考点巩固卷 21　统计与统计案例（七大考点）

考点预览

高端技巧及考点训练

考点 1 简单随机抽样

1．电影《孤注一掷》的上映引发了电信诈骗问题的热议，也加大了各个社区反电信诈骗的宣传力度.已知某社区共有居民 480 人，其中老年人 200 人，中年人 200 人，青少年 80 人，若按年龄进行分层随机抽样，共抽取 36 人作为代表，则中年人比青少年多（　）

A．6 人　　B．9 人　　C．12 人　　D．18 人

2．已知 A,B,C 三种不同型号的产品数量之比依次为 $4:3:7$，现用分层抽样的方法抽取容量为 N 的样本，若样本中 A 型号产品有 20 件，则 N 为（　）

A．60　　B．70　　C．80　　D．90

3．国内某优秀新能源电池制造企业在锂电池单位能量密度技术上取得了重大突破，该制造企业内的某车间有两条生产线，分别生产高能量密度锂电池和低能量密度锂电池，总产量为 400 个锂电池．质检人员采用分层随机抽样的方法随机抽取了一个容量为 80 的样本进行质量检测，已知样本中高能量密度锂电池有 35 个，则估计低能量密度锂电池的总产量为（　）.

A．325 个　　B．300 个　　C．225 个　　D．175 个

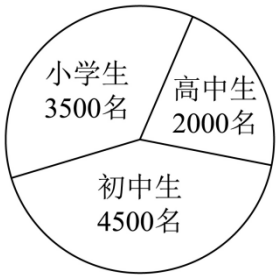
4．用按比例分配的分层随机抽样方法，从某学校的 600 名男生和 800 名女生中选取 14 人参与某项研学活动，则女生比男生多选取（　）

A．8 人　　B．6 人　　C．4 人　　D．2 人

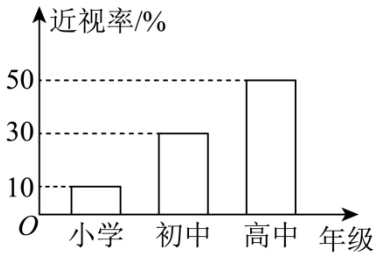
5．已知甲组数据：1，3，5，7，9，11，乙组数据：2，4，8，16，根据不同组别，用分层抽样的方法随机抽取一个容量为 5 的样本，则该样本的平均数不可能是（　）

A．5　　B．7　　C．9　　D．11

6．已知某地区中小学生人数和近视情况分别如图甲和图乙所示．为了了解该地区中小学生近视情况形成的原因，采用分层抽样的方法抽取部分学生进行调查，若抽取的小学生人数为 70，则抽取的高中生中近视人数为（　）



图甲



图乙

A．10　　B．20　　C．25　　D．40

7．为了检查某超市货架上的饮料是否含有塑化剂，要从编号依次为 1 到 100 的塑料瓶装饮料中抽取 5 瓶进行检验，用每部分选取的号码间隔一样的系统抽样方法确定所选取的 5 瓶饮料的编号可能是（　）

A．5，15，25，35，45　　B．10，25，40，55，70

C．10，20，30，40，50　　D．10，30，50，70，90

8．从一个含有 N 个个体的总体中抽取一容量为 n 的样本，当选取抽签法、随机数法和分层随机抽样三种不同方法时，总体中每个个体被抽中的概率分别为 p_1,p_2,p_3，三者关系可能是（　）

A．$p_1=p_2<p_3$　　B．$p_1=p_2=p_3$　　C．$p_1=p_3<p_2$　　D．$p_2=p_3<p_1$

9．下列说法中正确的个数有（　）

①对具有线性相关关系的变量 x，y，其回归方程为 $\hat{y}=0.3x-m$，若样本点的中心为 $(m,2.8)$，则实数 m 的值是 -4；

②某校共有学生 1003 人，用简单随机抽样的方法先剔除 3 人，再按简单随机抽样的方法抽取为 20 人，则每个学生被抽到的概率为 $\frac{1}{50}$；

③若随机事件 A，B 满足：$P(A)=\frac{1}{2}$，$P(B)=\frac{2}{3}$，$P(A\cup B)=\frac{5}{6}$，则事件 A 与 B 相互独立；

④若随机变量 ξ，η 满足 $\eta=2\xi+1$，则 $D(\eta)=2D(\xi)+1$.

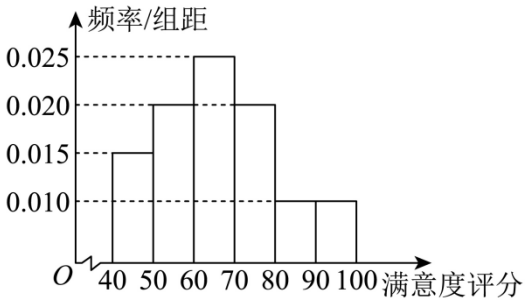
A．1　　B．2　　C．3　　D．4

10．为了解某校初中学生的近视情况，按年级用分层抽样的方法随机抽取 100 名学生进行视力检测，已知初一、初二、初三年级分别有 800 名，600 名，600 名学生，则不同的抽样结果共有（　）

A．$C_{800}^{40}\cdot\left(C_{600}^{30}\right)^2$　　B．$C_{800}^{30}\cdot C_{600}^{40}\cdot C_{600}^{30}$　　C．$\left(C_{800}^{30}\right)^2\cdot C_{600}^{40}$　　D．$\left(C_{800}^{40}\right)^2\cdot C_{600}^{20}$

考点 2　频率分布直方图

11．某公司为了解用户对其产品的满意度，从使用该产品的用户中随机调查了 100 个用户，根据用户对产品的满意度评分，得到如图所示的用户满意度评分的频率分布直方图：



根据此频率分布直方图，下面结论中不正确的是（　）

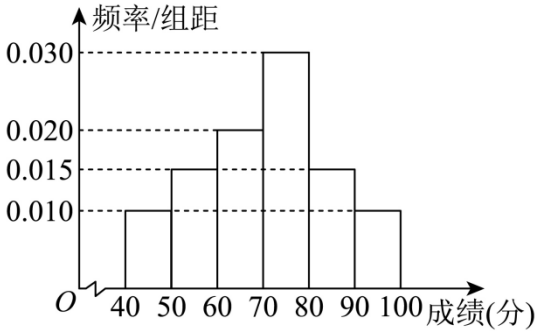
A．对该公司产品满意度评分低于 60 分的用户比例估计为 35%

B．对该公司产品满意度评分不低于 70 分的用户比例估计为 40%

C．估计该公司用户对产品的满意度评分的平均值不超过 60 分

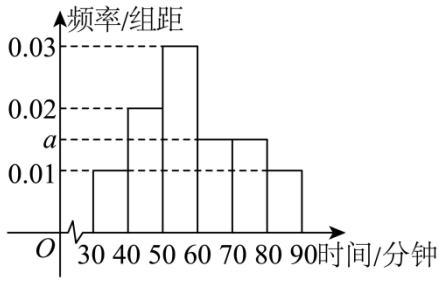
D．估计该公司有一半以上的用户，对产品的满意度评分介于 50 分至 80 分之间

12．在某次高中数学模拟考试中，对 800 名考生的考试成绩进行统计，得到如图所示的频率分布直方图，其中分组的区间分别为$[40,50)$，$[50,60)$，$[60,70)$，$[70,80)$，$[80,90)$，$[90,100]$.若考生成绩在$[70,80)$内的人数为m，考生成绩在$[80,100]$内的人数为n，则$m-n=$（　）



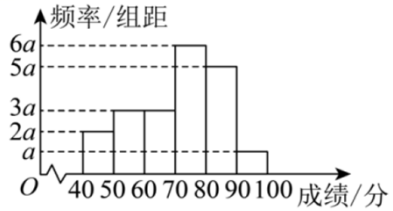
A．20　　B．10　　C．60　　D．40

13．为了解高中学生每天的体育活动时间,某市教育部门随机抽取1000高中学生进行调查,把每天进行体育活动的时间按照时长（单位：分钟）分成6组:$[30,40)$,$[40,50)$,$[50,60)$,$[60,70)$,$[70,80)$,$[80,90]$.然后对统计数据整理得到如图所示的频率分布直方图,则可估计这1000名学生每天体育活动时间的第25百分位数为（　）



A．47.5　　B．45.5　　C．43.5　　D．42.5

14．为了加深师生对党史的了解，激发广大师生知史爱党、知史爱国的热情，某校举办了“学党史、育新人”的党史知识竞赛，并将 1000 名师生的竞赛成绩（满分 100 分，成绩取整数）整理成如图所示的频率分布直方图，则下列说法错误的是（　）



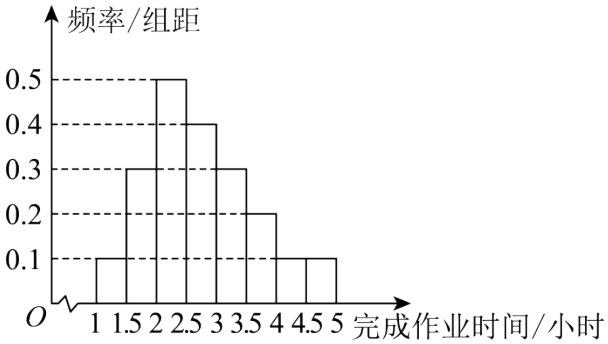
A．a 的值为 0.005

B．估计这组数据的众数为 75 分

C．估计成绩低于 60 分的有 250 人

D．估计这组数据的中位数为 $\frac{235}{3}$ 分

15．某教育机构为调查中小学生每日完成作业的时间，收集了某位学生 100 天每天完成作业的时间，并绘制了如图所示的频率分布直方图（每个区间均为左闭右开），根据此直方图得出了下列结论，其中正确的是（　）



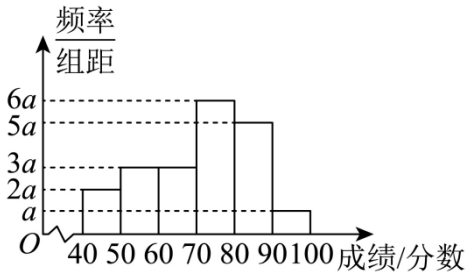
A．估计该学生每日完成作业的时间在 2 小时至 2.5 小时的有 50 天

B．估计该学生每日完成作业时间超过 3 小时的概率为 0.3

C．估计该学生每日完成作业时间的中位数为 2.625 小时

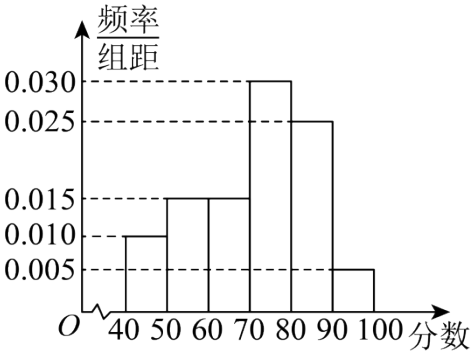
D．估计该学生每日完成作业时间的众数为 2.3 小时

16．为了加深师生对党史的了解，激发广大师生知史爱党、知史爱国的热情，某校举办了“学党史、育文化”的党史知识竞赛，并将 1000 名师生的竞赛成绩（满分 100 分，成绩取整数）整理成如图所示的频率分布直方图，估计这组数据的第 85 百分位数为（　）分



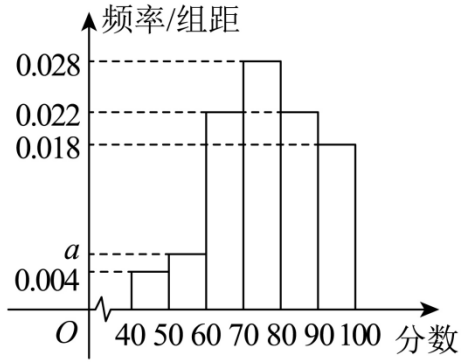
A．84　　B．85　　C．86　　D．87

17．某校高三共有 200 人参加体育测试，将体测得分情况进行了统计，把得分数据按照 $[40,50),[50,60),[60,70),[70,80),[80,90),[90,100]$ 分成 6 组，绘制了如图所示的频率分布直方图.根据规则，82 分以上的考生成绩等级为 A，则获得 A 的考生人数约为（　）



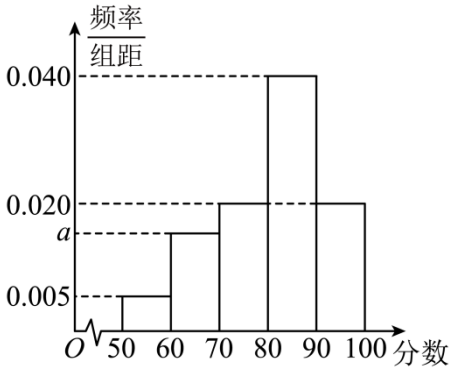
A．25　　B．50　　C．75　　D．100

18．为深入贯彻落实习近平总书记对天津工作“三个着力”重要要求，天津持续深化改革，创建全国文明城区，城市文明程度显著提升，人民群众的梦想不断实现.在创建文明城区的过程中，中央文明办对某小区居民进行了创建文明城区相关知识网络问卷调查，从本次问卷中随机抽取了 50 名居民的问卷结果，统计其得分数据，将所得 50 份数据的得分结果分为 6 组：$[40,50),[50,60),[60,70),[70,80),[80,90),[90,100]$，并整理得到如下的频率分布直方图，则该小区居民得分的第 70 百分位数为（　）



A．89.09　　B．86.52　　C．84.55　　D．81.32

19．某市为了解全市 12000 名高一学生的的体能素质情况，在全市高一学生中随机抽取了 1000 名学生进行体能测试，并将这 1000 名的体能测试成绩整理成如下频率分布直方图．根据此频率分布直方图，下列结论中正确的是（　）



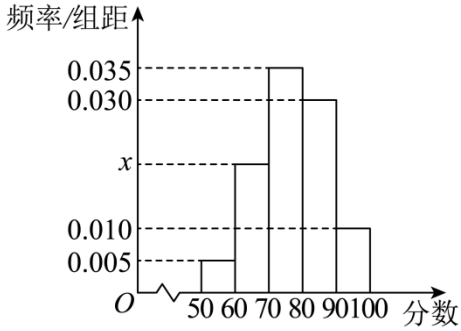
A．图中 a 的值为 0.020；

B．同一组中的数据用该组区间的中点值做代表，则这 1000 名学生的平均成绩约为 80.5；

C．估计样本数据的 75%分位数为 88；

D．由样本数据可估计全市高一学生体测成绩优异（80 分及以上）的人数约为 5000 人．

20．某校举行知识竞赛，对全校参赛的 1000 名学生的得分情况进行了统计，把得分数据按 $[50,60)$，$[60,70)$，$[70,80)$，$[80,90)$，$[90,100]$ 分成 5 组，得到如图所示的频率分布直方图，则下列说法不正确的是（　）



A．图中的 x 值为 0.020　　B．得分在 $[80,100]$ 的人数为 400

C．这组数据的极差为 50　　D．这组数据的平均数的估计值为 77

考点 3 均值及方差的性质

21．样本数据 $x_1,x_2,\cdots,x_n$ 的平均数 $\bar{x}=4$，方差 $s^2=1$，则样本数据 $2x_1+1$，$2x_2+1$，…，$2x_n+1$ 的平均数，方差分别为（　）

A．9，4　　B．9，2　　C．4，1　　D．2，1

22．若数据$x_1,x_2,x_3,\L,x_n$的标准差为s，则数据$3x_1+1$，$3x_2+1$，$3x_3+1$，…，$3x_n+1$的标准差为（　）

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

23．已知数据$x_1,x_2,x_3,\cdots,x_8$的平均数为10，方差为10，则$3x_1+2,3x_2+2,3x_3+2,\cdots,3x_8+2$的平均数和方差分别为（　）

A．32，90　　B．32，92　　C．30，90　　D．30，92

24．下列命题错误的是（　）

A．若数据$x_1,x_2,x_3,\cdots,x_n$的标准差为S，则数据$3x_1,3x_2,3x_3,\cdots,3x_n$的标准差为$3S$

B．若$X\sim B(4,p),D(X)=\frac{3}{4}$，则$P(X=2)=\frac{27}{128}$

C．若$X\sim N(1,\sigma^2),P(X>0)=0.75$，则$P(0<X<2)=0.5$

D．若X为取有限个值的离散型随机变量，则$\left[E(X)\right]^2\geq E(X^2)$

25．已知样本数据$x_1,x_2,\cdots,x_{100}$的平均数和标准差均为4，则数据$-x_1-1,-x_2-1,\cdots,-x_{100}-1$的平均数与方差分别为（　）

A．$-5,4$　　B．$-5,16$　　C．$4,16$　　D．$4,4$

26．已知一组数据x_1，x_2，x_3，x_4，x_5的平均数是2，方差是3，则对于以下数据：$2x_1+1$，$2x_2+1$，$2x_3+1$，$2x_4+1$，$2x_5+1$下列选项正确的是（　）

A．平均数是4，方差是6　　B．平均数是4，方差是7

C．平均数是5，方差是7　　D．平均数是5，方差是12

27．某人在“全球购”平台上购买了n件商品，这些商品的价格如果按美元计算，则平均数为A，标准差为s，如果按人民币计算（汇率按1美元＝7元人民币），则平均数和方差分别为（　）

A．A，s^2　　B．$7A$，$21s^2$　　C．$7A$，$14s^2$　　D．$7A$，$49s^2$

28．已知样本数据$x_1,x_2,\cdots,x_n$的平均数为$\overline{x}$、方差为s^2，若样本数据ax_1+5,ax_2+5，$\cdots$，ax_n+5的平均数为$4\overline{x}(a>0)$，方差为$4s^2$，则$\overline{x}=$（　）

A．$\frac{1}{4}$　　B．$-\frac{5}{12}$　　C．$\frac{5}{6}$　　D．$\frac{5}{2}$

29．一组数据x_1，x_2，x_3，$\cdots$，x_n的平均数和标准差分别为3和1，另一组数据ax_1+b，ax_2+b，ax_3+b，$\cdots$，ax_n+b（其中$a>0$）的平均数和标准差分别为10和4，则$a^b=$（　）

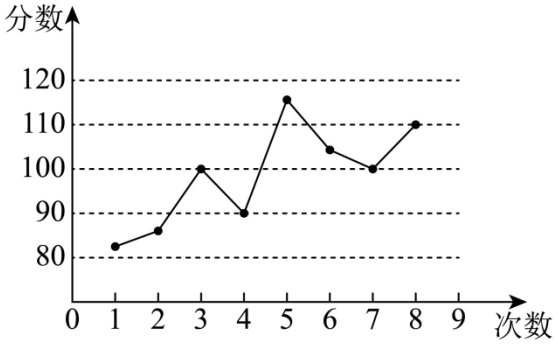
A．16　　B．8　　C．$\frac{1}{16}$　　D．$\frac{1}{8}$

30．已知数据$4x_1+1$，$4x_2+1$，…，$4x_{10}+1$的平均数和方差分别为4，10，那么数据x_1，x_2，…，x_{10}的平均数和方差分别为（　）

A．-1，$\frac{5}{2}$　　B．1，$\frac{5}{2}$　　C．1，$\frac{3}{2}$　　D．$\frac{3}{4}$，$\frac{5}{8}$

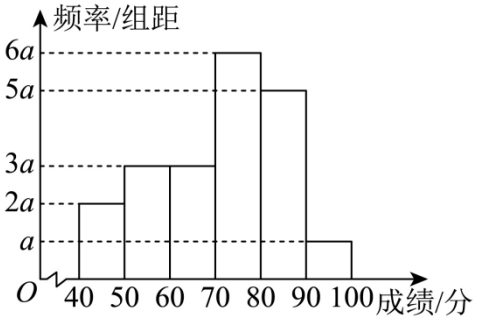
考点4 总体百分位数的估计

31．小明希望自己的高考数学成绩能超过120分，为了激励自己，他记录了近8次数学考试成绩，并绘制成折线统计图，如图，这8次成绩的第80百分位数是（　）



A．100　　B．105　　C．110　　D．120

32．某校高三年级举行数学知识竞赛，并将100名学生的竞赛成绩（满分100分，成绩取整数）整理成如图所示的频率分布直方图，则估计这组数据的第85百分位数为（　）



A．85　　B．86　　C．86.5　　D．87

33．某地气象部门统计了当地2024年3月前8天每天的最高气温T（单位：℃），数据如下：

时间	第1天	第2天	第3天	第4天	第5天	第6天	第7天	第8天
T（℃）	8	12	8	14	16	11	18	21

则这8天的气温数据的75%分位数为（　）

A．15　　B．16　　C．17　　D．18

34．已知某学校参加学科节数学竞赛决赛的8人的成绩（单位：分）为：72，78，80，81，83，86，88，90，则这组数据的第75百分位数是（　）

A．86　　B．87　　C．88　　D．90

35．已知一组数据：4，6，7，9，11，13，则这组数据的第 65 百分位数为（　）

A．6　　B．7　　C．9　　D．11

36．给出下列说法，其中正确的是（　　）

A．某病 8 位患者的潜伏期（天）分别为 3，3，8，4，2，7，10，18，则它们的第 50 百分位数为 4

B．已知数据 $x_1,x_2,\cdots$ 的平均数为 2，方差为 3，那么数据 $2x_1+1$，$2x_2+1$，… 的平均数和方差分别为 5，13

C．在回归直线方程 $\hat{y}=0.25x+1.5$ 中，相对于样本点 $(2,1.2)$ 的残差为 -0.8

D．样本相关系数 $r\in(-1,1)$

37．某台机器每天生产 10000 个零件，现连续 12 天检测，得到每天的次品零件个数依次为：8，12，9，18，16，17，15，9，18，20，13，11，则这组样本数据的中位数与第 60 百分位数之和是（　）

A．29　　B．30　　C．30.5　　D．31

38．样本数据 12，8，32，10，24，22，12，33 的第 60 百分位数为（　）

A．8　　B．12　　C．22　　D．24

39．样本数据 $36,26,17,23,33,106,42,31,30,33$ 的第 60 百分位数为（　）

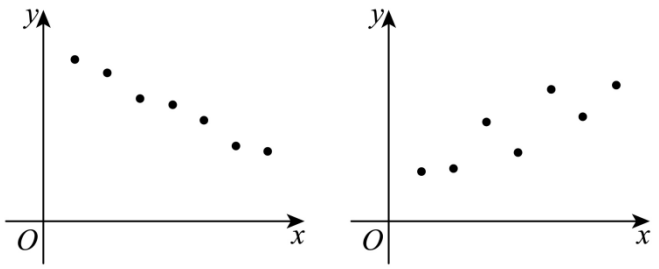
A．23　　B．31　　C．33　　D．36

40．样本数据 11，12，13，15，16，13，14，15，11 的第一四分位数为（　）

A．11.5　　B．12　　C．12.5　　D．13

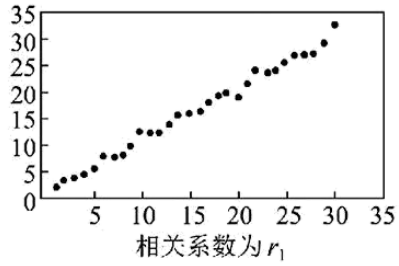
考点 5 相关关系与相关系数

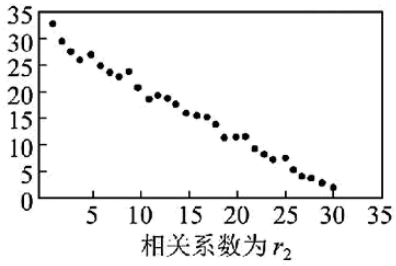
41．如图对两组数据 x，y 和 v，u 分别进行回归分析，得到散点图如图，并求得线性回归方程分别是 $y=b_1x+a_1$ 和 $u=b_2v+a_2$，并对变量 x，y 进行线性相关检验，得到相关系数 r_1，对变量 v，u 进行线性相关检验，得到相关系数 r_2，则下列判断正确的是（　）

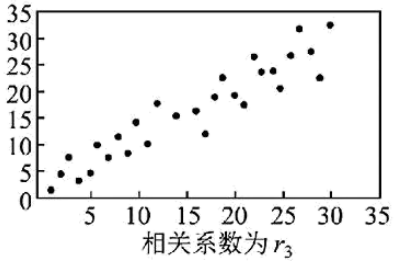


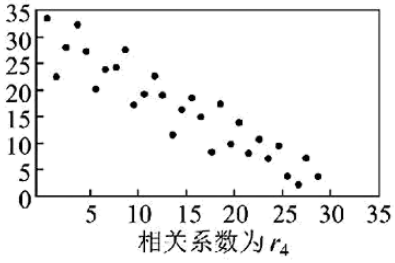
A．$b_1>0$　　B．$b_2<0$　　C．$|r_1|<|r_2|$　　D．$r_1+r_2<0$

42．对四组数据进行统计，获得如图散点图，关于其相关系数的比较，正确的是（　）





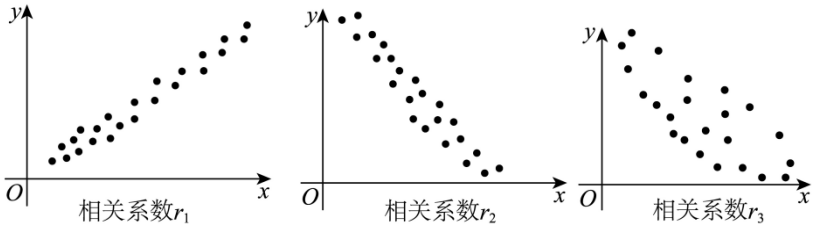




A．$r_4<r_2<0<r_1<r_3$　　B．$r_2<r_4<0<r_3<r_1$

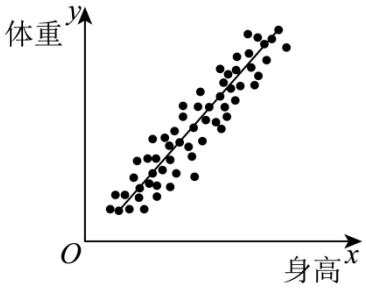
C．$r_2<r_4<0<r_1<r_3$　　D．$r_4<r_2<0<r_3<r_1$

43．上海百联集团对旗下若干门店的营业额与三个影响因素分别作了相关性分析，绘制了如下的散点图，则下述大小关系正确的为（　）.



A．$r_1>r_2>r_3$　　B．$r_2>r_3>r_1$　　C．$r_1>r_3>r_2$　　D．$r_3>r_2>r_1$

44．调查某校高三学生的身高x和体重y得到如图所示散点图，其中身高x和体重y相关系数$r=0.8255$，则下列说法正确的是（　）



A．学生身高和体重没有相关性

B．学生身高和体重呈正相关

C．学生身高和体重呈负相关

D．若从样本中抽取一部分，则这部分的相关系数一定是0.8255

45．已知变量x与y的回归直线方程为$y=3x-1$，变量y与z负相关，则（　）

A．x与y负相关，x与z负相关　　B．x与y正相关，x与z正相关

C．x与y负相关，x与z正相关　　D．x与y正相关，x与z负相关

46．某校数学建模兴趣小组为研究本地区儿子身高$y(\text{cm})$与父亲身高$x(\text{cm})$之间的关系，抽样调查后得出y与x线性相关，且经验回归方程为$\hat{y}=0.85x+29.5$.调查所得的部分样本数据如下：

父亲身高$x(\text{cm})$	164	166	170	173	173	174	180
儿子身高$y(\text{cm})$	165	168	176	170	172	176	178

则下列说法正确的是（　）

A．儿子身高$y(\text{cm})$是关于父亲身高$x(\text{cm})$的函数

B．当父亲身高增加1cm时，儿子身高增加0.85cm

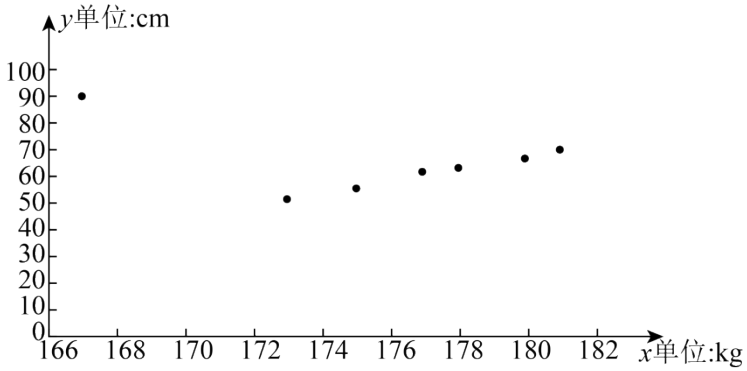
C．儿子身高为172cm时，父亲身高一定为173cm

D．父亲身高为170cm时，儿子身高的均值为174cm

47．某校为了解本校高一男生身高和体重的相关关系，在该校高一年级随机抽取了 7 名男生，测量了他们的身高和体重得下表：

身高x（单位：cm）	167	173	175	177	178	180	181
体重y（单位：kg）	90	54	59	64	67	72	76

由表格制作成如图所示的散点图：



由最小二乘法计算得到经验回归直线l_1的方程为$\hat{y}=b_1x+a_1$，其相关系数为r_1；经过残差分析，点$(167,90)$对应残差过大，把它去掉后，再用剩下的 6 组数据计算得到经验回归直线l_2的方程为$\hat{y}=\hat{b}_2x+\hat{a}_2$，相关系数为

r_2.则下列选项正确的是（　）

A．$\hat{b}_1<\hat{b}_2,\hat{a}_1>\hat{a}_2,r_1<r_2$　　B．$\hat{b}_1<\hat{b}_2,\hat{a}_1<\hat{a}_2,r_1>r_2$

C．$\hat{b}_1>\hat{b}_2,\hat{a}_1<\hat{a}_2,r_1>r_2$　　D．$\hat{b}_1>\hat{b}_2,\hat{a}_1>\hat{a}_2,r_1<r_2$

48．已知变量x，y之间的一组相关数据如下表所示：

x	6	8	10	12
y	6	m	3	2

据此得到变量x，y之间的线性回归方程为$\hat{y}=-0.7x+10.3$，则下列说法不正确的是（　）

A．变量x，y之间成负相关关系　　B．可以预测，当$x=20$时，$\hat{y}=-3.7$

C．$m=4$　　D．该回归直线必过点$(9,4)$

49．已知变量x和y满足经验回归方程$\hat{y}=-0.6x+10.4$，且变量x和y之间的一组相关数据如表所示，则下列说法错误的是（　）

x	6	8	10	12
y	7	m	4	3

A．变量x和y呈负相关　　B．当$x=10$时，$\hat{y}=4.4$

C．$m=5.6$　　D．该经验回归直线必过点$(9,5)$

50．对变量x,y有观测数据$(x_i,y_i)(i\in\mathbf{N}^*)$，得散点图1；对变量$u,v$有观测数据$(u_i,v_i)(i\in\mathbf{N}^*)$，得散点图2.

r_1表示变量x,y之间的线性相关系数，r_2表示变量u,v之间的线性相关系数，则下列说法正确的是（　）

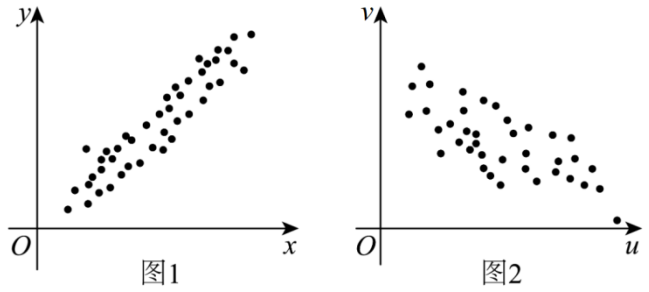


图1　　图2

A．变量x与y呈现正相关，且$|r_1|<|r_2|$　　B．变量x与y呈现负相关，且$|r_1|>|r_2|$

C．变量x与y呈现正相关，且$|r_1|>|r_2|$　　D．变量x与y呈现负相关，且$|r_1|<|r_2|$

考点 6 线性回归方程（非线性）

51．某零售行业为了解宣传对销售额的影响，在本市内随机抽取了 5 个大型零售卖场，得到其宣传费用 x（单位：万元）和销售额 y（单位：万元）的数据如下：

x（万元）	3	4	5	6	7
y（万元）	45	50	60	65	70

由统计数据知 y 与 x 满足线性回归方程 $\hat{y}=\hat{b}x+\hat{a}$，其中 $\hat{b}=6.5$，当宣传费用 $x=10$ 时，销售额 y 的估计值为（　）

A．89.5　　B．90.5　　C．92.5　　D．94.5

52．下列说法中，正确命题的个数为（　）

① 已知随机变量 X 服从二项分布 $B\left(n,\frac{1}{3}\right)$，若 $E(3X+1)=6$，则 $n=5$.

②对具有线性相关关系的变量 x，y，其线性回归方程为 $\hat{y}=0.3x-m$，若样本点的中心为 $(m,2.8)$，则实数 m 的值是 -4.

③以模型 $y=ce^{kx}$ 去拟合一组数据时，为了求出回归方程，设 $z=\ln y$，求得线性回归方程为 $z=0.3x+4$，则 c、k 的值分别是 e^4 和 0.3.

④若样本数据 $x_1,x_2,x_3,\cdots,x_{10}$ 的方差为 2，则数据：$2x_1-1,2x_2-1,\cdots,2x_{10}-1$ 的方差为 16

A．0 个　　B．1 个　　C．2 个　　D．3 个

53．下列说法正确的是（　）

A．若数据 x_1，x_2，…，x_n 的方差为 1，则数据 $2x_1$，$2x_2$，…，$2x_n$ 的标准差为 4

B．已知一组数据 2，3，5，7，8，9，9，11，则该组数据的第 40 百分位数为 6

C．一组样本数据的频率分布直方图是单峰的且形状是对称的，则该组数据的平均数和中位数应该大体上差不多

D．经验回归直线 $\hat{y}=\hat{b}x+\hat{a}$ 恒过 $(\bar{x},\bar{y})$，且在回归直线上的样本点越多，拟合效果越好

54．云计算是信息技术发展的集中体现，近年来，我国云计算市场规模持续增长. 已知某科技公司 2018 年至 2022 年云计算市场规模数据，且市场规模 y 与年份代码 x 的关系可以用模型 $y=c_1e^{c_2x}$（其中 e 为自然对数的底数）拟合，设 $z=\ln y$，得到数据统计表如下：

年份	2018 年	2019 年	2020 年	2021 年	2022 年
年份代码 x	1	2	3	4	5
$z=\ln y$	2	2.4	3	3.6	4

由上表可得经验回归方程 $z=0.52x+\hat{a}$，则 2026 年该科技公司云计算市场规模 y 的估计值为（　）

（参考公式：$\hat{a}=\overline{z}-\hat{b}\overline{x}$）

A．$e^{5.08}$　　B．$e^{5.6}$　　C．$e^{6.12}$　　D．$e^{6.5}$

55．下列说法中正确的是（　）

A．具有线性相关关系的变量 x，y，其线性回归方程为 $y=0.2x-m$，若样本的中心 $(m,3.2)$，则 $m=4$

B．数据 3，4，2，8，1，5，8，6 的中位数为 5

C．将一组数据中的每一个数据加上同一个正数后，方差变大

D．若甲、乙两组数据的相关系数分别为 –0.91 和 0.89，则甲组数据的线性相关性更强

56．下列命题错误的是（　）

A．两个随机变量的线性相关性越强，相关系数的绝对值越接近于 1

B．设 $\xi\sim B(n,p)$，若 $E(\xi)=30$，$D(\xi)=20$，则 $n=90$

C．线性回归直线 $y=\hat{b}x+\hat{a}$ 一定经过样本点的中心 $(\overline{x},\overline{y})$

D．一个袋子中有 100 个大小相同的球，其中有 40 个黄球、60 个白球，从中不放回地随机摸出 20 个球作为样本，用随机变量 X 表示样本中黄球的个数，则 X 服从二项分布，且 $E(X)=8$

57．已知变量 x 和 y 的统计数据如表：

x	1	2	3	4	5
y	6	6	7	8	8

根据上表可得回归直线方程 $y=0.6x+a$，据此可以预测当 $x=8$ 时，$y=$（　　）

A．8.5　　B．9　　C．9.5　　D．10

58．下列说法中正确的个数为（　）个

①对立事件一定是互斥事件；②在经验回归直线方程 $\hat{y}=0.1x+10$ 中，当解释变量 x 每增加一个单位时，预报变量 $\hat{y}$ 减少 0.1 个单位；③两个随机变量的线性相关性越强，相关系数绝对值越接近于 1；④在回归分析模型中，若相关指数 R^2 越小，则残差平方和越大，模型的拟合效果越好.

A．1　　B．2　　C．3　　D．4

59．下列说法不正确的是（　）．

A．一组数据 10，11，11，12，13，14，16，18，20，22 的第 60 百分位数为 14

B．若随机变量 X 服从正态分布 $N(3,\sigma^2)$，且 $P(X\le 4)=0.7$，则 $P(3<X<4)=0.2$

C．若线性相关系数 $|r|$ 越接近 1，则两个变量的线性相关程度越高

D．对具有线性相关关系的变量 x、y，且回归方程为 $y=0.3x-m$，若样本点的中心为 $(m,2.8)$，则实数 m 的值是 –4

60．为了研究 y 关于 x 的线性相关关系，收集了 5 组样本数据（见下表）：

x	1	2	3	4	5
y	0.5	0.9	1	1.1	1.5

若已求得一元线性回归方程为 $\hat{y}=\hat{a}x+0.34$，则下列选项中正确的是（　）

A．$\hat{a}=0.21$

B．当 $x=8$ 时，y 的预测值为 2.2

C．样本数据 y 的第 40 百分位数为 1

D．去掉样本点 $(3,1)$ 后，x 与 y 的样本相关系数 r 不会改变

考点 7 独立性检验

61．“民政送温暖，老人有饭吃”．近年来，各级政府，重视提高老年人的生活质量．在医疗、餐饮等多方面，为老人提供了方便．单从用餐方面，各社区，创建了“爱心食堂”、“爱心午餐”、“老人食堂”等等不同名称的食堂，解决了老人的吃饭问题．“爱心食堂 A”为了更好地服务老人，于 3 月 28 日 12 时，食堂管理层人员对这一时刻用餐的 118 人，对本食堂推出的 15 种菜品按性价比“满意”和“不满意”作问卷调查，其中，有 13 人来食堂用餐不足 5 次，另有儿童 5 人，他们对菜品不全了解，不予问卷统计，在被问卷的人员中男性比女性多 20 人．用餐者对 15 种菜品的性价比认为“满意”的菜品数记为 $x\left(0\le x\le 15,x\in\mathbf{N}\right)$，当 $x\ge 12$ 时，认为该用餐者对本食堂的菜品“满意”，否则，认为“不满意”.统计结果部分信息如下表：

	满意	不满意	合计
男	40		
女	20		

合计			

(1)①完成上面2×2列联表；

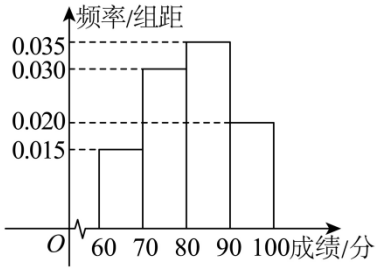
②能有多大（百分比）的把握认为用餐者对本食堂菜品的性价比是否满意与性别有关？

(2)用分层抽样在对菜品的性价比“满意”的人群中抽取 6 人，再从这 6 人中随机抽取 3 人，用X表示抽取的 3 人中的男性人数，求X的分布列和期望.

附：参考公式和临界值表$k^2=\dfrac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$，其中，$n=a+b+c+d$.

$P(k^2\geq k_0)$	0.100	0.050	0.010	0.001
k_0	2.706	3.841	6.635	10.828

62．某学校举办了一次主题为“科技兴国，强国有我”的知识竞赛，并从所有参赛学生中随机抽取了男、女生各 50 人，统计他们的竞赛成绩（满分 100 分，每名参赛学生至少得 60 分），并将成绩分成 4 组：$[60,70)$，$[70,80)$，$[80,90)$，$[90,100]$（单位：分），得到如下的频率分布直方图.



(1)现将竞赛成绩不低于 90 分的学生称为“科技知识达人”，成绩低于 90 分的学生称为“非科技知识达人”．把随机抽取的参赛学生数据统计如下，将下列2×2列联表补充完整，并判断是否有 95%的把握认为能否获得科技知识达人”称号与性别有关.

	科技知识达人	非科技知识达人	合计
男生	15		
女生			
合计			

(2)将频率视为概率，从所有参赛学生中随机抽取 3 人进行访谈，记这 3 人中是“科技知识达人”的人数为ξ，求ξ的分布列与数学期望.

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

$P\left(K^2\geq k_0\right)$	0.10	0.05	0.025	0.010	0.005	0.001
k_0	2.706	3.841	5.024	6.635	7.879	10.828

63．某数学老师在其任教的甲、乙两个班级中各抽取 30 名学生进行测试，分数分布如表：

分数区间	甲班人数	乙班人数
[0，30）	3	6
[30，60）	6	6
[60，90）	9	12
[90，120）	6	3
[120，150]	6	3

(1)若成绩在 120 分以上（含 120 分）为优秀，求从乙班参加测试的成绩在 90 分以上（含 90 分）的学生中，随机任取 2 名学生，恰有 1 名为优秀的概率；

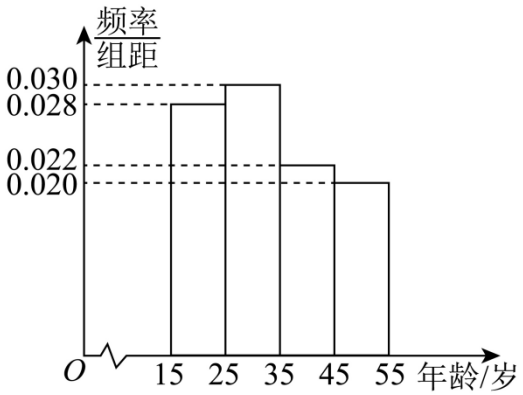
(2)根据以上数据完成下面的 2×2 列联表，则在犯错的概率不超过 0.1 的前提下，是否有足够的把握认为学生的数字成绩优秀与否和班级有关？

	优秀	不优秀	总计
甲班			
乙班			
总计			

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

k_0	2.072	2.706	3.841	5.024	6.635	7.879	10.828
$P\left(K^2 \geq k_0\right)$	0.15	0.10	0.05	0.025	0.010	0.005	0.001

64．同城配送是随即时物流发展而出现的非标准化服务，省时省力是消费者使用同城配送服务的主要目的.某同城配送服务公司随机统计了 800 名消费者的年龄（单位：岁）以及每月使用同城配送服务的次数，得到每月使用同城服务低于 5 次的有 550 人，并将每月使用同城配送服务次数不低于 5 次的消费者按照年龄$[15,25),[25,35),[35,45),[45,55]$进行分组，得到如图所示的频率分布直方图.



(1)估计每月使用同城配送服务不低于 5 次的消费者年龄的平均值和中位数（结果精确到 0.1，每组数据用该组区间的中点值代表）；

(2)若年龄在$[15,35)$内的人位于年龄段 I，年龄在$[35,55]$内的人位于年龄段 II，把每月使用同城配送服务低于 5 次的消费者称为“使用同城配送服务频率低”，否则称为“使用同城配送服务频率高”，若 800 名消费者中有 400 名在年龄段 I，补全2×2列联表，并判断是否有99%的把握认为消费者使用同城配送服务频率的高低与年龄段有关？

	年龄段 I	年龄段 II	合计
使用同城配送服务频率高			
使用同城配送服务频率低			
合计			

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

$P\left(K^2 \geq k_0\right)$	0.050	0.010	0.001
k_0	3.841	6.635	10.828

65．某地为调查年龄在 35—50 岁段人群每周的运动情况，从年龄在 35—50 岁段人群中随机抽取了 200 人的信息，将调查结果整理如下：

	女性	男性
每周运动超过 2 小时	60	80
每周运动不超过 2 小时	40	20

(1)根据以上信息，能否有 99%把握认为该地年龄在 35—50 岁段人群每周运动超过 2 小时与性别有关？

(2)在以上被抽取且每周运动不超过 2 小时的人中，按性别进行分层抽样，共抽 6 人．再从这 6 人中随机抽取 2 人进行访谈，求这 2 人中至少有 1 人是女性的概率．

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

$P\left(K^2 > k_0\right)$	0.10	0.05	0.025	0.010	0.001
k_0	2.706	3.841	5.024	6.635	10.828

66．为了有效预防流感，很多民众注射了流感疫苗.市防疫部门随机抽取了 1000 人进行调查，发现其中注射疫苗的 800 人中有 220 人感染流感，另外没注射疫苗的 200 人中有 80 人感染流感.医学研究表明，流感的检测结果有检错的可能，已知患流感的人其检测结果有95%呈阳性（流感），而没有患流感的人其检测结果有99%呈阴性（未感染）

(1)估计该市流感感染率是多少？

(2)根据所给的数据，判断是否有 99％的把握认为注射流感疫苗与预防流感有关；

(3)已知某人的流感检查结果呈阳性，求此人真的患有流感的概率.（精确到 0.001）

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

$P\left(K^2 \geq k\right)$	0.050	0.010	0.001
k	3.841	6.635	10.828

67．为了引导学生阅读世界经典文学名著，某学校举办“名著读书日”活动，每个月选择一天为“名著读书日”，并给出一些推荐书目．为了了解此活动促进学生阅读文学名著的情况，该校在此活动持续进行了一年之后，随机抽取了校内 100 名学生，调查他们在开始举办读书活动前后的一年时间内的名著阅读数量，所得数据如下表：

	不少于 5 本	少于 5 本	合计
活动前	35	65	100
活动后	60	40	100
合计	95	105	200

(1)依据小概率值$\alpha = 0.001$的独立性检验，分析举办该读书活动对学生阅读文学名著是否有促进作用；

(2)已知某学生计划在接下来的一年内阅读 6 本文学名著，其中 4 本国外名著，2 本国内名著，现从 6 本名著中随机抽取 3 本在上半年读完，求上半年读完的国内名著本数X的分布列及数学期望．

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

临界值表：

α	0.1	0.05	0.01	0.005	0.001
x_α	2.706	3.841	6.635	7.879	10.828

68．随着 AI 技术的不断发展，人工智能科技在越来越多的领域发挥着重要的作用．某校在寒假里给学生推荐了一套智能辅导系统，学生可自愿选择是否使用该系统完成假期的作业．开学时进行了入学测试，随机抽取了 100 名学生统计得到如下列联表：

	使用智能辅导系统	未使用智能辅导系统	合计
入学测试成绩优秀	20	20	40

入学测试成绩不优秀	40	20	60
合计	60	40	100

(1)判断是否有 95%的把握认为入学测试成绩优秀与使用智能辅导系统相关；

(2)若把这 100 名学生按照入学测试成绩是否优秀进行分层随机抽样，从中抽取 5 人，再从这 5 人中随机抽取 2 人，记抽取的 2 人中入学测试成绩优秀的人数为 X，求 X 的分布列及数学期望 $E(X)$.

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

$P\left(K^2\geq k_0\right)$	0.10	0.05	0.025	0.010
k_0	2.706	3.841	5.024	6.635

69．某海鲜餐厅在试营业期间，同时采用自助餐和团购套餐两种营销模式，其中自助餐模式是指顾客可随意享用餐厅内所有菜品，最长可用餐 2 小时；团购套餐是指顾客在 APP 上购买团购券后到店消费，只可享用套餐内所包含的菜品，用餐时间不限.该餐厅为了了解这两种营销模式的受欢迎程度，现随机调查了 130 位顾客对这两种营销模式的意见反馈，统计结果如下表：

	认为自助餐更有性价比	认为团购套餐更有性价比
男性顾客	40	20
女性顾客	30	40

(1)依据小概率值 $\alpha=0.001$ 的独立性检验，推断能否认为顾客对这两种营销模式的意见与顾客的性别有关；

(2)店长统计了第 x_1，x_2，x_3，x_4 天自助餐的用餐人数 y，统计结果如下（已知 $\sum_{i=1}^{4}\left(x_i-\overline{x}\right)^2=5$）：

（天）x	x_1	x_2	x_3	x_4
（用餐人数）y	32	52	73	95

经计算得经验回归方程为 $\hat{y}=21x+10.5$，以样本 $\left(x_i,y_i\right)\left(\mathrm{i}=1,2,3,4\right)$ 的相关系数 r 为标准，对该经验回归方程的拟合效果进行说明.

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

（ii）相关系数 $r=\dfrac{\sum\limits_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\sum\limits_{i=1}^{n}(x_i-\overline{x})^2\sum\limits_{i=1}^{n}(y_i-\overline{y})^2}}$ 若 $|r|>0.75$，可认为该模型拟合效果良好，反之，则认为该模型拟合效果不好.

（iii）$\chi^2=\dfrac{n\left(ad-bc\right)^2}{\left(a+b\right)\left(c+d\right)\left(a+c\right)\left(b+d\right)}$，其中 $n=a+b+c+d$.

α	0.050	0.010	0.001
x_α	3.841	6.635	10.828

70．2024 年 3 月，某校语文教师对学生提出“3 月读一本书”的要求，每位学生都选择且只能选择《红楼梦》和《三国演义》中的一本，现随机调查该校男、女生各 100 人，整理得到 2×2 列联表如下.

	《红楼梦》	《三国演义》
男生	30	70
女生	60	40

(1)依据小概率值 $\alpha=0.001$ 的独立性检验，能否认为学生选择《红楼梦》还是《三国演义》与性别有关？

(2)已知学生选择哪本书是相互独立的，用频率代替概率，从该校选择《红楼梦》的学生中随机抽取 3 人，抽到的女生人数设为 ξ，求 ξ 的分布列和数学期望.

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

参考数据：

α	0.1	0.05	0.01	0.005	0.001
x_α	2.706	3.841	6.635	7.879	10.828

考点巩固卷 21　统计与统计案例（七大考点）

考点预览

高端技巧及考点训练

考点 1　简单随机抽样

1、抽样调查

（1）总体：统计中所考察对象的某一数值指标的全体构成的集合称为总体.

（2）个体：构成总体的每一个元素叫做个体.

（3）样本：从总体中抽取若干个个体进行考察，这若干个个体所构成的集合叫做总体的一个样本，样本中个体的数目叫做样本容量.

2、简单随机抽样

（1）定义

一般地，设一个总体含有 N 个个体，从中逐个不放回地抽取 n 个个体作为样本（$n \le N$），如果每次抽取时总体内的各个个体被抽到的机会都相等，就把这种抽样方法叫做简单随机抽样．这样抽取的样本，叫做简单随机样本.

（2）两种常用的简单随机抽样方法

①抽签法：一般地，抽签法就是把总体中的 N 个个体编号，把号码写在号签上，将号签放在一个容器中，搅拌均匀后，每次从中抽取一个号签，连续抽取 n 次，就得到一个容量为 n 的样本.

②随机数法：即利用随机数表、随机数骰子或计算机产生的随机数进行抽样．这里仅介绍随机数表法．随机数表由数字 0，1，2，…，9 组成，并且每个数字在表中各个位置出现的机会都是一样的.

注意：为了保证所选数字的随机性，需在查看随机数表前就指出开始数字的横、纵位置.

（3）抽签法与随机数法的适用情况

抽签法适用于总体中个体数较少的情况，随机数法适用于总体中个体数较多的情况，但是当总体容量很大时，需要的样本容量也很大时，利用随机数法抽取样本仍不方便.

（4）简单随机抽样的特征

①有限性：简单随机抽样要求被抽取的样本的总体个数是有限的，便于通过样本对总体进行分析.

②逐一性：简单随机抽样是从总体中逐个地进行抽取，便于实践中操作.

③不放回性：简单随机抽样是一种不放回抽样，便于进行有关的分析和计算.

④等可能性：简单单随机抽样中各个个体被抽到的机会都相等，从而保证了抽样方法的公平.

只有四个特点都满足的抽样才是简单随机抽样.

3、分层抽样

（1）定义

一般地，在抽样时，将总体分成互不交叉的层，然后按照一定的比例，从各层独立地抽取一定数量的个体，将各层取出的个体合在一起作为样本，这种抽样方法叫做分层抽样.

分层抽样适用于已知总体是由差异明显的几部分组成的.

（2）分层抽样问题类型及解题思路

①求某层应抽个体数量：按该层所占总体的比例计算.

②已知某层个体数量，求总体容量或反之求解：根据分层抽样就是按比例抽样，列比例式进行计算.

③分层抽样的计算应根据抽样比构造方程求解，其中“抽样比$=\dfrac{\text{样本容量}}{\text{总体容量}}=\dfrac{\text{各层样本数量}}{\text{各层个体数量}}$”

注意：分层抽样时，每层抽取的个体可以不一样多，但必须满足抽取$n_i=n\cdot\dfrac{N_i}{N}$（$i=1,2,\cdots,k$）个个体（其中i是层数，n是抽取的样本容量，N_i是第i层中个体的个数，N是总体容量）.

1．电影《孤注一掷》的上映引发了电信诈骗问题的热议，也加大了各个社区反电信诈骗的宣传力度.已知某社区共有居民480人，其中老年人200人，中年人200人，青少年80人，若按年龄进行分层随机抽样，共抽取36人作为代表，则中年人比青少年多（　）

A．6人　　B．9人　　C．12人　　D．18人

【答案】B

【分析】根据题意可以计算出分层随机抽样的抽样比例，进而计算出中年人和青年人的人数，进而可以知道中年人比青少年多多少个.

【详解】设中年人抽取x人，青少年抽取y人，由分层随机抽样可知$\dfrac{200}{480}=\dfrac{x}{36},\dfrac{80}{480}=\dfrac{y}{36}$，

解得$x=15,y=6$，故中年人比青少年多9人.

故选:B.

2．已知A,B,C三种不同型号的产品数量之比依次为$4:3:7$，现用分层抽样的方法抽取容量为N的样本，若样本中A型号产品有20件，则N为（　）

A．60　　B．70　　C．80　　D．90

【答案】B

【分析】由条件确定A型号产品的抽样比，再根据频数，频率，样本容量的关系求N.

【详解】因为A,B,C三种不同型号的产品数量之比依次为$4:3:7$，

且用分层抽样的方法抽取一个容量为N的样本，

所以A型号产品被抽的抽样比为：$\frac{4}{4+3+7}=\frac{2}{7}$，

因为A型号产品有20件，所以$\frac{20}{N}=\frac{2}{7}$，解得$N=70$.

故选：B.

3．国内某优秀新能源电池制造企业在锂电池单位能量密度技术上取得了重大突破，该制造企业内的某车间有两条生产线，分别生产高能量密度锂电池和低能量密度锂电池，总产量为400个锂电池．质检人员采用分层随机抽样的方法随机抽取了一个容量为80的样本进行质量检测，已知样本中高能量密度锂电池有35个，则估计低能量密度锂电池的总产量为（　）.

A．325个　　B．300个　　C．225个　　D．175个

【答案】C

【分析】根据分层抽样计算规则计算可得.

【详解】根据分层随机抽样可知低能量密度锂电池的产量为$400\times\frac{80-35}{80}=225$（个）.

故选：C

4．用按比例分配的分层随机抽样方法，从某学校的600名男生和800名女生中选取14人参与某项研学活动，则女生比男生多选取（　）

A．8人　　B．6人　　C．4人　　D．2人

【答案】D

【分析】确定抽样比计算出男生和女生的人数即可得出结论.

【详解】依题意可知，分层抽样比为$100:1$，

因此可得选取的男生为6人，女生为8人，

所以女生比男生多选取2人.

故选：D

5．已知甲组数据：1，3，5，7，9，11，乙组数据：2，4，8，16，根据不同组别，用分层抽样的方法随机抽取一个容量为5的样本，则该样本的平均数不可能是（　）

A．5　　B．7　　C．9　　D．11

【答案】D

【分析】先根据分层抽样算出甲乙两组数据抽到的数据个数，列出表格，在结合平均数公式计算得出答案；

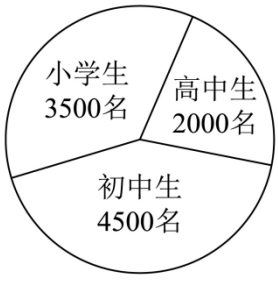
【详解】根据分层抽样可知甲组数据抽取 3 个数据，乙组数据抽取 2 个数据，具体情况如下表：

甲组抽样	乙组抽样	平均数
3，5，7	2，8	5
5，7，11	4，8	7
5，7，9	8，16	9

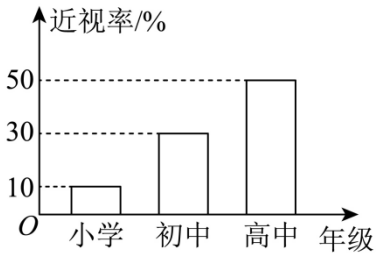
平均数为 11 时，需 5 个样本数字之和为 55，而样本之和最大值为 $7+9+11+8+16=51$.

故选：D.

6．已知某地区中小学生人数和近视情况分别如图甲和图乙所示．为了了解该地区中小学生近视情况形成的原因，采用分层抽样的方法抽取部分学生进行调查，若抽取的小学生人数为 70，则抽取的高中生中近视人数为（　）



图甲



图乙

A．10　　B．20　　C．25　　D．40

【答案】B

【分析】根据题意，求得抽取的高中生人数是 40 人，再结合图乙可知高中生的近视率为 50%，即可求解.

【详解】由图甲可知抽取的高中生人数是 $70\times\frac{2000}{3500}=40$，

又由图乙可知高中生的近视率为 50%，所以抽取的高中生中近视人数为 $40'50\%=20$ 人.

故选：B.

7．为了检查某超市货架上的饮料是否含有塑化剂，要从编号依次为 1 到 100 的塑料瓶装饮料中抽取 5 瓶进行检验，用每部分选取的号码间隔一样的系统抽样方法确定所选取的 5 瓶饮料的编号可能是（　）

A．5，15，25，35，45　　B．10，25，40，55，70

C．10，20，30，40，50　　D．10，30，50，70，90

【答案】D

【分析】求出分段间隔，然后验证每个选项中样本编号的间隔即可得出结论.

【详解】利用系统抽样，把编号分为 5 段，每段 20 个，每段抽取 1 个，号码间隔为 20.

选项 A 中样本间隔为 10，选项 B 中样本间隔为 15，选项 C 中样本间隔为 10，

选项 D 中样本间隔为 20.

故选：D

8．从一个含有 N 个个体的总体中抽取一容量为 n 的样本，当选取抽签法、随机数法和分层随机抽样三种不同方法时，总体中每个个体被抽中的概率分别为 p_1, p_2, p_3，三者关系可能是（　）

A．$p_1 = p_2 < p_3$　　B．$p_1 = p_2 = p_3$　　C．$p_1 = p_3 < p_2$　　D．$p_2 = p_3 < p_1$

【答案】B

【分析】根据抽样的概念，每个个体被抽中的概率是均等的，进而即可选择答案.

【详解】因为在抽签法抽样、随机数法抽样和分层随机抽样中，每个个体被抽中的概率均为 $\frac{n}{N}$，

所以 $p_1 = p_2 = p_3$.

故选：B.

9．下列说法中正确的个数有（　）

①对具有线性相关关系的变量 x，y，其回归方程为 $\hat{y} = 0.3x - m$，若样本点的中心为 $(m, 2.8)$，则实数 m 的值是 -4；

②某校共有学生 1003 人，用简单随机抽样的方法先剔除 3 人，再按简单随机抽样的方法抽取为 20 人，则每个学生被抽到的概率为 $\frac{1}{50}$；

③若随机事件 A，B 满足：$P(A) = \frac{1}{2}$，$P(B) = \frac{2}{3}$，$P(A \cup B) = \frac{5}{6}$，则事件 A 与 B 相互独立；

④若随机变量 ξ，η 满足 $\eta = 2\xi + 1$，则 $D(\eta) = 2D(\xi) + 1$.

A．1　　B．2　　C．3　　D．4

【答案】B

【分析】根据根据回归直线过样本中心点，计算可判断①正确；据简单随机抽样概率均等计算可知②错误；由 $P(A \cup B) = P(A) + P(B) - P(A \cap B)$，可求得 $P(A \mid B) = \frac{1}{3}$，可判断③正确；根据方差的计算公式可知④错.

【详解】对于①：因为回归方程为 $\hat{y} = 0.3x - m$，又样本点中心为 $(m, 2.8)$，

所以 $2.8 = 0.3m - m$，解得 $m = -4$，则实数 m 的值是 -4，故①正确；

对于②：根据简单随机抽样概率均等可知，每个学生被抽到的概率为 $\frac{1000}{1003} \cdot \frac{20}{1000} = \frac{20}{1003}$，故②错误.

对于③：由 $P(A\cup B)=P(\mathrm{A})+P(B)-P(A\cap B)$，可得 $\frac{5}{6}=\frac{1}{2}+\frac{2}{3}-P(A\mid B)$，

解得 $P(A\mid B)=\frac{1}{3}$，$P(A)P(B)=\frac{1}{2}\times\frac{2}{3}=\frac{1}{3}$，所以 $P(A\cap B)=P(A)P(B)$，

所以事件 A 与 B 相互独立，故③正确；

对于④：由 $\eta=2\xi+1$，可得 $D(\eta)=4D(\xi)$，故④错误.

故正确的命题有 2 个.

故选：B.

10．为了解某校初中学生的近视情况，按年级用分层抽样的方法随机抽取 100 名学生进行视力检测，已知初一、初二、初三年级分别有 800 名，600 名，600 名学生，则不同的抽样结果共有（　）

A．$C_{800}^{40}\cdot\left(C_{600}^{30}\right)^2$　B．$C_{800}^{30}\cdot C_{600}^{40}\cdot C_{600}^{30}$　C．$\left(C_{800}^{30}\right)^2\cdot C_{600}^{40}$　D．$\left(C_{800}^{40}\right)^2\cdot C_{600}^{20}$

【答案】A

【分析】根据分层抽样可知抽取初一学生 40 名，初二、初三学生各 30 名，由分步乘法计数原理即可求解.

【详解】由初一、初二、初三年级分别有 800 名，600 名，600 名学生可知，

抽样比为 $\frac{100}{800+600+600}=\frac{1}{20}$，

按年级用分层抽样的方法随机抽取初一学生 40 名，初二、初三学生各 30 名，

根据分步乘法计数原理可知，

不同的抽样结果共有 $C_{800}^{40}\cdot\left(C_{600}^{30}\right)^2$.

故选：A.

考点 2　频率分布直方图

1、频率分布直方图

（1）频率、频数、样本容量的计算方法

①$\frac{\text{频率}}{\text{组距}}$×组距＝频率.

②$\frac{\text{频数}}{\text{样本容量}}$＝频率，$\frac{\text{频数}}{\text{频率}}$＝样本容量，样本容量×频率＝频数.

③频率分布直方图中各个小方形的面积总和等于1 .

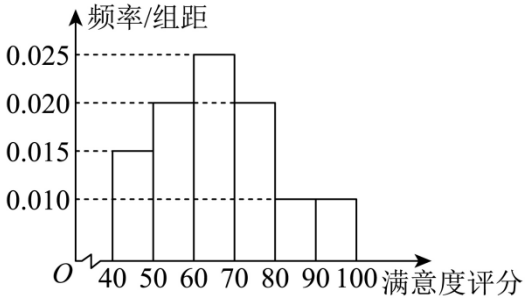
2、频率分布直方图中数字特征的计算

（1）最高的小长方形底边中点的横坐标即是众数.

（2）中位数左边和右边的小长方形的面积和是相等的．设中位数为 x，利用 x 左（右）侧矩形面积之和等于 0.5，即可求出 x.

（3）平均数是频率分布直方图的“重心”，等于频率分布直方图中每个小长方形的面积乘以小长方形底边中点的横坐标之和，即有 $\bar{x}=x_1p_1+x_1p_1+\cdots+x_np_n$，其中 x_n 为每个小长方形底边的中点， p_n 为每个小长方形的面积.

11．某公司为了解用户对其产品的满意度，从使用该产品的用户中随机调查了 100 个用户，根据用户对产品的满意度评分，得到如图所示的用户满意度评分的频率分布直方图：



根据此频率分布直方图，下面结论中不正确的是（　）

A．对该公司产品满意度评分低于 60 分的用户比例估计为 35%

B．对该公司产品满意度评分不低于 70 分的用户比例估计为 40%

C．估计该公司用户对产品的满意度评分的平均值不超过 60 分

D．估计该公司有一半以上的用户，对产品的满意度评分介于 50 分至 80 分之间

【答案】C

【分析】由频率分布直方图计算频率逐项判断 A，B，D 即可，计算平均数判断 C 即可.

【详解】对于 A，对该公司产品满意度评分低于 60 分的用户比例估计为：

$(0.015+0.020)\times10\times100\%=35\%$，故 A 正确；

对于 B，对该公司产品满意度评分不低于 70 分的用户比例估计为：

$(0.020+0.010+0.010)\times10\times100\%=40\%$，故 B 正确；

对于 C，估计该公司用户对产品的满意度评分的平均值为：

$\bar{x}=45\times0.15+55\times0.2+65\times0.25+75\times0.2+85\times0.1+95\times0.1=67>60$，故 C 错误；

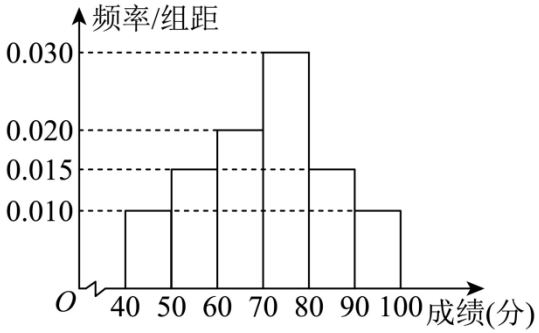
对于 D，对产品的满意度评分介于 50 分至 80 分之间的用户比例为：

$(0.025+0.020+0.020)\times10\times100\%=65\%$，

估计该公司有一半以上的用户，对产品的满意度评分介于 50 分至 80 分之间，故 D 正确.

故选：C.

12．在某次高中数学模拟考试中，对 800 名考生的考试成绩进行统计，得到如图所示的频率分布直方图，其中分组的区间分别为$[40,50)$，$[50,60)$，$[60,70)$，$[70,80)$，$[80,90)$，$[90,100]$.若考生成绩在$[70,80)$内的人数为m，考生成绩在$[80,100]$内的人数为n，则$m-n=$（　）



A．20　　B．10　　C．60　　D．40

【答案】D

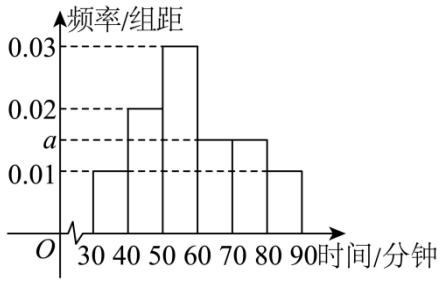
【分析】由频率分布直方图求出m、n，即可得解.

【详解】由频率分布直方图可得$m=800\times0.03\times10=240$，$n=800\times(0.01+0.015)\times10=200$，

所以$m-n=240-200=40$.

故选：D.

13．为了解高中学生每天的体育活动时间,某市教育部门随机抽取1000高中学生进行调查,把每天进行体育活动的时间按照时长（单位：分钟）分成6组:$[30,40)$,$[40,50)$,$[50,60)$,$[60,70)$,$[70,80)$,$[80,90]$.然后对统计数据整理得到如图所示的频率分布直方图,则可估计这1000名学生每天体育活动时间的第25百分位数为（　）



A．47.5　　B．45.5　　C．43.5　　D．42.5

【答案】A

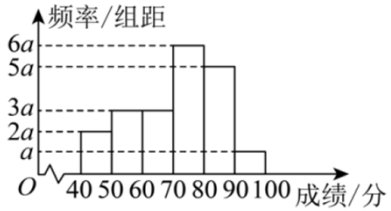
【分析】根据第25百分位数的概念，知道它在第二组$[40,50)$里.运用概率之和为0.25，构造方程，解出即可.

【详解】第25百分位数设为x，而$0.1<0.25<0.1+0.2$，则所求百分位数在第二组，

则可列方程$0.1+0.02(x-40)=0.25$,解得$x=47.5$.

故选:A.

14．为了加深师生对党史的了解，激发广大师生知史爱党、知史爱国的热情，某校举办了“学党史、育新人”的党史知识竞赛，并将 1000 名师生的竞赛成绩（满分 100 分，成绩取整数）整理成如图所示的频率分布直方图，则下列说法错误的是（　）



A．a 的值为 0.005

B．估计这组数据的众数为 75 分

C．估计成绩低于 60 分的有 250 人

D．估计这组数据的中位数为 $\frac{235}{3}$ 分

【答案】D

【分析】对 A，根据频率和为 1 求解即可；对 B，根据频率分布直方图的众数判断即可；对 C，计算成绩低于 60 分的频率，进而可得人数；对 D，根据成绩低于中位数的频率为 0.5 计算即可.

【详解】对 A，由题意，$10\times(2a+3a+3a+6a+5a+a)=1$，解得 $a=0.005$，故 A 正确；

对 B，由直方图可得估计这组数据的众数为 $\frac{70+80}{2}=75$ 分，故 B 正确；

对 C，由直方图可得成绩低于 60 分的频率为 $10\times(0.01+0.015)=0.25$，故估计成绩低于 60 分的有 $1000\times0.25=250$ 人，故 C 正确；

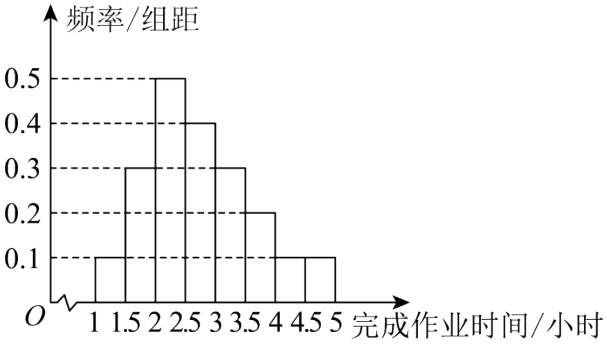
对 D，由 A 可得区间 $[40,50),[50,60),[60,70),[70,80),[80,90),[90,100]$ 的频率分别为 $0.1,0.15,0.15,0.3,0.25,0.05$，

因为 $0.1+0.15+0.15+0.3>0.5$，$0.1+0.15+0.15<0.5$，故中位数位于 $[70,80)$ 内.

设中位数为 x，则 $0.1+0.15+0.15+0.03\times(x-70)=0.5$，解得 $x=\frac{220}{3}$，故 D 错误.

故选：D

15．某教育机构为调查中小学生每日完成作业的时间，收集了某位学生 100 天每天完成作业的时间，并绘制了如图所示的频率分布直方图（每个区间均为左闭右开），根据此直方图得出了下列结论，其中正确的是（　）



A．估计该学生每日完成作业的时间在 2 小时至 2.5 小时的有 50 天

B．估计该学生每日完成作业时间超过 3 小时的概率为 0.3

C．估计该学生每日完成作业时间的中位数为 2.625 小时

D．估计该学生每日完成作业时间的众数为 2.3 小时

【答案】C

【分析】利用频率分别直方图、频数、频率、中位数、众数直接求解.

【详解】对于 A，该学生每日完成作业的时间在 2 小时至 2.5 小时的天数为：$0.5\times0.5\times100=25$ 天，故 A 错误；

对于 B，估计该学生每日完成作业时间超过 3 小时的概率为$(0.3+0.2+0.1+0.1)\times0.5=0.35$，故 B 错误；

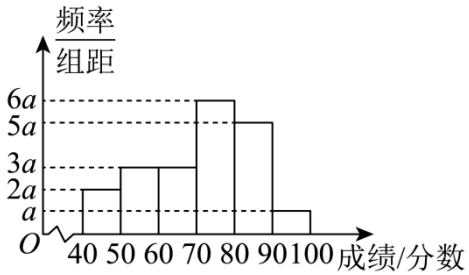
对于 C，$[1,2.5)$的频率为$(0.1+0.3+0.5)\times0.5=0.45$，$[1,3)$的频率为$0.45+0.4\times0.5=0.65$，

则该学生每日完成作业时间的中位数为$2.5+\dfrac{0.5-0.45}{0.2}\times0.5=2.625$，故 C 正确；

对于 D，估计该学生每日完成作业时间的众数为$\dfrac{2+2.5}{2}=2.25$，故 D 错误；

故选：C

16．为了加深师生对党史的了解，激发广大师生知史爱党、知史爱国的热情，某校举办了“学党史、育文化”的党史知识竞赛，并将 1000 名师生的竞赛成绩（满分 100 分，成绩取整数）整理成如图所示的频率分布直方图，估计这组数据的第 85 百分位数为（　）分



A．84　　B．85　　C．86　　D．87

【答案】C

【分析】根据百分位数定义，结合数据求解即可.

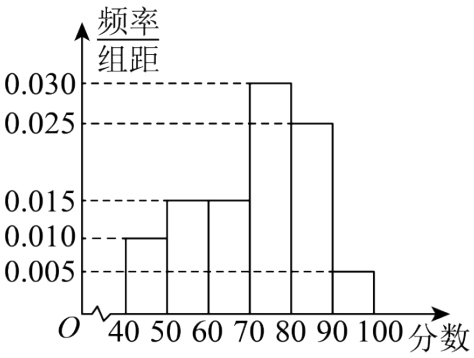
【详解】由$10\times(2a+3a+3a+6a+5a+a)=1$，解得：$a=0.005$，

所以前 4 组频率之和为$14\times0.005\times10=0.7$，前 5 组频率之和为$19\times0.005\times10=0.95$，

设这组数据的第 85 百分位数为x，则$0.7+(x-80)\times0.025=0.85$，解得：$x=86$，

故选：C

17．某校高三共有 200 人参加体育测试，将体测得分情况进行了统计，把得分数据按照$[40,50),[50,60),[60,70),[70,80),[80,90),[90,100]$分成 6 组，绘制了如图所示的频率分布直方图.根据规则，82 分以上的考生成绩等级为 A，则获得 A 的考生人数约为（　）



A．25　　B．50　　C．75　　D．100

【答案】B

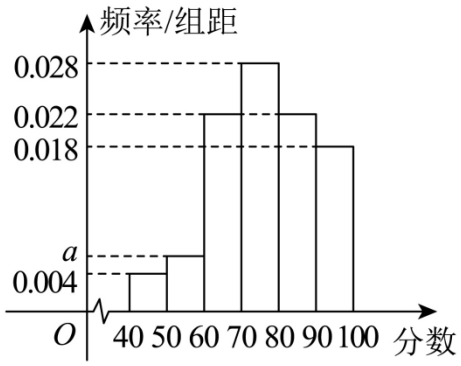
【分析】根据频率分布直方图求获得 A 的频率，进而可得相应的人数.

【详解】由题意可知：估计获得 A 的频率为$0.025\times(90-82)+0.005\times10=0.25$，

所以获得 A 的考生人数约为$0.25\times200=50$.

故选：B.

18．为深入贯彻落实习近平总书记对天津工作“三个着力”重要要求，天津持续深化改革，创建全国文明城区，城市文明程度显著提升，人民群众的梦想不断实现.在创建文明城区的过程中，中央文明办对某小区居民进行了创建文明城区相关知识网络问卷调查，从本次问卷中随机抽取了 50 名居民的问卷结果，统计其得分数据，将所得 50 份数据的得分结果分为 6 组：$[40,50),[50,60),[60,70),[70,80),[80,90),[90,100]$，并整理得到如下的频率分布直方图，则该小区居民得分的第 70 百分位数为（　）



A．89.09　　B．86.52　　C．84.55　　D．81.32

【答案】C

【分析】利用百分位数的概念以及频率分布直方图求解.

【详解】由题意得$(0.004+a+0.018+2\times0.022+0.028)\times10=1$，

解得$a=0.006$，

因为前 4 组数据的频率之和为$0.04+0.06+0.22+0.28=0.6$，

前 5 组数据的频率之和为$0.04+0.06+0.22+0.28+0.22=0.82$，

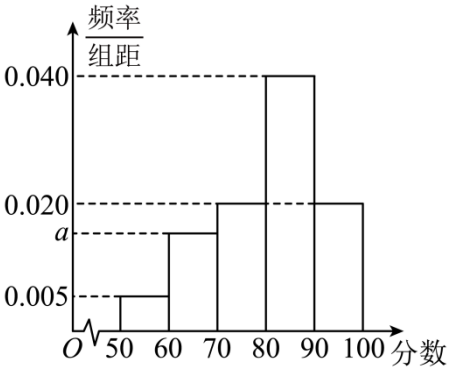
则 70% 分位数在$[80,90)$内，设 70% 分位数为 x，

则$0.6+(x-80)\times0.022=0.7$，解得$x\approx84.55$，

所以 70% 分位数约为84.55 .

故选：C.

19．某市为了解全市 12000 名高一学生的的体能素质情况，在全市高一学生中随机抽取了 1000 名学生进行体能测试，并将这 1000 名的体能测试成绩整理成如下频率分布直方图．根据此频率分布直方图，下列结论中正确的是（　）



A．图中a的值为 0.020；

B．同一组中的数据用该组区间的中点值做代表，则这 1000 名学生的平均成绩约为 80.5；

C．估计样本数据的 75%分位数为 88；

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：https://d.book118.com/925033343310012102