专题检测（十二） 统计、统计案例

A 组——“6+3+3”考点落实练

一、选择题

1.利用系统抽样法从编号分别为 1，2，3，…，80 的 80 件不同产品中抽出一个容量为 16 的样本，如果抽出的产品中有一件产品的编号为 13，则抽到产品的最大编号为(　　)

A.73　　B.78　　C.77　　D.76

2.(2019·**全国卷Ⅱ**)演讲比赛共有 9 位评委分别给出某选手的原始评分，评定该选手的成绩时，从 9 个原始评分中去掉 1 个最高分、1 个最低分，得到 7 个有效评分.7 个有效评分与 9 个原始评分相比，不变的数字特征是(　　)

A.中位数　　B.平均数　　C.方差　　D.极差

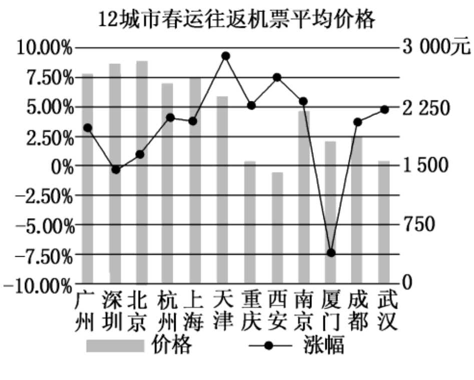
3.(2019·**广东六校第一次联考**)某单位为了落实“绿水青山就是金山银山”理念，制定节能减排的目标，先调查了用电量 y(单位：kW·h)与气温 x(单位：℃)之间的关系，随机选取了 4 天的用电量与当天气温，并制作了如下对照表：

x(单位：℃)	17	14	10	－1
y(单位：kW·h)	24	34	38	a

由表中数据得线性回归方程：$\hat{y}=-2x+60$，则 a 的值为(　　)

A.48　　B.62　　C.64　　D.68

4.如图是民航部门统计的 2019 年春运期间十二个城市售出的往返机票的平均价格以及相比去年同期变化幅度的数据统计图表，根据图表，下面叙述不正确的是(　　)



A.深圳的变化幅度最小，北京的平均价格最高

B.深圳和厦门的春运期间往返机票价格同去年相比有所下降

C.平均价格从高到低居于前三位的城市为北京、深圳、广州

D.平均价格的涨幅从高到低居于前三位的城市为天津、西安、厦门

5.一个样本容量为 10 的样本数据，它们组成一个公差不为 0 的等差数列$\{a_n\}$，若 $a_3=8$，且

a_1，a_3，a_7成等比数列，则此样本的平均数和中位数分别是(　　)

A.13，12　　B.13，13　　C.12，13　　D.13，14

6.(2019·成都市第二次诊断性检测)为比较甲、乙两名篮球运动员的近期竞技状态，选取这两名球员最近五场比赛的得分，制成如图所示的茎叶图.有下列结论：①甲最近五场比赛得分的中位数高于乙最近五场比赛得分的中位数；②甲最近五场比赛得分的平均数低于乙最近五场比赛得分的平均数；③从最近五场比赛的得分看，乙比甲更稳定；④从最近五场比赛的得分看，甲比乙更稳定.

甲		乙
9 8 5	2	8 9
2 1	3	0 1 2

其中所有正确结论的编号为(　　)

A.①③　　B.①④　　C.②③　　D.②④

二、填空题

7.(2019·全国卷Ⅱ)我国高铁发展迅速，技术先进.经统计，在经停某站的高铁列车中，有 10 个车次的正点率为 0.97，有 20 个车次的正点率为 0.98，有 10 个车次的正点率为 0.99，则经停该站高铁列车所有车次的平均正点率的估计值为________.

8.(2019·安徽五校联盟第二次质检)数据 a_1，a_2，a_3，…，a_n的方差为 σ^2，则数据 $2a_1$，$2a_2$，$2a_3$，…，$2a_n$的方差为________.

9.某新闻媒体为了了解观众对央视《开门大吉》节目的喜爱与性别是否有关系，随机调查了观看该节目的观众 110 名，得到如下的列联表：

	女	男	总计
喜爱	40	20	60
不喜爱	20	30	50
总计	60	50	110

试根据样本估计总体的思想，估计在犯错误的概率不超过________的前提下(约有________的把握)认为“喜爱该节目与否和性别有关”.

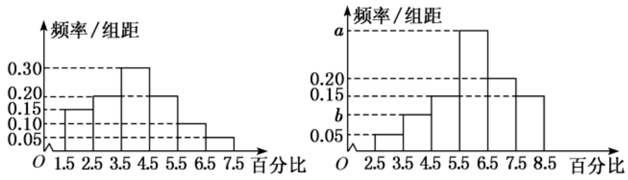
参考附表：

$P(K^2\geqslant k_0)$	0.050	0.010	0.001
k_0	3.841	6.635	10.828

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

三、解答题

10.(2019·全国卷Ⅲ)为了解甲、乙两种离子在小鼠体内的残留程度，进行如下试验：将 200 只小鼠随机分成 A，B 两组，每组 100 只，其中 A 组小鼠给服甲离子溶液，B 组小鼠给服乙离子溶液.每只小鼠给服的溶液体积相同、摩尔浓度相同.经过一段时间后用某种科学方法测算出残留在小鼠体内离子的百分比.根据试验数据分别得到如下直方图：



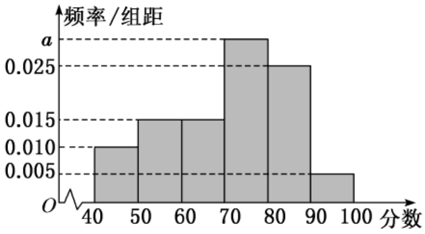
甲离子残留百分比直方图　　乙离子残留百分比直方图

记 C 为事件：“乙离子残留在体内的百分比不低于 5.5”，根据直方图得到 $P(C)$的估计值为 0.70.

(1)求乙离子残留百分比直方图中 a，b 的值；

(2)分别估计甲、乙离子残留百分比的平均值(同一组中的数据用该组区间的中点值为代表).

11.某市教育学院从参加市级高中数学竞赛的考生中随机抽取 60 名学生，将其竞赛成绩(均为整数)分成六段：[40，50)，[50，60)，[60，70)，…，[90，100]，得到如图所示的频率分布直方图.



(1)根据频率分布直方图，估计参加高中数学竞赛的考生的成绩的平均数、众数、中位数(小数点后保留一位有效数字)；

(2)用分层抽样的方法在各分数段的考生中抽取一个容量为 20 的样本，则各分数段抽取的人数分别是多少？

12.(2019·沈阳市质量监测(一))某篮球运动员的投篮命中率为 50%，他想提高自己的投篮水平，制定了一个夏季训练计划，为了了解训练效果，执行训练前，他统计了 10 场比赛的得分，计算出得分的中位数为 15，平均得分为 15，得分的方差为 46.3.执行训练后也统计了 10 场比赛的得分，茎叶图如图所示：

0	8	9				
1	2	4	4	5	6	8
2	1	3				

(1)请计算该篮球运动员执行训练后统计的 10 场比赛得分的中位数、平均得分与方差.

(2)如果仅从执行训练前后统计的各 10 场比赛得分数据分析，你认为训练计划对该运动员的投篮水平的提高是否有帮助？为什么？

B组——大题专攻强化练

1.(2019·武汉市调研测试)一个工厂在某年里连续10个月每月产品的总成本y(万元)与该月产量x(万件)之间有如下一组数据：

x	1.08	1.12	1.19	1.28	1.36	1.48	1.59	1.68	1.80	1.87
y	2.25	2.37	2.40	2.55	2.64	2.75	2.92	3.03	3.14	3.26

(1)通过画散点图，发现可用线性回归模型拟合y与x的关系，请用相关系数加以说明.

(2)①建立月总成本y与月产量x之间的回归方程；

②通过建立的y关于x的回归方程，估计某月产量为1.98万件时，产品的总成本为多少万元？(均精确到0.001)

附注：①参考数据：**错误!**$_i$＝27.31，

$\sqrt{\sum\limits_{i=1}^{10}x_i^2-10\overline{x}^2}\approx0.850$，$\sqrt{\sum\limits_{i=1}^{10}y_i^2-10\overline{y}^2}\approx1.042$，$\hat{b}\approx1.223$.

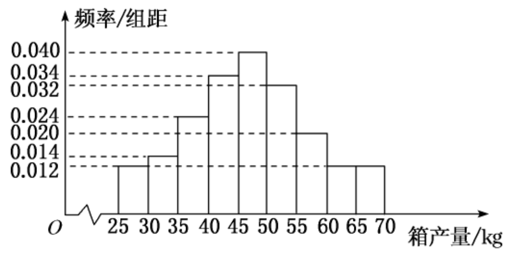
②参考公式：相关系数

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

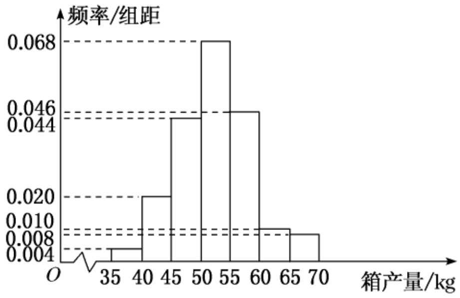
回归直线$\hat{y}=\hat{a}+\hat{b}x$中斜率和截距的最小二乘估计公式分别为：

$$\hat{b}=\frac{\sum\limits_{i=1}^{n}x_iy_i-n\overline{x}\,\overline{y}}{\sum\limits_{i=1}^{n}x_i^2-n\overline{x}^2},\hat{a}=\overline{y}-\hat{b}\,\overline{x}.$$

2.海水养殖场进行某水产品的新、旧网箱养殖方法的产量对比，收获时各随机抽取了 100 个网箱，测量各箱水产品的产量(单位：kg)，其频率分布直方图如下：



旧养殖法



新养殖法

(1)估计旧养殖法的箱产量低于 50 kg 的概率并估计新养殖法的箱产量的平均值；

(2)填写下面的 2×2 列联表，并根据列联表判断是否有 99%的把握认为箱产量与养殖方法有关.

	箱产量<50 kg	箱产量≥50 kg	总计
旧养殖法			
新养殖法			
总计			

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

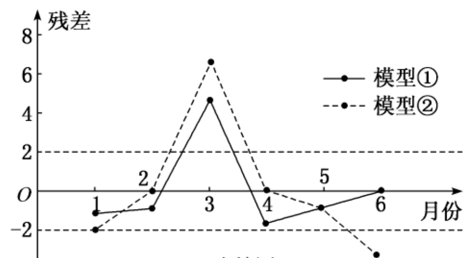
$P(K^2\geqslant k_0)$	0.050	0.010	0.001
k_0	3.841	6.635	10.828

3.(2019·长沙市统一模拟考试)某互联网公司为了确定下一季度的前期广告投入计划，收集了近 6 个月广告投入量 x(单位：万元)和收益 y(单位：万元)的数据如下表：

月份	1	2	3	4	5	6
广告投入量/万元	2	4	6	8	10	12
收益/万元	14.21	20.31	31.8	31.18	37.83	44.67

他们用两种模型①$y=bx+a$，②$y=ae^{bx}$ 分别进行拟合，得到相应的回归方程并进行残差分析，得到如图所示的残差图及一些统计量的值：

$\overline{x}$	$\overline{y}$	$\sum\limits_{i=1}^{6}x_iy_i$	$\sum\limits_{i=1}^{6}x_i^2$
7	30	1 464.24	364



残差图

(1)根据残差图，比较模型①，②的拟合效果，应选择哪个模型？并说明理由.

(2)残差绝对值大于 2 的数据被认为是异常数据，需要剔除：(ⅰ)剔除异常数据后，求出(1)中所选模型的回归方程；(ⅱ)广告投入量 $x=18$ 时，(1)中所选模型收益的预报值是多少？

附：对于一组数据$(x_1，y_1)$，$(x_2，y_2)$，…，$(x_n，y_n)$，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：

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

4.每年 10 月中上旬是小麦的最佳种植时间，但小麦的发芽会受到土壤、气候等多方面因素的影响.某科技兴趣小组为了解昼夜温差的大小与小麦发芽的多少之间的关系，在不同的温差下统计了 100 颗小麦种子的发芽数，得到了如下数据：

温差 x(℃)	8	10	11	13	12
发芽数 y(颗)	79	81	85	90	86

(1)请根据统计的最后三组数据，求出 y 关于 x 的线性回归方程$\hat{y}=\hat{b}x+\hat{a}$；

(2)若由(1)中的线性回归方程得到的估计值与前两组数据的实际值误差均不超过两颗，则认为线性回归方程是可靠的，试判断(1)中得到的线性回归方程是否可靠；

(3)若 100 颗小麦种子的发芽数为 n 颗，则记 n%的发芽率，当发芽率为 n%时，平均每亩地的收益为 10n 元，某农场有土地 10 万亩，小麦种植期间昼夜温差大约为 9 ℃，根据(1)中得到的线性回归方程估计该农场种植小麦所获得的收益.

附：在线性回归方程$\hat{y}=\hat{b}x+\hat{a}$中，$\hat{b}=\dfrac{\sum_{i=1}^{n}x_iy_i-n\overline{x}\,\overline{y}}{\sum_{i=1}^{n}x_i^2-n\overline{x}^2}$.