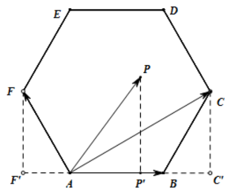
2023年高考数学考前复习材料——新教材中的新增知识查漏（1）

【真题展示】

1．(2020新高考I(山东)第7题)已知P是边长为2的正六边形$ABCDEF$内的一点，则$\overrightarrow{AP}\cdot\overrightarrow{AB}$的取值范用是（　　）

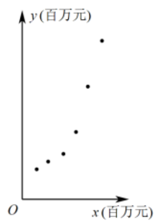
A．$(-2,6)$　　B．$(-6,2)$　　C．$(-2,4)$　　D．$(-4,6)$

【答案】A

【解析】

$\overrightarrow{AB}$的模为2，根据正六边形的特征，可以得到$\overrightarrow{AP}$在$\overrightarrow{AB}$方向上的投影的取值范围是$(-1,3)$，结合向量数量积的定义式，可知$\overrightarrow{AP}\cdot\overrightarrow{AB}$等于$\overrightarrow{AB}$的模与$\overrightarrow{AP}$在$\overrightarrow{AB}$方向上的投影的乘积，所以$\overrightarrow{AP}\cdot\overrightarrow{AB}$的取值范围是$(-2,6)$，故选：A.

2．（2023河北秦皇岛检测）某企业秉承“科学技术是第一生产力”的发展理念，投入大量科研经费进行技术革新，该企业统计了最近6年投入的年科研经费x（单位：百万元）和年利润y（单位：百万元）的数据，并绘制成如图所示的散点图．已知x，y的平均值分别为$\bar{x}=7$，$\bar{y}=10$．甲统计员得到的回归方程为$\hat{y}=1.69x+\hat{a}$；乙统计员得到的回归方程为$\hat{y}=2.52e^{0.17x}$；若甲、乙二人计算均未出现错误，则以下结论正确的为（　　）



A．当投入年科研经费为20（百万元）时，按乙统计员的回归方程可得年利润估计值为75.6（百万元）（取$e^{3.4}=30$）

B．$\hat{a}=-1.83$

C．方程$\hat{y}=1.69x+\hat{a}$比方程$\hat{y}=2.52e^{0.17x}$拟合效果好

D．y与x正相关

【答案】ABD

【分析】将$x=20$代入对应的回归方程，判断A，结合样本中心点过回归直线方程判断B，由散点图判断C，根据正相关的定义判断D.

【解析】将$x=20$代入$\hat{y}=2.52e^{0.17x}$得$\hat{y}=75.6$，A正确；

将$\bar{x}=7,\bar{y}=10$代入$\hat{y}=1.69x+\hat{a}$得$\hat{a}=-1.83$，B正确；

由散点图可知，回归方程$\hat{y}=2.52e^{0.17x}$比$\hat{y}=1.69x+\hat{a}$的拟合效果更好，C错误；

因为y随x的增大而增大，所以y与x正相关，D正确.
因为y随x的增大而增大，所以y与x正相关，D正确.
故选：ABD.

3．（多选）（2023湖南常德一模）下列说法正确的是（　　）

A．数据6，5，3，4，2，7，8，9的上四分位数为7

B．若$\xi \sim N(\mu,\sigma^2)$，且函数$f(x)=P(x\le\xi\le x+2)$为偶函数，则$\mu=1$

C．若随机事件A，B满足：$P(A|B)+P(\overline{A})=1$，则A，B相互独立

D．已知采用分层抽样得到的样本数据由两部分组成，第一部分样本数据$x_i\left(i=1,2,\cdots,n\right)$的平均数为$\overline{x}$，方差为$s_x^2$；第二部分样本数据$y_i\left(i=1,2,\cdots,n\right)$的平均数为$\overline{y}$，方差为$s_y^2$，若总的样本方差为$S^2=\dfrac{S_x^2+S_y^2}{2}$，则$\overline{x}=\overline{y}$

【答案】BCD

【分析】根据偶函数的定义及正态曲线的对称性即可判断B,根据百分位数的概念即可判断A，根据对立事件的概率公式，条件概率公式，独立事件的积事件的乘法公式即可求解C；根据平均数和方差的计算公式即可化简求解D.

【解析】

对A，∵数据6，5，3，4，2，7，8，9按从小到大排列为：2，3，4，5，6，7，8，9一共8个数，又$8\times\dfrac{3}{4}=6$，∴该数据的上四分位数为$\dfrac{8+7}{2}=7.5$，故A错误；

对于B,∵函数$f(x)=P(x\le\xi\le x+2)$为偶函数，$\therefore f(-x)=f(x)$，

$\therefore P(-x\le\xi\le -x+2)=P(x\le\xi\le x+2)$，又$\xi\sim N(\mu,\sigma^2)$，∴区间$\left[-x,-x+2\right]$与区间$\left[x,x+2\right]$关于$x=\mu$对称，$\therefore \mu=\dfrac{-x+x+2}{2}=\dfrac{x+2-x}{2}=1$，故B正确；

对C，$\because 0<P(A)<1$，$0<P(B)<1$，且$P(A|B)+P(\overline{A})=1$，$\therefore P(A|B)=1-P(\overline{A})=P(A)$，

$\therefore \dfrac{P(AB)}{P(B)}=P(A)$，$\therefore P(AB)=P(A)P(B)$，故$A,B$ 相互独立，所以C正确；

对 D，第一部分样本数据$x_i\left(i=1,2,\cdots,n\right)$的平均数为$\overline{x}$，方差为$s_x^2$；则$s_x^2=\dfrac{1}{n}(\sum\limits_{i=1}^{n}x_i^2-n\overline{x}^2)$第二部分样本数据$y_i\left(i=1,2,\cdots,n\right)$的平均数为$\overline{y}$，方差为$s_y^2$，则$s_y^2=\dfrac{1}{n}\left(\sum\limits_{i=1}^{n}y_i^2-n\overline{y}^2\right)$，若总的样本方差为$s^2=\dfrac{1}{2n}\left[\sum\limits_{i=1}^{n}y_i^2+\sum\limits_{i=1}^{n}x_i^2-2n\left(\dfrac{\overline{x}+\overline{y}}{2}\right)^2\right]=\dfrac{1}{2n}\left[\sum\limits_{i=1}^{n}y_i^2+\sum\limits_{i=1}^{n}x_i^2-2n\left(\dfrac{\overline{x}+\overline{y}}{2}\right)^2\right]$，若$s^2=\dfrac{S_x^2+S_y^2}{2}$，即

$\dfrac{1}{2n}\left[\sum\limits_{i=1}^{n}y_i^2+\sum\limits_{i=1}^{n}x_i^2-2n\left(\dfrac{\overline{x}+\overline{y}}{2}\right)^2\right]=\dfrac{1}{2}\left[\dfrac{1}{n}\left(\sum\limits_{i=1}^{n}x_i^2-n\overline{x}^2\right)+\dfrac{1}{n}\left(\sum\limits_{i=1}^{n}y_i^2-n\overline{y}^2\right)\right]$,、

$$\sum_{i=1}^{n}y_i^2+\sum_{i=1}^{n}x_i^2-2n\left(\frac{\overline{x}+\overline{y}}{2}\right)^2=\left(\sum_{i=1}^{n}x_i^2-n\overline{x}^2\right)+\left(\sum_{i=1}^{n}y_i^2-n\overline{y}^2\right)\Rightarrow\left(\overline{x}-\overline{y}\right)^2=0\Rightarrow\overline{x}=\overline{y}$$

故D正确，故选：BCD

4．（2023广东深圳一模）某企业因技术升级，决定从2023年起实现新的绩效方案．方案起草后，为了解员工对新绩效方案是否满意，决定采取如下“随机化回答技术”进行问卷调查：

一个袋子中装有三个大小相同的小球，其中 1 个黑球，2 个白球．企业所有员工从袋子中有放回的随机摸两次球，每次摸出一球．约定“若两次摸到的球的颜色不同，则按方式Ⅰ回答问卷，否则按方式Ⅱ回答问卷”．

方式Ⅰ：若第一次摸到的是白球，则在问卷中画“○”，否则画“×”；

方式Ⅱ：若你对新绩效方案满意，则在问卷中画“○”，否则画“×”．

当所有员工完成问卷调查后，统计画○，画×的比例．用频率估计概率，由所学概率知识即可求得该企业员工对新绩效方案的满意度的估计值．其中满意度$=\dfrac{\text{企业所有对新绩效方案满意的员工人数}}{\text{企业所有员工人数}}\times 100\%$．

（1）若该企业某部门有 9 名员工，用 X 表示其中按方式Ⅰ回答问卷的人数，求 X 的数学期望；

（2）若该企业的所有调查问卷中，画“○”与画“×”的比例为 $4:5$，试估计该企业员工对新绩效方案的满意度．

【解】（1）每次摸到白球的概率 $\dfrac{2}{3}$，摸到黑球的概率为 $\dfrac{1}{3}$，

每名员工两次摸到的球的颜色不同的概率 $P=\mathrm{C}_2^1\times\dfrac{2}{3}\times\dfrac{1}{3}=\dfrac{4}{9}$，

由题意可得：该部门 9 名员工中按方式Ⅰ回答问卷的人数 $X\sim B(9,\dfrac{4}{9})$，

所以 X 的数学期望 $E\left(X\right)=9\times\dfrac{4}{9}=4$．

（2）记事件 A 为“按方式I回答问卷”，事件 B 为“按方式II回答问卷”，事件 C 为“在问卷中画○”．

由（1）知 $P\left(A\right)=\dfrac{4}{9}$，$P\left(B\right)=1-P\left(A\right)=\dfrac{5}{9}$，$P\left(A\right)P\left(C|A\right)=P\left(AC\right)=\dfrac{2}{3}\times\dfrac{1}{3}=\dfrac{2}{9}$．

$\because P\left(C\right)=\dfrac{4}{4+5}=\dfrac{4}{9}$，

由全概率公式 $P\left(C\right)=P\left(A\right)P\left(C|A\right)+P\left(B\right)P\left(C|B\right)$，

则 $\dfrac{4}{9}=\dfrac{2}{9}+\dfrac{5}{9}P\left(C|B\right)$，解得 $P\left(C|B\right)=\dfrac{2}{5}=0.4$，

故根据调查问卷估计，该企业员工对新绩效方案的满意度为 40%．

5．（2022 新高考I）一医疗团队为研究某地的一种地方性疾病与当地居民的卫生习惯（卫生习惯分为良好和不够良好两类）的关系，在已患该疾病的病例中随机调查了 100 例（称为病例组），同时在未患该疾病的人群中随机调查了 100 人（称为对照组），得到如下数据：

	不够良好	良好
病例组	40	60
对照组	10	90

（1）能否有 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）从该地的人群中任选一人，A 表示事件“选到的人卫生习惯不够良好”，B 表示事件“选到的人患有该疾病”，$\dfrac{P(B|A)}{P(\bar{B}|A)}$ 与 $\dfrac{P(B|\bar{A})}{P(\bar{B}|\bar{A})}$ 的比值是卫生习惯不够良好对患该疾病风险程度的一项度量指标，记该指标为 R．

（i）证明：$R=\dfrac{P(A|B)}{P(\bar{A}|B)}\cdot\dfrac{P(\bar{A}|\bar{B})}{P(A|\bar{B})}$；

（ii）利用该调查数据，给出 $P(A|B)$，$P(A|\bar{B})$ 的估计值，并利用（i）的结果给出 R 的估计值．

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

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

【思路分析】（1）补充列联表，根据表中数据计算 K^2，对照附表得出结论.
（2）（ⅰ）根据条件概率的定义与运算性质，证明即可；
（ⅱ）利用调查数据和对立事件的概率公式，计算即可.
【解】（1）补充列联表为：

	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

计算 $K^2=\dfrac{200\times(40\times90-10\times60)^2}{100\times100\times50\times150}=24>6.635$，

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：

$$R=\frac{P(B|A)}{P(\bar{B}|A)}:\frac{P(B|\bar{A})}{P(\bar{B}|\bar{A})}=\frac{P(B|A)}{P(\bar{B}|A)}\cdot\frac{P(\bar{B}|\bar{A})}{P(B|\bar{A})}=\frac{\frac{P(AB)}{P(A)}}{\frac{P(A\bar{B})}{P(A)}}\cdot\frac{\frac{P(\bar{A}\bar{B})}{P(\bar{A})}}{\frac{P(\bar{A}B)}{P(\bar{A})}}=\frac{P(AB)\cdot P(\bar{A}\bar{B})}{P(A\bar{B})\cdot P(\bar{A}B)}=\frac{\frac{P(AB)}{P(B)}}{\frac{P(\bar{A}B)}{P(B)}}\cdot\frac{\frac{P(\bar{A}\bar{B})}{P(\bar{B})}}{\frac{P(A\bar{B})}{P(\bar{B})}}=\frac{P(A|B)}{P(\bar{A}|B)}\cdot\frac{P(\bar{A}|\bar{B})}{P(A|\bar{B})};$$

（ⅱ）利用调查数据，$P(A|B)=\dfrac{40}{100}=\dfrac{2}{5}$，$P(A|\bar{B})=\dfrac{10}{100}=\dfrac{1}{10}$，$P(\bar{A}|B)=1-P(A|B)=\dfrac{3}{5}$，

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

【解读】

新教材（人教A）就课程内容而言，与旧教材比，新增知识：

1. 平面向量投影的概念与投影向量的意义（必修第二册）；
   空间向量投影的概念与投影向量的意义（选择性必修第一册）；
   用向量方法解决空间中的距离问题（选择性必修第一册）；
2. 分层抽样的样本均值与样本方差（必修第二册）；
3. 用样本估计百分位数及百分位数的含义（必修第二册）；
4. 全概率公式（选择性必修第三册）；
5. 残差、相关系数 r、决定系数 R^2（选择性必修第三册）；
   （相关系数 r 提高了要求，增加了样本相关系数与标准化数据向量夹角的关系内容）；
6. 复数的三角形式（必修第二册）（选学）；
7. 贝叶斯公式（选择性必修第三册）（选学）；
8. 必修和选择性必修均增加了数学建模与数学探究活动.

【投影向量】

例1.（1）（2023湖南常德一模）已知向量 $\vec{a},\vec{b}$ 满足 $\vec{a}\cdot\vec{b}=2$，且 $\vec{b}=(3,-4)$，则向量 $\vec{a}$ 在向量 $\vec{b}$ 上的投影向量为（　　）

A. $(\frac{6}{5},-\frac{8}{5})$　　B. $(-\frac{6}{5},\frac{8}{5})$　　C. $(-\frac{6}{25},\frac{8}{25})$　　D. $(\frac{6}{25},-\frac{8}{25})$

【答案】D

2023年高考数学考前复习材料——新教材中的新增知识查漏（2）

【解析】因为$\vec{a},\vec{b}$满足$\vec{a}\cdot\vec{b}=2$，且$\vec{b}=(3,-4)$，所以$|\vec{b}|=5$，向量$\vec{a}$在向量$\vec{b}$上的投影向量为$\frac{\vec{b}}{|\vec{b}|}\cdot\frac{\vec{a}\cdot\vec{b}}{|\vec{b}|}=\frac{2}{25}\vec{b}=(\frac{6}{25},-\frac{8}{25})$，故选：D.

（2）（2023南师附中等四校联考）在ΔABC中，若向量$\overrightarrow{AC}$在$\overrightarrow{AB}$上的投影向量为$\frac{1}{4}\overrightarrow{AB}$，则$A-B$的最大值为（　　）

A．$\frac{\pi}{3}$　　B．$\frac{\pi}{4}$　　C．$\frac{\pi}{6}$　　D．$\frac{\pi}{12}$

【答案】C

【解析】记ΔABC的内角A，B，C的对边分别为a，b，c，则因为向量$\overrightarrow{AC}$在$\overrightarrow{AB}$上的投影向量为$\frac{1}{4}\overrightarrow{AB}$，所以$\overrightarrow{AC}\cdot\overrightarrow{AB}=\frac{1}{4}\overrightarrow{AB}^2$，所以$bc\cos A=\frac{1}{4}c^2$，即$c=4b\cos A$，结合正弦定理，得$\sin C=4\sin B\cos A$，所以$\sin(A+B)=4\sin B\cos A$，所以$\sin A\cos B=3\sin B\cos A$，所以$\tan A=3\tan B>0$，所以$\tan(A-B)=\frac{\tan A-\tan B}{1+\tan A\tan B}=\frac{2\tan B}{1+3\tan^2 B}\le\frac{\sqrt{3}}{3}$，当且仅当$\tan B=\frac{\sqrt{3}}{3}$时等号成立．结合$0<A-B<\frac{\pi}{2}$，得$A-B$的最大值为$\frac{\pi}{6}$，选C.

【变式演练】

（2023湖南一模）已知ΔABC的外接圆的圆心为O，半径为1，$\overrightarrow{AO}=\frac{1}{2}\overrightarrow{AB}+\frac{1}{2}\overrightarrow{AC}$，$\overrightarrow{BA}$在$\overrightarrow{BC}$上的投影向量为$\frac{1}{4}\overrightarrow{BC}$，则$\overrightarrow{OA}\cdot\overrightarrow{BC}=$（　　）

A．$-\sqrt{3}$　　B．-1　　C．1　　D．$\sqrt{3}$

【答案】B

【解析】$\overrightarrow{AO}=\frac{1}{2}\overrightarrow{AB}+\frac{1}{2}\overrightarrow{AC}$，则$O$为$BC$中点，$O$又是外接圆圆心，

则ΔABC为直角三角形，$\frac{1}{4}\overrightarrow{BC}$为$\overrightarrow{BA}$在$\overrightarrow{BC}$上的投影向量，

$\therefore\frac{|\overrightarrow{BA}|\cos B}{|\overrightarrow{BC}|}\cdot\overrightarrow{BC}=\frac{1}{4}\overrightarrow{BC}$，$\therefore\frac{|\overrightarrow{BA}|\cos B}{|\overrightarrow{BC}|}=\frac{1}{4}$，

$\therefore\cos^2 B=\frac{1}{4}$，$\therefore\cos B=\frac{1}{2}$，$\therefore B=\frac{\pi}{3}$，$C=\frac{\pi}{6}$，

$\therefore\overrightarrow{OA}\cdot\overrightarrow{BC}=-\overrightarrow{AO}\cdot\overrightarrow{BC}=-\frac{1}{2}(\overrightarrow{AB}+\overrightarrow{AC})(\overrightarrow{AC}-\overrightarrow{AB})=-\frac{1}{2}(\overrightarrow{AC}^2-\overrightarrow{AB}^2)$，

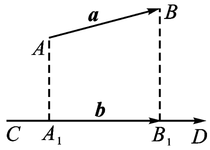
ΔABC的外接圆半径为1，$\therefore BC=2$，$\therefore AB=1$，$AC=\sqrt{3}$，

$\therefore\overrightarrow{OA}\cdot\overrightarrow{BC}=-\frac{1}{2}(3-1)=-1$，故选：B.

【总结提炼】

（1）定义：如图，设$\vec{a}$，$\vec{b}$是两个非零向量，$\overrightarrow{AB}=\vec{a}$，$\overrightarrow{CD}=\vec{b}$，作如下的变换：过$\overrightarrow{AB}$的起

点 A 和终点 B，分别作 $\overrightarrow{CD}$ 所在直线的垂线，垂足分别为 A_1，B_1，得到 $\overrightarrow{A_1B_1}$，则称上述变换为向量 $\vec{a}$ 向向量 $\vec{b}$ 投影，$\overrightarrow{A_1B_1}$ 叫做向量 $\vec{a}$ 在向量 $\vec{b}$ 上的投影向量；



（2）计算：设与 $\vec{b}$ 方向相同的单位向量为 $\vec{e}$，$\vec{a}$ 与 $\vec{b}$ 的夹角为 θ，则向量 $\vec{a}$ 在向量 $\vec{b}$ 上的投影向量是 $|\vec{a}|\cos\theta\vec{e}$；

（3）$\vec{a}\cdot\vec{b}=\overrightarrow{AB}\cdot\overrightarrow{CD}=\overrightarrow{A_1B_1}\cdot\overrightarrow{CD}$.

【分层抽样的样本均值与样本方差】

例 2.（武昌区 2023 元月质量检测）某校采用分层随机抽样采集了高一、高二、高三年级学生的身高情况，部分调查数据如下：

项目	样本量	样本平均数	样本方差
高一	100	167	120
高二	100	170	150
高三	100	173	150

则总的样本方差 $s^2=$______.

【答案】146

【分析】由分层抽样后的样本方差公式计算可得结果.

【解析】由题意知，总的样本平均数为 $\bar{x}=\frac{100}{300}\times167+\frac{100}{300}\times170+\frac{100}{300}\times173=170$，

∴总的样本方差为：

$$s^2=\frac{100}{300}\times[120+(167-170)^2]+\frac{100}{300}\times[150+(170-170)^2]+\frac{100}{300}\times[150+(173-170)^2]$$

$$=\frac{1}{3}\times(120+9)+\frac{1}{3}\times150+\frac{1}{3}\times(150+9)=146,$$

故答案为：146.

【变式演练】

在对某中学高一年级学生身高的调查中，采用样本量比例分配的分层随机抽样，若只知道抽取了男生 24 人，其平均数和方差分别为 170.5 和 12.96，抽取了女生 26 人，其平均数和方差分别为 160.5 和 36.96，则据此可估计高一年级全体学生的身高的均值为________，方差为________.

【答案】165.3，50.4

【解析】设 24 名男生的身高分别为 x_1，x_2，x_3，…，x_{24}，平均数为 $\bar{x}$，26 名女生的身高分别为 y_1，y_2，y_3，…，y_{26}，平均数为 $\bar{y}$，样本中 50 人的身高平均为 $\bar{z}$，

$\bar{x}=\frac{1}{24}\sum_{i=1}^{24}x_i=170.5$，可得 $\sum_{i=1}^{24}x_i=24\times170.5=4\ 092$，$\bar{y}=\frac{1}{26}\sum_{i=1}^{26}y_i=160.5$，

$\bar{z}=\frac{1}{50}\sum_{i=1}^{50}z_i=\frac{1}{50}(\sum_{i=1}^{24}x_i+\sum_{i=1}^{26}y_i)=\frac{1}{50}(24\bar{x}+26\bar{y})=\frac{1}{50}(4\ 092+4\ 173)=165.3$，

可得$\sum_{i=1}^{26} y_i = 26\overline{y} = 26\times 160.5 = 4173$，

$s_x^2 = \frac{1}{24}\sum_{i=1}^{24}(x_i - \overline{x})^2 = \frac{1}{24}\sum_{i=1}^{24}x_i^2 - \overline{x}^2 = 12.96$，

可得$\sum_{i=1}^{24}x_i^2 = 24(s_x^2 + \overline{x}^2)$，$s_y^2 = \frac{1}{26}\sum_{i=1}^{26}(y_i - \overline{y})^2 = \frac{1}{26}\sum_{i=1}^{26}y_i^2 - \overline{y}^2 = 36.96$，

可得$\sum_{i=1}^{26}y_i^2 = 26(s_y^2 + \overline{y}^2)$，

$s_z^2 = \frac{1}{50}\sum_{i=1}^{50}(z_i - \overline{z})^2 = \frac{1}{50}[\sum_{i=1}^{24}(x_i - \overline{z})^2 + \sum_{i=1}^{26}(y_i - \overline{z})^2]$

$= \frac{1}{50}[24(s_x^2 + (\overline{x} - \overline{z})^2) + 26(s_y^2 + (\overline{y} - \overline{z})^2)]$

$= \frac{1}{50}[24(12.96 + (170.5 - 165.3)^2) + 26(36.96 + (160.5 - 165.3)^2)]$

$= \frac{1}{50}[24(12.96 + 27.04) + 26(36.96 + 23.04)]$

$= \frac{1}{50}(24\times 40 + 26\times 60) = \frac{1}{50}\times 2\,520 = 50.4.$

【总结提炼】

（1）在分层抽样中，如果层数分为2层，第1层和第2层包含的个体数分别为M和N，抽取的样本容量分别为m和n，第1层和第2层的样本平均数分别为$\overline{x}$和$\overline{y}$，则样本的平均数$\overline{\omega} = \frac{m}{m+n}\overline{x} + \frac{n}{m+n}\overline{y} = \frac{M}{M+N}\overline{x} + \frac{N}{M+N}\overline{y}$.

（2）在分层抽样中，我们可以直接用样本平均数$\overline{\omega}$估计总体平均数$\overline{W}$.

（3）设样本容量为n，平均数为$\overline{x}$，其中两层的个体数分别为n_1和n_2，两层的平均数分别为$\overline{x_1}$和$\overline{x_2}$，方差分别为s_1^2，s_2^2，则这个样本的方差$s^2 = \frac{n_1}{n}[s_1^2 + (\overline{x_1} - \overline{x})^2] + \frac{n_2}{n}[s_2^2 + (\overline{x_2} - \overline{x})^2]$.

【用样本估计百分位数及百分位数的含义】

例 3.（1）（2023 辽宁校联考模拟）某地有 9 个快递收件点，在某天接收到的快递个数分别为 360，284，290，300，402，188，240，260，288，则这组数据的第 72 百分位数为（　　）

A．290　　B．295　　C．300　　D．330

【答案】C

【解析】将这组数据按照从小到大的顺序排列得 188，240，260，284，288，290，300，360，402，因为$9\times 0.72 = 6.48$，所以这组数据的第 72 百分位数为 300.故选：C

（2）（2022 广东广州一模）为了养成良好的运动习惯，某人记录了自己一周内每天的运动时长（单位：分钟），分别为 53，57，45，61，79，49，x，若这组数据的第 80 百分位数与第 60 百分位数的差为 3，则$x =$（　　）

A．58 或 64　　B．59 或 64　　C．58　　D．59

【答案】A

【解析】将已知的 6 个数从小到大排序为 45，49，53，57，61，79.

若$x \le 57$，则这组数据的第 80 百分位数与第 60 百分位数分别为 61 和 57，他们的差为 4，不符合条件；若$x \ge 79$，则这组数据的第 80 百分位数与第 60 百分位数分别为 79 和 61，它们的差为 18，不符合条件；若$57 < x < 79$，则这组数据的第 80 百分位数与第 60 百分位数分别为x和 61（或 61 和x），则$|x - 61| = 3$，解得$x = 58$或$x = 64$，故选：A.

【变式演练】

（2023 湖南模拟）洞庭湿地保护区于长江中游的湖南省，面积 168000 公顷，为了保护该湿地保护区内的渔业资源和生物多样性，从 2003 年起全面实施禁渔期制度.该湿地保护区的渔业资源科学研究培殖了一批珍稀类银鱼鱼苗，从中随机抽取 100 尾测量鱼苗的体长（单位：毫米），所得的数据如下表：

分组（单位：毫米）	$[70,75)$	$[75,80)$	$[80,85)$	$[85,90)$	$[90,95)$	$[95,100)$
频数	10	10	m	35	15	n

若依上述 6 组数据绘制的频率分布直方图中，$[95,100)$ 分组对应小矩形的高为 0.01，则该样本中的90%分位数的银鱼鱼苗的体长为（保留一位小数）（　　）

A．87 毫米　　　B．88 毫米　　　C．90.5 毫米　　　D．93.3 毫米

【答案】D

【解析】由题意可知，$[95,100)$ 内的频率为 0.05，

所以 $n=100\times0.05$，$m=100-10-10-35-15-5=25$，

鱼苗体长在 $[70,90)$ 内的频率为 0.80，在 $[70,95)$ 内的频率为 0.95，

所以 90%分位数在区间 $[90,95)$ 内，大小为 $90+5\times\frac{2}{3}\approx93.3$.故选：D

【总结提炼】

（1）定义

一组数据的第 p 百分位数是这样一个值，它使得这组数据中至少有 $p\%$ 的数据小于或等于这个值，且至少有 $(100-p)\%$ 的数据大于或等于这个值.

（2）计算一组 n 个数据的的第 p 百分位数的步骤

①按从小到大排列原始数据.

②计算 $i=n\times p\%$.

③若 i 不是整数而大于 i 的比邻整数 j，则第 p 百分位数为第 j 项数据；若 i 是整数，则第 p 百分位数为第 i 项与第 $i+1$ 项数据的平均数.

（3）四分位数

我们之前学过的中位数，相当于是第50百分位数．在实际应用中，除了中位数外，常用的分位数还有第25百分位数，第75百分位数．这三个分位数把一组由小到大排列后的数据分成四等份，因此称为四分位数.

【残差、相关系数 r 、决定系数 R^2】

例 4.（2023 山东潍坊一模）某学校研究性学习小组在学习生物遗传学的过程中，为验证高尔顿提出的关于儿子成年后身高 y（单位：cm）与父亲身高 x（单位：cm）之间的关系及存在的遗传规律，随机抽取了 5 对父子的身高数据，如下表：

父亲身高 x	160	170	175	185	190
儿子身高 y	170	174	175	180	186

（1）根据表中数据，求出 y 关于 x 的线性回归方程，并利用回归直线方程分别确定儿子比父亲高和儿子比父亲矮的条件，由此可得到怎样的遗传规律？

（2）记 $\hat{e}_i=y_i-\hat{y}_i=y_i-\hat{b}x_i-\hat{a},(i=1,2,\cdots,n)$，其中 y_i 为观测值，$\hat{y}_i$ 为预测值，$\hat{e}_i$ 为对应 (x_i,y_i) 的残差.求（1）中儿子身高的残差的和、并探究这个结果是否对任意具有线性相关关系的两个变量都成立？若成立加以证明；若不成立说明理由.

参考数据及公式： $\sum_{i=1}^{5}x_i=880,\sum_{i=1}^{5}x_i^2=155450,\sum_{i=1}^{5}y_i=885,\sum_{i=1}^{5}x_iy_i=156045$，

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

2023年高考数学考前复习材料——新教材中的新增知识查漏（3）

【解】（1）由题意得 $\overline{x}=\dfrac{160+170+175+185+190}{5}=176,\overline{y}=\dfrac{170+174+175+180+186}{5}=177$，

$$\hat{b}=\frac{\sum_{i=1}^{5}x_iy_i-5\overline{xy}}{\sum_{i=1}^{5}x_i^2-5\overline{x}^2}=\frac{156045-5\times176\times177}{155450-5\times176^2}=\frac{156045-155760}{155450-154880}=\frac{285}{570}=0.5,$$

$\hat{a}=\overline{y}-\hat{b}\overline{x}=177-0.5\times176=89$，所以回归直线方程为 $y=0.5x+89$，

令 $0.5x+89-x>0$ 得 $x<178$，即 $x<178$ 时，儿子比父亲高；

令 $0.5x-89-x<0$ 得 $x>178$，即 $x>178$ 时，儿子比父亲矮，

可得当父亲身高较高时，儿子平均身高要矮于父亲，即儿子身高有一个回归，

回归到全种群平均高度的趋势.

（2）由 $y=0.5x+89$ 可得 $y_1=0.5\times160+89=169,y_2=174,y_3=176.5,y_4=181.5,y_5=184$，

所以 $\sum_{i=1}^{5}\hat{y}_i=885$，

又 $\sum_{i=1}^{5}y_i=885$，所以 $\sum_{i=1}^{5}\hat{e}_i=\sum_{i=1}^{5}\left(y_i-\hat{y}_i\right)=\sum_{i=1}^{5}y_i-\sum_{i=1}^{5}\hat{y}_i=0$，

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

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

例 5.（2023 山东淄博一模）随着科技进步，近来年，新能源汽车产业迅速发展．以下是汽车工业协会 2023 年 2 月公布的近六年新能源乘用车的年销售量数据：

年份	2017	2018	2019	2020	2021	2022
年份代码 x	1	2	3	4	5	6
新能源乘用车年销售 y（万辆）	50	78	126	121	137	352

（1）根据表中数据，求出 y 关于 x 的线性回归方程；（结果保留整数）

（2）若用 $y=me^{nx}$ 模型拟合 y 与 x 的关系，可得回归方程为 $y=37.71e^{0.33x}$，经计算该模型和第（1）问中模型的 R^2（R^2 为相关指数）分别为 0.87 和 0.71，请分别利用这两个模型，求 2023 年新能源乘用车的年销售量的预测值；

（3）你认为（2）中用哪个模型得到的预测值更可靠？请说明理由．

参考数据：设 $u=\ln y$，其中 $u_i=\ln y_i$．

$\overline{y}$	$\overline{u}$	$\sum_{i=1}^{6}\left(x_i-\overline{x}\right)\left(y_i-\overline{y}\right)$	$\sum_{i=1}^{6}\left(x_i-\overline{x}\right)\left(u_i-\overline{u}\right)$	$e^{3.63}$	$e^{5.94}$	$e^{6.27}$
144	4.78	841	5.70	37.71	380	528

参考公式：对于一组具有线性相关关系的数据 $(x_i, y_i)(i=1,2,3,\cdots,n)$，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公式分别为 $\hat{b}=\dfrac{\sum_{i=1}^{n}\left(x_i-\overline{x}\right)\left(y_i-\overline{y}\right)}{\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2}$，$\hat{a}=\overline{y}-\hat{b}\overline{x}$．

【解】（1）由表中数据得，$\bar{x}=\dfrac{1+2+3+4+5+6}{6}=3.5$，$\bar{y}=144$，$\sum_{i=1}^{6}\left(x_i-\bar{x}\right)\left(y_i-\bar{y}\right)=841$，

$$\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2=\left(x_1-\bar{x}\right)^2+\left(x_2-\bar{x}\right)^2+\left(x_3-\bar{x}\right)^2+\left(x_4-\bar{x}\right)^2+\left(x_5-\bar{x}\right)^2+\left(x_6-\bar{x}\right)^2$$

$$=\left(1-3.5\right)^2+\left(2-3.5\right)^2+\left(3-3.5\right)^2+\left(4-3.5\right)^2+\left(5-3.5\right)^2+\left(6-3.5\right)^2$$

$$=17.5$$

$$\therefore \hat{b}=\frac{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)\left(y_i-\bar{y}\right)}{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}=\frac{841}{17.5}\approx 48,\quad \hat{a}=\bar{y}-\hat{b}\bar{x}=144-48\times 3.5=-24$$

$\therefore$ y 关于 x 的线性回归方程为：$\hat{y}=48x-24$.

（2）由（1）知，y 关于 x 的线性回归方程为：$\hat{y}=48x-24$，

当 $x=7$ 时，2023 年新能源乘用车的年销售量的预测值：

$\hat{y}=48\times 7-24=312$（万辆）；

对于回归方程 $y=37.71\mathrm{e}^{0.33x}$，

当 $x=7$ 时，2023 年新能源乘用车的年销售量的预测值：

$y=37.71\mathrm{e}^{0.33\times 7}=\mathrm{e}^{3.63}\times \mathrm{e}^{2.31}=\mathrm{e}^{5.94}=380$（万辆）.

（3）依题意：$y=37.71\mathrm{e}^{0.33x}$ 模型和第（1）问中模型的 R^2（R^2 为相关指数）分别为 0.87 和 0.71，由于相关指数越接近于1，两个变量之间的关系就强，相应的拟合程度也越好，所以 $y=37.71\mathrm{e}^{0.33x}$ 模型得到的预测值更可靠.

【总结提炼】

线性回归分析的原理、方法和步骤：

（1）利用图表和数字特征可以对数据做简单的分析，但是用回归直线方程可以对数据的未来值进行预测．在选取数据观察的时候，要注意大量相对稳定的数据比不稳定的数据更有价值，近期的数据比过去久远的数据更有价值.

（2）判断两组数据是否具有线性相关关系的方法：散点图，相关系数.

（3）相关指数 R^2 与相关系数 r 在含有一个解释变量的线性回归模型中是等价的量 $(R^2=r^2)$，都是用来判断线性回归模型拟合效果好不好的量.

【全概率公式】

例 6.（2023 浙江高三开学考）设甲乘汽车、动车前往某目的地的概率分别为 $0.4, 0.6$，汽车和动车正点到达目的地的概率分别为 $0.7, 0.9$，则甲正点到达目的地的概率为（　　）

A．0.78　　B．0.8　　C．0.82　　D．0.84

【答案】C

【解析】设事件 A 表示甲正点到达目的地，事件 B 表示甲乘动车到达目的地，事件 C 表示甲乘汽车到达目的地，

由题意知 $P(B)=0.6, P(C)=0.4, P(A|B)=0.9, P(A|C)=0.7$.

由全概率公式得 $P(A)=P(B)P(A|B)+P(C)P(A|C)=0.6\times 0.9+0.4\times 0.7=0.28+0.54=0.82$.

故选：C

【变式演练】

（2023 福建统考一模）校园师生安全重于泰山，越来越多的学校纷纷引进各类急救设备．某

学校引进 M，N 两种类型的自动体外除颤器（简称 AED）若干，并组织全校师生学习 AED 的使用规则及方法．经过短期的强化培训，在单位时间内，选择 M，N 两种类型 AED 操作成功的概率分别为 $\frac{2}{3}$ 和 $\frac{1}{2}$，假设每次操作能否成功相互独立．

（1）现有某受训学生进行急救演练，假定他每次随机等可能选择 M 或 N 型 AED 进行操作，求他恰好在第二次操作成功的概率；

（2）为激发师生学习并正确操作 AED 的热情，学校选择一名教师代表进行连续两次设备操作展示，下面是两种方案：

方案甲：在第一次操作时，随机等可能的选择 M 或 N 型 AED 中的一种，若第一次对某类型 AED 操作成功，则第二次继续使用该类型设备；若第一次对某类型 AED 操作不成功，则第二次使用另一类型 AED 进行操作．

方案乙：在第一次操作时，随机等可能的选择 M 或 N 型 AED 中的一种，无论第一次操作是否成功，第二次均使用第一次所选择的设备．

假定方案选择及操作不相互影响，以成功操作累积次数的期望值为决策依据，分析哪种方案更好？

【解】（1）设“操作成功”为事件 S，“选择设备 M”为事件 A，“选择设备 N”为事件 B

由题意， $P(A)=P(B)=\frac{1}{2},P(S|A)=\frac{2}{3},P(S|B)=\frac{1}{2}$

恰在第二次操作才成功的概率 $P=P(\bar{S})P(S)$，

$P(S)=P(A)P(S\mid A)+P(B)P(S\mid B)=\frac{1}{2}\times\frac{2}{3}+\frac{1}{2}\times\frac{1}{2}=\frac{7}{12}$，

$P(\bar{S})=1-P(S)=\frac{5}{12}$

所以恰在第二次操作才成功的概率为 $\frac{5}{12}\times\frac{7}{12}=\frac{35}{144}$．

（2）设方案甲和方案乙成功操作累计次数分别为 X，Y，则 X，Y 可能取值均为 0，1，2，

$P(X=0)=P(A)P(\bar{S}\mid A)P(\bar{S}\mid B)+P(B)P(\bar{S}\mid B)P(\bar{S}\mid A)$

$=\frac{1}{2}\times\left(1-\frac{2}{3}\right)\times\left(1-\frac{1}{2}\right)+\frac{1}{2}\times\left(1-\frac{1}{2}\right)\times\left(1-\frac{2}{3}\right)=\frac{1}{6}$；

$P(X=1)=P(A)P(\bar{S}\mid A)P(S\mid B)+P(A)P(S\mid A)P(\bar{S}\mid A)$

$+P(B)P(\bar{S}\mid B)P(S\mid A)+P(B)P(S\mid B)P(\bar{S}\mid B)$

$=\frac{1}{2}\times\left(1-\frac{2}{3}\right)\times\frac{1}{2}+\frac{1}{2}\times\frac{2}{3}\times\left(1-\frac{2}{3}\right)+\frac{1}{2}\times\left(1-\frac{1}{2}\right)\times\frac{2}{3}+\frac{1}{2}\times\frac{1}{2}\times\left(1-\frac{1}{2}\right)=\frac{35}{72}$；

$P(X=2)=P(A)P(S\mid A)P(S\mid A)+P(B)P(S\mid B)P(S\mid B)$

$=\frac{1}{2}\times\frac{2}{3}\times\frac{2}{3}+\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}=\frac{25}{72}$；

所以 $E(X)=0\times\frac{1}{6}+1\times\frac{35}{72}+2\times\frac{25}{72}=\frac{85}{72}$

方法一：$P(Y=0)=P(A)P(\bar{S}\mid A)P(\bar{S}\mid A)+P(B)P(\bar{S}\mid B)P(\bar{S}\mid B)$

$=\frac{1}{2}\times\left(1-\frac{2}{3}\right)\times\left(1-\frac{2}{3}\right)+\frac{1}{2}\times\left(1-\frac{1}{2}\right)\times\left(1-\frac{1}{2}\right)=\frac{13}{72}$；

$P(Y=1)=P(A)P(\overline{S}\mid A)P(S\mid A)+P(A)P(S\mid A)P(\overline{S}\mid A)$

$+P(B)P(\overline{S}\mid B)P(S\mid B)+P(B)P(S\mid B)P(\overline{S}\mid B)$

$=\frac{1}{2}\times\left(1-\frac{2}{3}\right)\times\frac{2}{3}+\frac{1}{2}\times\frac{2}{3}\times\left(1-\frac{2}{3}\right)+\frac{1}{2}\times\left(1-\frac{1}{2}\right)\times\frac{1}{2}+\frac{1}{2}\times\frac{1}{2}\times\left(1-\frac{1}{2}\right)=\frac{17}{36}$

$P(Y=2)=P(A)P(S\mid A)P(S\mid A)+P(B)P(S\mid B)P(S\mid B)$

$=\frac{1}{2}\times\frac{2}{3}\times\frac{2}{3}+\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}=\frac{25}{72}$；

所以$E\left(Y\right)=0\times\frac{13}{72}+1\times\frac{17}{36}+2\times\frac{25}{72}=\frac{7}{6}$

方法二：方案乙选择其中一种操作设备后，进行2次独立重复试验，

所以$E(Y)=\frac{1}{2}\times2\times\frac{2}{3}+\frac{1}{2}\times2\times\frac{1}{2}=\frac{7}{6}$，

决策一：因为$E(X)>E(Y)$，故方案甲更好．

决策二：因为$E(X)$与$E(Y)$差距非常小，所以两种方案均可．

【总结提炼】

全概率公式$P(B)=\sum_{i=1}^{n}P(A_i)P(B\mid A_i)$在解题中体现了“化整为零、各个击破”的转化思想，可将较为复杂的概率计算分解为一些较为容易的情况分别进行考虑．

（一）全概率公式

（1）$P(B)=P(A)P(B\mid A)+P(\overline{A})P(B\mid\overline{A})$；

（2）**定理**1若样本空间Ω中的事件A_1，A_2，…，A_n满足：

①任意两个事件均互斥，即$A_iA_j=\varnothing$，i，$j=1,2,\cdots,n$，$i\neq j$；

②$A_1+A_2+\cdots+A_n=\Omega$；

③$P\left(A_i\right)>0$，$i=1,2,\cdots,n$．

则对Ω中的任意事件B，都有$B=BA_1+BA_2+\cdots+BA_n$，且

$P(B)=\sum_{i=1}^{n}P(BA_i)=\sum_{i=1}^{n}P(A_i)P(B\mid A_i)$．

注意：（1）全概率公式是用来计算一个复杂事件的概率，它需要将复杂事件分解成若干简单事件的概率计算，即运用了“化整为零”的思想处理问题．

（2）什么样的问题适用于这个公式？所研究的事件试验前提或前一步骤试验有多种可能，在这多种可能中均有所研究的事件发生，这时要求所研究事件的概率就可用全概率公式．

****（二）贝叶斯公式**

（1）一般地，当$0<P(A)<1$且$P(B)>0$时，有

$$P(A|B)=\frac{P(A)P(B\mid A)}{P(B)}=\frac{P(A)P(B\mid A)}{P(A)P(B\mid A)+P(\overline{A})P(B\mid\overline{A})}$$

（2）定理2若样本空间Ω中的事件$A_1,A_2,\cdots,A_n$满足：

①任意两个事件均互斥，即$A_iA_j=\varnothing$，i，$j=1,2,\cdots,n$，$i\neq j$；

②$A_1+A_2+\cdots+A_n=\Omega$；

③$0<P\left(A_i\right)<1$，$i=1,2,\cdots,n$．则对Ω中的任意概率非零的事件B，都有$B=BA_1+BA_2+\cdots+BA_n$，

且$P(A_j|B)=\frac{P(A_j)P(B\mid A_j)}{P(B)}=\frac{P(A_j)P(B\mid A_j)}{\sum_{i=1}^{n}P(A_i)P(B\mid A_i)}$

2023年高考数学考前复习材料——新教材中的新增知识查漏（4）

注意：（1）在理论研究和实际中还会遇到一类问题，这就是需要根据试验发生的结果寻找原因，看看导致这一试验结果的各种可能的原因中哪个起主要作用，解决这类问题的方法就是使用贝叶斯公式．贝叶斯公式的意义是导致事件B发生的各种原因可能性的大小，称之为后验概率．

（2）贝叶斯公式充分体现了$P(A|B)$，$P(A)$，$P(B)$，$P(B|A)$，$P(B|\overline{A})$，$P(AB)$之间的转关系，即$P(A|B)=\dfrac{P(AB)}{P(B)}$，$P(AB)=P(A|B)P(B)=P(B|A)P(A)$，$P(B)=P(A)P(B|A)+P(\overline{A})P(B|\overline{A})$之间的内在联系．

【巩固练习】

1．（2023南京盐城一模）ΔABC中，AH为BC边上的高，且$\overrightarrow{BH}=3\overrightarrow{HC}$，动点$P$满足$\overrightarrow{AP}\cdot\overrightarrow{BC}=-\dfrac{1}{4}\overrightarrow{BC}^2$，则点$P$的轨迹一定过$\Delta ABC$的（　　）

A．外心　　B．内心　　C．垂心　　D．重心

【答案】A

【解析】法一：设动点P在$\overrightarrow{BC}$的投影为D，则由AH为BC边上的高，得$\overrightarrow{AP}$在$\overrightarrow{BC}$的投影为$\overrightarrow{HD}$，结合$\overrightarrow{AP}\cdot\overrightarrow{BC}=-\dfrac{1}{4}\overrightarrow{BC}^2$，得$\overrightarrow{HD}\cdot\overrightarrow{BC}=-\dfrac{1}{4}\overrightarrow{BC}^2$，因为$\overrightarrow{BH}=3\overrightarrow{HC}$，所以$D$是$BC$的中点，所以点$P$的轨迹一定过$\Delta ABC$的外心，选择A．

法二：建系求轨迹方程．

2．为庆祝中国共产党成立100周年，深入推进党史学习教育，引导干部学史明理、学史增信、学史崇德、学史力行，某中学党支部组织学校初、高中两个学部的党员参加了全省教育系统的党史知识竞赛活动，其中初中部20名党员竞赛成绩的平均分为a，方差为2；高中部50名党员竞赛成绩的平均分为b，方差为$\dfrac{12}{5}$．若$a=b$，则该学校全体参赛党员竞赛成绩的方差为（　）

A．$\dfrac{16}{7}$　　B．$\dfrac{21}{10}$　　C．$\dfrac{43}{20}$　　D．$\dfrac{33}{14}$

【答案】A

【解析】设初中部20名的党员竞赛成绩分别为x_1，x_2，x_3，$\cdots$，x_{20}，高中部50名的党员竞赛成绩分别为y_1，y_2，y_3，$\cdots$，y_{50}，则结合题设，得$\dfrac{(x_1-a)^2+(x_2-a)^2+\cdots+(x_{20}-a)^2}{20}=2$，$\dfrac{(y_1-b)^2+(y_2-b)^2+\cdots+(y_{50}-b)^2}{50}=\dfrac{12}{5}$，所以$(x_1-a)^2+(x_2-a)^2+\cdots+(x_{20}-a)^2=40$。$(y_1-b)^2+(y_2-b)^2+\cdots+(y_{50}-b)^2=120$．因为$a=b$，所以该学校全体参赛党员竞赛成绩的平均分为$a$，所以该学校全体参赛党员竞赛成绩的方差为$\dfrac{1}{70}[(x_1-a)^2+(x_2-a)^2+\cdots+(x_{20}-a)^2+(y_1-b)^2+(y_2-b)^2+\cdots+(y_{50}-b)^2]=\dfrac{40+120}{70}=\dfrac{16}{7}$，故选A．

3．（2022湖南岳阳三模）中国青年志愿者协会成立于1994年12月5日，此后广大志愿者、志愿服务组织不断蓬勃发展，目前高校青年志愿者组织就有132个.为了解某大学学生参加志愿者工作的情况，随机抽取某高校志愿者协会的40名成员，就他们2022年第2季度参加志愿服务的次数进行了统计，数据如表所示.则这40名学生本季度参加志愿活动的第40百

分数位为（　）

次数	7	8	9	10	11
人数	6	10	9	8	7

A．9　　B．8　　C．8.5　　D．9.5

【答案】C

【解析】∵ $40\times 0.4=16$ 为整数

∴第 40 百分数位为第 16 位和第 17 位的平均数，即为 $\frac{8+9}{2}=8.5$，故选：C.

4．（多选）（2023 山东济南一模）某公司为了解用户对其产品的满意度，随机调查了 10 个用户，得到用户对产品的满意度评分如下表所示，评分用区间[0，10]内的一个数来表示，该数越接近 10 表示满意度越高，则下列说法正确的是（　）

7	8	9	7	5	4	10	9	4	7

A．这组数据的平均数为 6　　B．这组数据的众数为 7

C．这组数据的极差为 6　　D．这组数据的 75%分位数为 9

【答案】BCD

【分析】由平均数、众数、极差、百分位数的定义即可得出答案.

【解析】这组数从小到大排列为：4，4，5，7，7，7，8，9，9，10，

计算这组数据的平均数为 $\frac{1}{10}\times(4+4+5+7+7+7+8+9+9+10)=7$，选项 A 错误；

这组数据的众数是 7，选项 B 正确；

这组数据的极差是 $10-4=6$，选项 C 正确；

因为 10×75%=7.5，且第 8 个数是 9，所以这组数据的 75%分位数为 9，选项 D 正确.

故选：BCD.

5．（2023 河北衡水一模）2022 年 10 月 1 日，女篮世界杯落幕，时隔 28 年，中国队再次获得亚军，追平历史最佳成绩．为考察某队员甲对球队的贡献，教练对近两年甲参加过的 100 场比赛进行统计：甲在前锋位置出场 20 次，其中球队获胜 14 次；中锋位置出场 30 次，其中球队获胜 21 次；后卫位置出场 50 次，其中球队获胜 40 次．用该样本的频率估计概率，则甲参加比赛时，该球队某场比赛获胜的概率________.

【答案】0.75

【解析】设 $A_1=$“甲担任前锋”； $A_2=$“甲担任中锋”； $A_3=$“甲担任后卫”； $B=$“某场比赛中该球队获胜”；则 $P\left(A_1\right)=\frac{20}{100}=0.2$， $P\left(A_2\right)=\frac{30}{100}=0.3$， $P\left(A_3\right)=\frac{50}{100}=0.5$， $P\left(B\mid A_1\right)=\frac{14}{20}=0.7$， $P\left(B\mid A_2\right)=\frac{21}{30}=0.7$，

$P\left(B\mid A_3\right)=\frac{40}{50}=0.8$，由全概率公式可得：

$P(B)=P\left(A_1\right)P\left(B\mid A_1\right)+P\left(A_2\right)P\left(B\mid A_2\right)+P\left(A_3\right)P\left(B\mid A_3\right)=0.2\times0.7+0.3\times0.7+0.5\times0.8=0.75$.

所以甲参加比赛时，该球队某场比赛获胜的概率是 0.75.

6．由于身体及心理方面的差异，人们往往认为女性驾驶员比男性驾驶员更容易发生交通事故.为调查女性驾驶员是否比男性驾驶员更容易发生交通事故，橙子辅导的同学组成了调查小组，对其所在城市进行了调查研究，结果却显示为：该市 2021 年男女驾驶员的比例为 $7:3$，男性驾驶员平均万人的发案率为 2.20，女性驾驶员平均万人的发案率为 0.25.（发案即发生

了交通事故，暂不区分其是否为肇事责任人）.

（1）若在全市驾驶员中随机抽取 3 人，则恰有 1 位女驾驶员的概率是多少？

（2）若该市一名驾驶员在 2021 年发生了交通事故，则其为女性的概率是多少？（结果保留到小数点后第三位）.

【分析】

（1）设在全市驾驶员中随机抽取 3 人，女驾驶员的人数为 X，则 $X\sim B(3,0.3)$，再根据二项分布的概率公式求解即可；

（2）设事件 A：驾驶员为女性，事件 B：驾驶员发生的交通事故，进而结合全概率公式和条件概率公式求解即可.

【解】（1）因为该市 2021 年男女驾驶员的比例为 $7:3$，

所以，在全市驾驶员中随机抽取 1 人是女驾驶员的概率为 0.3，

设在全市驾驶员中随机抽取 3 人，女驾驶员的人数为 X，

所以，$X\sim B(3,0.3)$

所以，恰有 1 位女驾驶员的概率是 $P(X=1)=C_3^1\times0.3\times0.7^2=0.441$.

（2）设事件 A：驾驶员为女性，事件 B：驾驶员发生的交通事故.

所以 $P(A)=0.3$，$P(\overline{A})=0.7$，$P(B|A)=0.25\times10^{-4}$，$P(B|\overline{A})=2.20\times10^{-4}$,

所以，根据全概率公式，

$$P(B)=P(A)P(B|A)+P(\overline{A})P(B|\overline{A})=0.075\times10^{-4}+2.2\times0.7\times10^{-4}=1.615\times10^{-4},$$

所以，$P(A|B)=\dfrac{P(AB)}{P(B)}=\dfrac{P(A)P(B|A)}{P(A)P(B|A)+P(\overline{A})P(B|\overline{A})}=\dfrac{0.075\times10^{-4}}{1.615\times10^{-4}}\approx0.046.$

所以，该市一名驾驶员在 2021 年发生了交通事故，则其为女性　概率是多少 0.046.

7.（2023 河南郑州一模）自主创新是我国经济发展的核心动力，科技自立自强已被赋予国家发展战略支点的功能．目前实现科技自立自强我们仍面临巨大挑战，越来越多的企业主动谋划、加快发展，推动我国科技创新迈上新台阶．某企业拟对某芯片进行科技升级，根据市场调研与模拟，得到科技升级投入 x（亿元）与科技升级直接收益 y（亿元）的数据统计如下：

序号	1	2	3	4	5	6	7
x	2	3	4	6	8	10	13
y	13	22	31	42	50	56	58

根据表格中的数据，建立了 y 与 x 的两个回归模型：模型①：$\hat{y}=4.1x+11.8$；模型②：$\hat{y}=21.3\sqrt{x}-14.4$.

（1）根据下列表格中的数据，比较模型①、②的相关指数的大小，并选择拟合精度更高、更可靠的模型；

（2）根据（1）选择的模型，预测对芯片科技升级的投入为 17 亿元时的直接收益.

回归模型	模型①	模型②
回归方程	$\hat{y}=4.1x+11.8$	$\hat{y}=21.3\sqrt{x}-14.4$
$\sum_{i=1}^{7}(y_i-\hat{y}_i)^2$	182.4	79.2

（**附**：刻画回归效果的相关指数 $R^2=1-\dfrac{\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i-\overline{y}\right)^2}$，$\sqrt{17}\approx 4.1$）

【解】(1）由表格中的数据，182.4＞79.2，

$\therefore \dfrac{182.4}{\sum_{i=1}^{7}(y_i-\overline{y})^2}>\dfrac{79.2}{\sum_{i=1}^{7}(y_i-\overline{y})^2}$，$1-\dfrac{182.4}{\sum_{i=1}^{7}(y_i-\overline{y})^2}<1-\dfrac{79.2}{\sum_{i=1}^{7}(y_i-\overline{y})^2}$，

∴模型①的相关指数小于模型②的相关指数，∴回归模型②的拟合效果更好

（2）当 x＝17 亿时，科技升级直接收益的预测值为：$\hat{y}=21.3\sqrt{x}-14.4\approx 72.93(亿元)$.

8.（2023 山东烟台一模）我国5G技术给直播行业带来了很多发展空间，加上受疫情影响，直播这种成本较低的获客渠道备受商家青睐，某商场统计了 2022 年 1~5 月某商品的线上月销售量 y（单位：千件）与售价 x（单位：元/件）的情况如下表示.

月份	1	2	3	4	5
售价 x（元/件）	60	56	58	57	54
月销售量 y（千件）	5	9	7	10	9

（1）求相关系数 r，并说明是否可以用线性回归模型拟合 y 与 x 的关系（当 $|r|\in[0.75,1]$ 时，可以认为两个变量有很强的线性相关性；否则，没有很强的线性相关性）（精确到 0.01）;

（2）建立 y 关于 x 的线性回归方程，并估计当售价为 55 元/件时，该商品的线上月销售量估计为多少千件？

（3）若每件商品的购进价格为 $(0.5x+25)$ 元/件，如果不考虑其他费用，由（2）中结论，当商品售价为多少时，可使得该商品的月利润最大？（该结果保留整数）

参考公式：对于一组数据 $\left(x_i,y_i\right)(i=1,2,3,\cdots,n)$，相关系数 $r=\dfrac{\sum_{i=1}^{n}\left(x_i-\overline{x}\right)\left(y_i-\overline{y}\right)}{\sqrt{\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2}\sqrt{\sum_{i=1}^{n}\left(y_i-\overline{y}\right)^2}}$，$\hat{b}=\dfrac{\sum_{i=1}^{n}\left(x_i-\overline{x}\right)\left(y_i-\overline{y}\right)}{\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2},\hat{a}=\overline{y}-\hat{b}\overline{x}$.

参考数据：$\sqrt{5}\approx 2.236$.

【解】(1）由已知数据可得 $\overline{x}=\dfrac{60+56+58+57+54}{5}=57,\overline{y}=\dfrac{5+9+7+10+9}{5}=8$，

$\sqrt{\sum_{i=1}^{5}\left(x_i-\overline{x}\right)^2}=\sqrt{3^2+(-1)^2+1^2+0^2+(-3)^2}=2\sqrt{5}$，$\sqrt{\sum_{i=1}^{5}\left(y_i-\overline{y}\right)^2}=\sqrt{(-3)^2+1^2+(-1)^2+2^2+1^2}=4$，

$\sum_{i=1}^{5}\left(x_i-\overline{x}\right)\left(y_i-\overline{y}\right)=3\times(-3)+(-1)\times 1+1\times(-1)+0\times 2+(-3)\times 1=-14$，

所以相关系数 $r=\dfrac{\sum_{i=1}^{5}\left(x_i-\overline{x}\right)\left(y_i-\overline{y}\right)}{\sqrt{\sum_{i=1}^{5}\left(x_i-\overline{x}\right)^2}\sqrt{\sum_{i=1}^{5}\left(y_i-\overline{y}\right)^2}}=\dfrac{-14}{2\sqrt{5}\times 4}\approx -0.78$，

因为 $|r|>0.75$，所以 y 与 x 有很强的线性相关性，可以用线性回归模型拟合.

（2）由于 $\hat{b}=\dfrac{\sum_{i=1}^{5}\left(x_i-\overline{x}\right)\left(y_i-\overline{y}\right)}{\sum_{i=1}^{5}\left(x_i-\overline{x}\right)^2}=\dfrac{-14}{20}=-0.7$，$a=\overline{y}-\hat{b}\overline{x}=8-(-0.7)\times 57=47.9$，

所以 y 关于 x 的线性回归方程为 $\hat{y}=-0.7x+47.9$，当 $x=55$ 时，$\hat{y}=-0.7\times 55+47.9=9.4$，

故当售价为 $x=55$ 元/件时，该商品的线上月销售量估计为 9.4 千件.

（3）设每月的利润为 Z 元，则 $Z=1000(x-0.5x-25)(-0.7x+47.9)=50\left(-7x^2+829x-23950\right)$，

当 $x=\dfrac{829}{14}\approx 59$ 时，Z 取得最大值．即当商品售价为 59 元/件时，可使得该商品的月利润最大.

2023 年高考数学考前复习材料——新教材中的新增知识查漏（5）

【备用题】

1．（2023 湖南长沙一模）据一组样本数据$(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)$，求得经验回归方程为$y=1.2x+0.4$，且$\bar{x}=3$．现发现这组样本数据中有两个样本点$(1.2,0.5)$和$(4.8,7.5)$误差较大，去除后重新求得的经验回归直线$l$的斜率为$1.1$，则（　　）

A．去除两个误差较大的样本点后，y的估计值增加速度变快

B．去除两个误差较大的样本点后，重新求得的回归方程对应直线一定过点$(3,5)$

C．去除两个误差较大的样本点后，重新求得的回归方程为$y=1.1x+0.7$

D．去除两个误差较大的样本点后，相应于样本点$(2,2.7)$的残差为 0.1

【答案】C

【解析】对于 A，因为去除两个误差较大的样本点后，经验回归直线l的斜率变小，则y的估计值增加速度变慢，A 错误；对于 B，由$y=1.2x+0.4$及$\bar{x}=3$得：$\bar{y}=4$，因为去除的两个样本点$(1.2,0.5)$和$(4.8,7.5)$，并且$\frac{1.2+4.8}{2}=3,\frac{0.5+7.5}{2}=4$，因此去除两个样本点后，样本的中心点仍为$(3,4)$，因此重新求得的回归方程对应直线一定过点$(3,4)$，B 错误；对于 C，设去除后重新求得的经验回归直线l的方程为$y=1.1x+\hat{a}$，由选项 B 知，$4=1.1\times3+\hat{a}$，解得$\hat{a}=0.7$，所以重新求得的回归方程为$y=1.1x+0.7$，C 正确；对于 D，由选项 C 知，$y=1.1x+0.7$，当$x=2$时，$y=1.1\times2+0.7=2.9$，则$2.7-2.9=-0.2$，因此去除两个误差较大的样本点后，相应于样本点$(2,2.7)$的残差为-0.2，D 错误．故选：C.

2．（2023 重庆模拟）（多选）小明在家独自用下表分析高三前 5 次月考中数学的班级排名y与考试次数x的相关性时，忘记了第二次和第四次月考排名，但小明记得平均排名$\bar{y}=6$，于是分别用 m＝6 和 m＝8 得到了两条回归直线方程：$y=b_1x+a_1$，$y=b_2x+a_2$，对应的相关系数分别为r_1、r_2，排名y对应的方差分别为s_1^2、s_2^2，则下列结论正确的是（　　）

x	1	2	3	4	5
y	10	m	6	n	2

（附：$b=\frac{\sum_{i=1}^{n}x_iy_i-n\bar{x}\bar{y}}{\sum_{i=1}^{n}x_i^2-n\left(\bar{x}\right)^2}$，$a=\bar{y}-b\bar{x}$）

A．$r_1<r_2$　　　B．$s_1^2<s_2^2$　　　C．$b_1<b_2$　　　D．$a_1<a_2$

【答案】BD

【解析】当$m=6$时，$\bar{x}=\frac{1+2+3+4+5}{5}=3,\bar{y}=\frac{10+6+6+n_1+2}{5}=6$，解得$n_1=6$，

则$\sum_{i=1}^{5}x_iy_i=1\times10+2\times6+3\times6+4\times6+5\times2=74$，$\sum_{i=1}^{5}x_i^2=1^2+2^2+3^2+4^2+5^2=55$，$\bar{x}\bar{y}=18$，

$\sum_{i=1}^{5}(x_i-\bar{x})(y_i-\bar{y})=(1-3)(10-6)+(2-3)(6-6)+(3-3)(6-6)+(4-3)(6-6)+(5-3)(2-6)=-16$，

$\sum_{i=1}^{n}(x_i-\bar{x})^2(y_i-\bar{y})^2=(1-3)^2(10-6)^2+(2-3)^2(6-6)^2+(3-3)^2(6-6)^2+(4-3)^2(6-6)^2+(5-3)^2(2-6)^2=128$

所以$b_1=\frac{\sum_{i=1}^{n}x_iy_i-n\bar{x}\cdot\bar{y}}{\sum_{i=1}^{n}x_i^2-n\bar{x}^2}=\frac{74-5\times18}{55-5\times3^2}=-\frac{8}{5}$，得$a_1=\bar{y}-b_1\bar{x}=\frac{54}{5}$，$r_1=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2(y_i-\bar{y})^2}}=\frac{-16}{\sqrt{128}}=-\sqrt{2}$，

$s_1^2=\frac{1}{n}\sum_{i=1}^{n}(y_i-\overline{y})^2=\frac{(10-6)^2+(6-6)^2+(6-6)^2+(6-6)^2+(2-6)^2}{5}=\frac{32}{5}$；

同理，当$m=8$时，$b_2=-2,a_2=12,r_2=-\frac{5\sqrt{34}}{17}$，$s_2^2=8$，所以$r_1>r_2,s_1^2<s_2^2,b_1>b_2,a_1<a_2$，故选：BD.

3．（多选）（2022 福建福州一模）下列命题中，正确的命题有（　　）

A. 已知随机变量X服从正态分布$N(2,\sigma^2)$且$P(X<4)=0.9$，则$P(0<X<2)=0.3$

B. 设随机变量$X\sim B(20,\frac{1}{2})$，则$D(X)=5$

C. 在抛骰子试验中，事件$A=\{1,2,3,5,6\}$，事件$B=\{2,4,5,6\}$，则$P(A|B)=\frac{3}{5}$

D. 在线性回归模型中，R^2表示解释变量对于预报变量变化的贡献率，R^2越接近于 1，表示回归的效果越好

【答案】BD

【解析】A：因为$X\sim N(2,\sigma^2)$且$P(X<4)=0.9$，所以$P(X>4)=1-0.9=0.1$，

所以$P(0<X<2)=P(2<X<4)=0.5-P(X>4)=0.5-0.1=0.4$，A 错误；

B：因为$X\sim B(20,\frac{1}{2})$，所以$D(X)=np(1-p)=20\times\frac{1}{2}\times(1-\frac{1}{2})=5$，B 正确；

C：由题知，事件$AB=\{2,5,6\}$，所以$P(A|B)=\frac{P(AB)}{P(B)}=\frac{3}{4}$，C 错误；

D：由R^2的意义可知 D 正确.