}(t_i-\bar{t})^2\sum\limits_{i=1}^{7}(y_i-\bar{y})^2}}=\frac{\sum\limits_{i=1}^{7}t_iy_i-7\bar{t}\bar{y}}{\sqrt{\sum\limits_{i=1}^{7}(t_i-\bar{t})^2\sum\limits_{i=1}^{7}(y_i-\bar{y})^2}}\approx\frac{40.17-4\times9.32}{2\sqrt{7}\times0.55}\approx\frac{2.89}{2.9106}\approx0.996$$

$\because 0.996>0.75$

故 $y$ 与 $t$ 之间存在较强的正相关关系；

(2) 解：$\because \hat{b}=\dfrac{\sum\limits_{i=1}^{n}(t_i-\bar{t})(y_i-\bar{y})}{\sum\limits_{i=1}^{7}(t_i-\bar{t})^2}=\dfrac{\sum\limits_{i=1}^{7}t_iy_i-7\bar{t}\bar{y}}{\sum\limits_{i=1}^{7}t_i^2-7\bar{t}^2}\approx\dfrac{2.89}{28}\approx0.103$，$\hat{a}=\hat{y}-\hat{b}\bar{t}\approx1.331-0.103\times4\approx0.93$，

$\therefore \hat{y}=0.103t+0.93$，2016 年对应的 $t$ 值为 9，故 $\hat{y}=0.103\times9+0.93=1.83$

预测 2016 年我国生活垃圾无害化处理量为 1.83 亿吨.