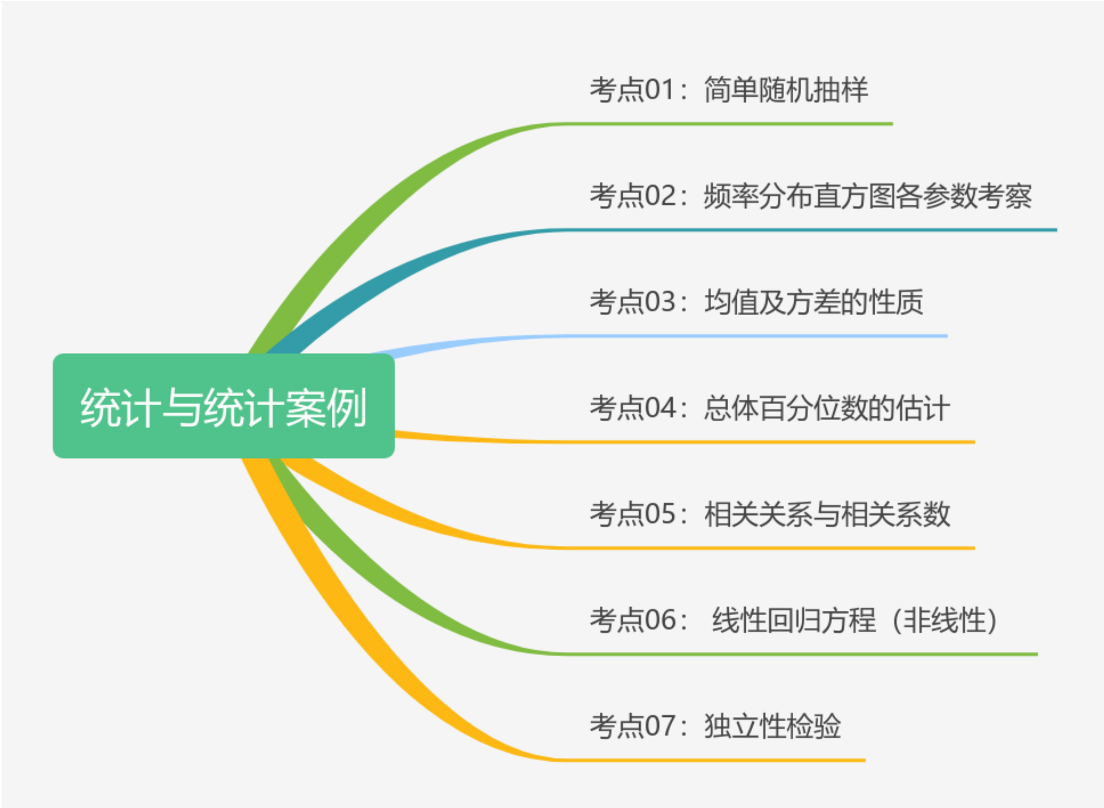
考点巩固卷 21　统计与统计案例（七大考点）

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：$\dfrac{4}{4+3+7}=\dfrac{2}{7}$，

因为A型号产品有20件，所以$\dfrac{20}{N}=\dfrac{2}{7}$，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

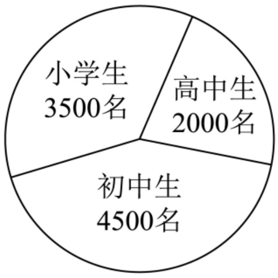
3，5，7	2，8	5
5，7，11	4，8	7

5，7，9	8，16	9

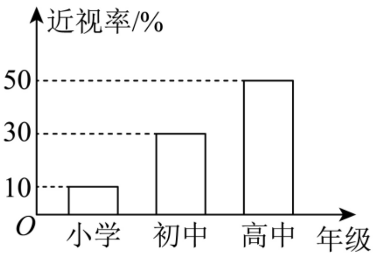
平均数为 11 时，需 5 个样本数字之和为 55，而样本之和最大值为$7+9+11+8+16=51$.

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$70\times\dfrac{2000}{3500}=40$，

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$40\times 50\%=20$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$P(A \cup B) = P(\mathrm{A}) + P(B) - P(A \cap B)$，可求得$P(AB) = \frac{1}{3}$，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$\frac{5}{6}=\frac{1}{2}+\frac{2}{3}-P(A\cap B)$，

解得$P(A\cap B)=\frac{1}{3}$，$P(A)P(B)=\frac{1}{2}\times\frac{2}{3}=\frac{1}{3}$，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．$\mathrm{C}_{800}^{40}\cdot\left(\mathrm{C}_{600}^{30}\right)^2$　B．$\mathrm{C}_{800}^{30}\cdot\mathrm{C}_{600}^{40}\cdot\mathrm{C}_{600}^{30}$　C．$\left(\mathrm{C}_{800}^{30}\right)^2\cdot\mathrm{C}_{600}^{40}$　D．$\left(\mathrm{C}_{800}^{40}\right)^2\cdot\mathrm{C}_{600}^{20}$

【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$\mathrm{C}_{800}^{40}\cdot\left(\mathrm{C}_{600}^{30}\right)^2$.

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$\frac{频率}{组距}\times$组距＝频率.

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

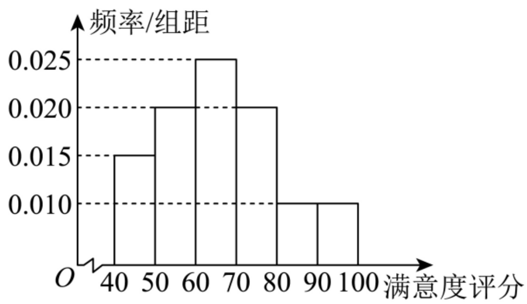
2、频率分布直方图中数字特征的计算

（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 $\overline{x}=x_1p_1+x_1p_1+\cdots+x_np_n$，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：

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

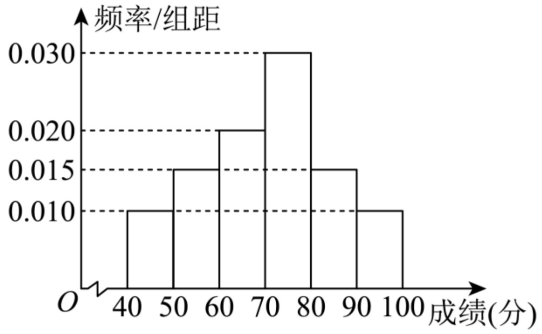
对于 D，对产品的满意度评分介于 50 分至 80 分之间的用户比例为：

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

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

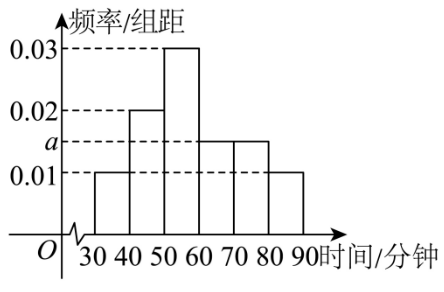
【答案】D

【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

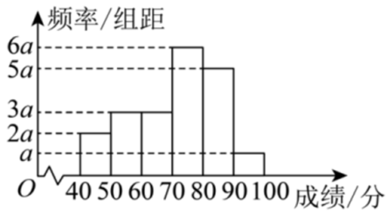
，构造方程，解出即可.

【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 $\dfrac{235}{3}$ 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 $\dfrac{70+80}{2}=75$ 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

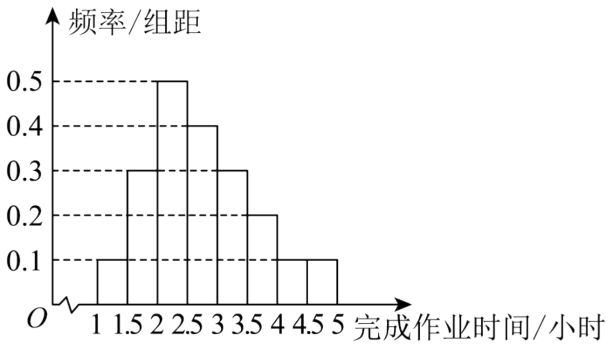
$0.1,0.15,0.15,0.3,0.25,0.05$，

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 $x=\dfrac{220}{3}$，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；

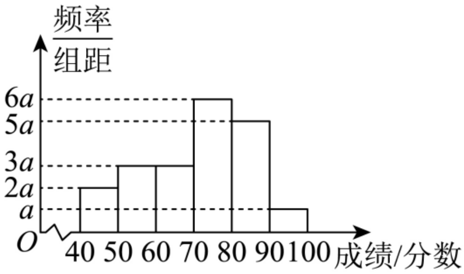
对于 C，$[1,2.5)$ 的频率为 $(0.1+0.3+0.5)\times0.5=0.45$，$[1,3)$ 的频率为 $0.45+0.4\times0.5=0.65$，

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.

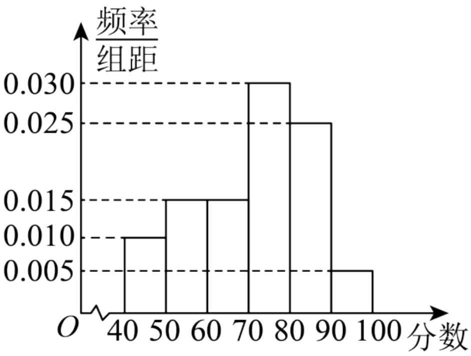
【详解】由$10\times(2a+3a+3a+6a+5a+a)=1$，解得：$a=0.005$，

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

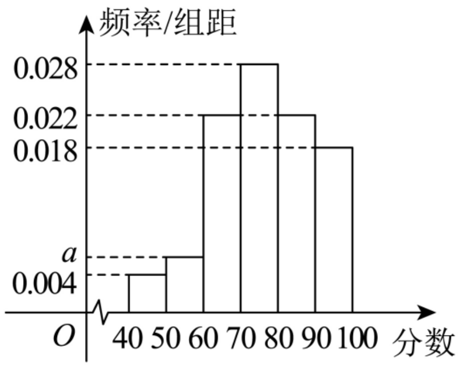
【分析】根据频率分布直方图求获得 A 的频率，进而可得相应的人数.

【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$0.04+0.06+0.22+0.28+0.22=0.82$，

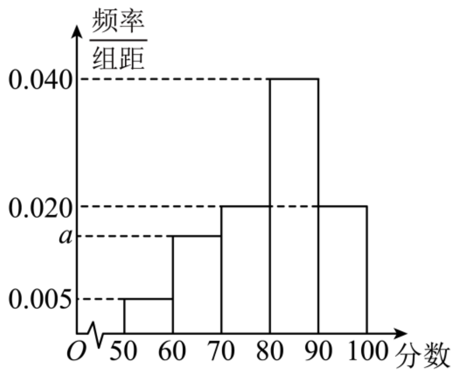
则70%分位数在[80,90)内，设70%分位数为 x，

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；

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．

【答案】B

【分析】A.根据频率和为 1，计算 a 的值；B.根据平均数公式，判断 B；C.根据百分位数公式，判断 C；计算体测成绩在 $[80,100]$ 内的频率，再结合总人数，即可判断 D.

【详解】A.由频率分布直方图可知，$10\times(0.005+a+0.02+0.04+0.02)=1$，

得：$a=0.015$，故 A 错误；

B.$(55\times0.005+65\times0.015+75\times0.02+85\times0.04+95\times0.02)\times10=80.5$，故 B 正确；

C.设 75% 百分位数 x，易得 $x\in[80,90)$，

则 $10\times0.005+10\times0.015+10\times0.02+(x-80)\times0.04=0.75$，

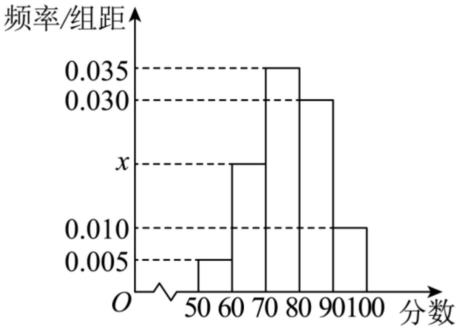
解得：$x=88.75$，故 C 错误；

D.则体测成绩在 $[80,100]$ 的频率为 $10\times0.04+10\times0.02=0.6$，

估计全市高一学生体测成绩优异（80 分及以上）的人数约为 $12000\times0.6=7200$ 人，故 D 错误.

故选：B.

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

【答案】C

【分析】根据频率分布直方图中所有长方形的面积和为 1，以及极值、频数以及平均数的计算，对每个选项进行逐一分析，即可判断和选择.

【详解】对于 A，由 $(0.005+x+0.035+0.030+0.010)\times 10=1$，可解得 $x=0.020$，故选项 A 正确；

对于 B，得分在 80 分及以上的人数的频率为 $(0.030+0.010)\times 10=0.4$，

故人数为 $1000\times 0.4=400$，故选项 B 正确；

对于 C，频率分布直方图无法看出这组数据的最大值和最小值，故选项 C 不正确；

对于 D，这组数据的平均数的估计值为：

$55\times 0.05+65\times 0.2+75\times 0.35+85\times 0.3+95\times 0.1=77$，故选项 D 正确.

故选：C.

考点 3 均值及方差的性质

平均数、方差的性质

如果数据 $x_1,x_2,\ldots\ldots,x_n$ 的平均数为 $\overline{x}$，方差为 s^2，那么

①一组新数据 $x_1+b,x_2+b,\ldots\ldots x_n+b$ 的平均数为 $\overline{x}+b$，方差是 s^2.

②一组新数据 $ax_1,ax_2,\ldots\ldots,ax_n$ 的平均数为 $a\overline{x}$，方差是 a^2s^2.

③一组新数据 $ax_1+b,ax_2+b,\ldots\ldots,ax_n+b$ 的平均数为 $a\overline{x}+b$，方差是 a^2s^2.

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

A．9，4　　B．9，2　　C．4，1　　D．2，1

【答案】A

【分析】由平均值、方差的性质求新数据的平均数和方差.

【详解】由 $\overline{x}=4$，得样本数据 $2x_1+1$，$2x_2+1$，$\cdots$，$2x_n+1$ 的平均数为 $2\overline{x}+1=2\times4+1=9$，

由 $s^2=1$，得样本数据 $2x_1+1$，$2x_2+1$，$\cdots$，$2x_n+1$ 的方差为 $4s^2=4$.

故选：A

22．若数据 $x_1,x_2,x_3,\cdots,x_n$ 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．$3s$

【答案】D

【分析】根据线性变化前后数据的方差的关系求解.

【详解】因为数据 $x_1,x_2,x_3,\cdots,x_n$ 的标准差为 s，

由数据方差的性质，可得数据 $3x_1+1$，$3x_2+1$，…，$3x_n+1$ 的标准差为 $\sqrt{3^2s^2}=3s$，

故选：D.

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

【答案】A

【分析】根据平均数、方差的性质计算可得.

【详解】因为 $x_1,x_2,x_3,\cdots,x_8$ 的平均数是 10，方差是 10，

所以 $3x_1+2,3x_2+2,3x_3+2,\cdots,3x_8+2$ 的平均数是 $3\times10+2=32$，方差是 $3^2\times10=90$.

故选：A.

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 $\left[E(X)\right]^2\geq E(X^2)$

【答案】D

【分析】根据方差以及标准差的性质即可求解 A；结合二项分布的概率公式，即可求解 B；结合正态分布的对称性，即可求解 C；结合方差的非负性，即可求解 D.

【详解】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 $\sqrt{9S^2}=3S$，故 A 正确；

$X\sim B(4,p)$，$D(X)=\dfrac{3}{4}$，则 $4p(1-p)=\dfrac{3}{4}$，得 $p(1-p)=\dfrac{3}{16}$，

$P(X=2)=\mathrm{C}_4^2p^2(1-p)^2=6\left[p(1-p)\right]^2=6\times\left(\dfrac{3}{16}\right)^2=\dfrac{27}{128}$，故 B 正确；

$X\sim N(1,\sigma^2)$，$P(X>0)=0.75$，

则 $P(0<X<2)=2P(0<X<1)=2[P(X>0)-P(X>1)]=2\times(0.75-0.5)=0.5$，故 C 正确；

X 为取有限个值的离散型随机变量，

则 $D(X)=E(X^2)-[E(X)]^2\geq 0$，故 D 错误.

故选：D.

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．$4,4$

【答案】B

【分析】根据样本数据同加上一个数和同乘以一个数后的新数据的平均值和方差的性质，即可求得答案.

【详解】由题意知样本数据 $x_1,x_2,\cdots,x_{100}$ 的平均数和标准差均为 4，则 $x_1,x_2,\cdots,x_{100}$ 的方差为 16，

则 $-x_1,-x_2,\cdots,-x_{100}$ 的平均数为 -4，方差为 $(-1)^2\times16=16$，

故 $-x_1-1,-x_2-1,\cdots,-x_{100}-1$ 的平均数为 $-4-1=-5$，方差 16，

故选：B

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

【答案】D

【分析】根据平均数以及方差的性质即可求解.

【详解】由于数据x_1，x_2，x_3，x_4，x_5的平均数是2，方差是3，故数据：$2x_1+1$，$2x_2+1$，$2x_3+1$，$2x_4+1$，$2x_5+1$的平均数是$2\times 2+1=5$，方差是$2^2\times 3=12$，

故选：D

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．$7A$，$49s^2$

【答案】D

【分析】根据一组数据同乘以一个数后的平均数以及方差的性质计算即可.

【详解】由题意知这些商品的价格如果按人民币计算，价格是按美元计算的价格的7倍，故按人民币计，则平均数和方差分别为$7A$, $7^2\times s^2=49s^2$.

故选：D.

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．$\frac{5}{2}$

【答案】D

【分析】由平均数和方差的运算性质即可求解.

【详解】由方差的性质，得ax_1+5，ax_2+5，…，ax_n+5的方差为a^2s^2，

故$a^2=4$，解得$a=\pm 2$．由$a>0$，可知$a=2$.

由平均数的性质，得ax_1+5，ax_2+5，…，ax_n+5的平均数为$a\overline{x}+5$，

故$a\overline{x}+5=2\overline{x}+5=4\overline{x}$，

解得$\overline{x}=\frac{5}{2}$.

故选：D.

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，L ，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．$\frac{1}{8}$

【答案】C

【分析】

根据两组数据的线性关系确定它们的平均数与标准差的关系列方程，即可得a,b的值，从而可得答案.

【详解】由题可知，$\begin{cases}3a+b=10\\a^2\times1=4^2\\a>0\end{cases}$，解得$\begin{cases}a=4\\b=-2\end{cases}$，则$a^b=4^{-2}=\frac{1}{16}$.

故选：C.

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．$\frac{3}{4}$，$\frac{5}{8}$

【答案】D

【分析】

利用平均数与方差的运算性质求解即可.

【详解】设数据x_1，x_2，…，x_{10}的平均数和方差分别为μ和s^2，

则数据$4x_1+1$，$4x_2+1$，…，$4x_{10}+1$的平均数为$4\times\mu+1=4$，方差为$4^2\times s^2=10$，

得$\mu=\frac{3}{4}$，$s^2=\frac{5}{8}$，

故选：D.

考点 4 总体百分位数的估计

百 分 位 数

（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.

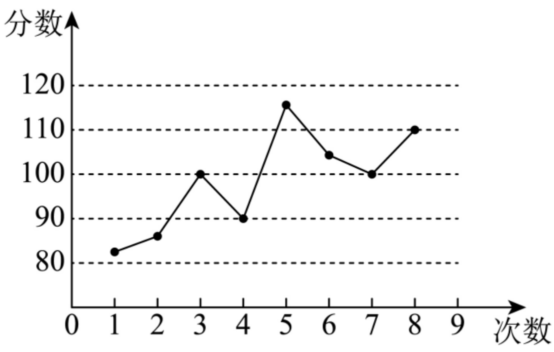
②计算$i = n \times p\%$.

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.

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

【答案】C

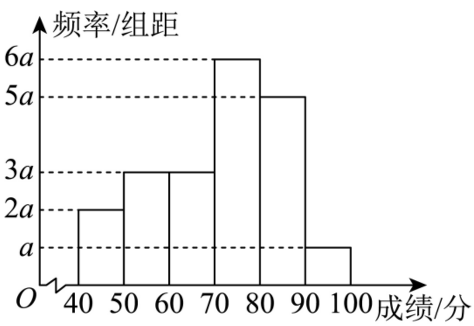
【分析】根据百分位数定义求解即可.

【详解】因为$8 \times 80\% = 6.4$，由图可知8次成绩由小到大排序，

第7个位置的数是110，所以这8次成绩的第80百分位数是110.

故选：C.

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

【答案】B

【分析】由频率分布直方图性质求a，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组频率和为$14\times0.005\times10=0.7$，前5组频率和为$19\times0.005\times10=0.95$，

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：B

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

【答案】C

【分析】由小到大排列数据，再由百分数求法按步骤求解即可.

【详解】将8天的数据由小到大排列：$8,8,11,12,14,16,18,21$.

因为$8\times75\%=6$，6是整数，

故第这8天的气温数据的75%分位数为$\frac{16+18}{2}=17$.

故选：C.

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

【答案】B

【分析】根据样本数据百分位数的定义求解即可.

【详解】将数据从小到大排序得$72,78,80,81,83,86,88,90$，

因为$8\times75\%=6$，

所以第75百分位数是$\frac{86+88}{2}=87$.

故选：B.

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

【答案】C

【分析】由百分位数的定义，求出第 65 百分位数是这组数据从小到大排列的第几个数,即可得到答案.

【详解】已知一组数据：4，6，7，9，11，13，共 6 个数，

则 $6\times 65\% = 3.9$，

所以这组数据的第 65 百分位数为从小到大排列的第四个数 9.

故选：C.

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 $2x_1+1$，$2x_2+1$，$\cdots$ 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 $r \in (-1,1)$

【答案】C

【分析】根据百分位数的概念可判断 A 的真假；根据两组相关数据的平均数和方差的计算方法判断 B 的真假；计算残差判断 C 的真假；根据相关系数的取值范围判断 D.

【详解】对 A：将 3，3，8，4，2，7，10，18 由小到大排列为 2，3，3，4，7，8，10，18，第 50 百分位数即为中位数，这组数的中位数为 $\frac{1}{2}\times(4+7) = 5.5$，所以 A 错误；

对 B：由数据 $x_1, x_2, \cdots$ 的平均数为 2，方差为 3，则数据 $2x_1+1$，$2x_2+1$，$\cdots$ 的平均数为 $2\times 2+1=5$，方差为 $2^2\times 3 = 12$，所以 B 错误；

对 C：残差 $\hat{e}_i = y_i - \hat{y}_i = y_i - \hat{b}x_i - \hat{a} = 1.2 - 0.25\times 2 - 1.5 = -0.8$，故 C 正确；

对 D：样本的相关系数应满足 $-1 \le r \le 1$，所以 D 错误.

故选：C

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

【答案】B

【分析】由百分位数、中位数的定义即可求解.

【详解】将这 12 个数据从小到大排列为$8,9,9,11,12,13,15,16,17,18,18,20$，

$60\%\times 12=7.2$，所以排列后的第 8 个数即为第 60 百分位数：16，

中位数为$\frac{13+15}{2}=14$，故所求为：$14+16=30$.

故选：B.

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

【答案】C

【分析】根据给定条件，利用第 60 百分位数的定义求解即得.

【详解】样本数据 12，8，32，10，24，22，12，33，按从小到大排序为 8，10，12，12，22，24，32，33，

由$8\times 60\%=4.8$，得样本数据的第 60 百分位数为升序排列的第五个数，即 22.

故选：C

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

【答案】C

【分析】由百分位数的定义，先将样本数据从小到大排列，再计算第 60 百分位数为第 6 和第 7 个数的平均数即可.

【详解】将这组数据从小到大排列为$17,23,26,30,31,33,33,36,42,106$，

数据的第 60 百分位数为$\frac{33+33}{2}=33$，

故选：C.

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

【答案】B

【分析】把样本数据由小到大排列，再利用第一四分位数的定义求解即得.

【详解】样本数据由小到大排列为 11 ，11，12 ，13 ，13 ，14 ，15 ，15，16 ，

由$9\times 25\%=2.25$，得样本数据的第一四分位数为 12.

故选：B

考点5 相关关系与相关系数

1、变量之间的相关关系

当自变量取值一定时，因变量的取值带有一定的随机性，则这两个变量之间的关系叫相关关系．由于相关关系的不确定性，在寻找变量之间相关关系的过程中，统计发挥着非常重要的作用．我们可以通过收集大量的数据，在对数据进行统计分析的基础上，发现其中的规律，对它们的关系作出判断．

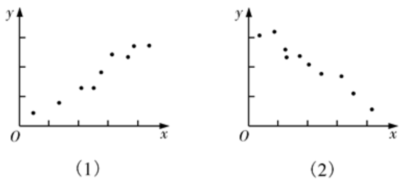
注意：相关关系与函数关系是不同的，相关关系是一种非确定的关系，函数关系是一种确定的关系，而且函数关系是一种因果关系，但相关关系不一定是因果关系，也可能是伴随关系．

2、散点图

将样本中的n个数据点$(x_i,y_i)(i=1,2,\cdots,n)$描在平面直角坐标系中，所得图形叫做散点图．根据散点图中点的分布可以直观地判断两个变量之间的关系．

（1）如果散点图中的点散布在从左下角到右上角的区域内，对于两个变量的这种相关关系，我们将它称为正相关，如图（1）所示；

（2）如果散点图中的点散布在从左上角到右下角的区域内，对于两个变量的这种相关关系，我们将它称为负相关，如图（2）所示．



3、相关系数

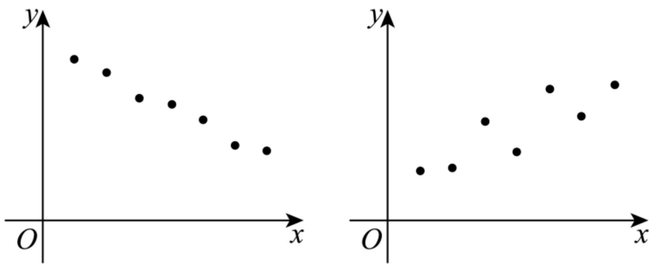
若相应于变量x的取值x_i，变量y的观测值为$y_i(1\le i\le n)$，则变量x与y的相关系数$r=\dfrac{\sum\limits_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum\limits_{i=1}^{n}(x_i-\bar{x})^2\sum\limits_{i=1}^{n}(y_i-\bar{y})^2}}=\dfrac{\sum\limits_{i=1}^{n}x_iy_i-n\bar{x}\bar{y}}{\sqrt{\sum\limits_{i=1}^{n}x^2{}_i-n\bar{x}^2}\sqrt{\sum\limits_{i=1}^{n}y^2{}_i-n\bar{y}^2}}$，通常用$r$来衡量$x$与$y$之间的线性关系的强弱，$r$的范围为$-1\le r\le 1$．

（1）当$r>0$时，表示两个变量正相关；当$r<0$时，表示两个变量负相关．

（2）$|r|$越接近1，表示两个变量的线性相关性越强；$|r|$越接近0，表示两个变量间几乎不存在线性相关关系．当$|r|=1$时，所有数据点都在一条直线上．

（3）通常当$|r|>0.75$时，认为两个变量具有很强的线性相关关系.

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．$r_1+r_2<0$

【答案】D

【分析】由两散点图中散点的位置关系直接得答案.

【详解】由散点图可知，x与y负相关，v与u正相关，则$b_1<0$，$b_2>0$，故A、B错误；

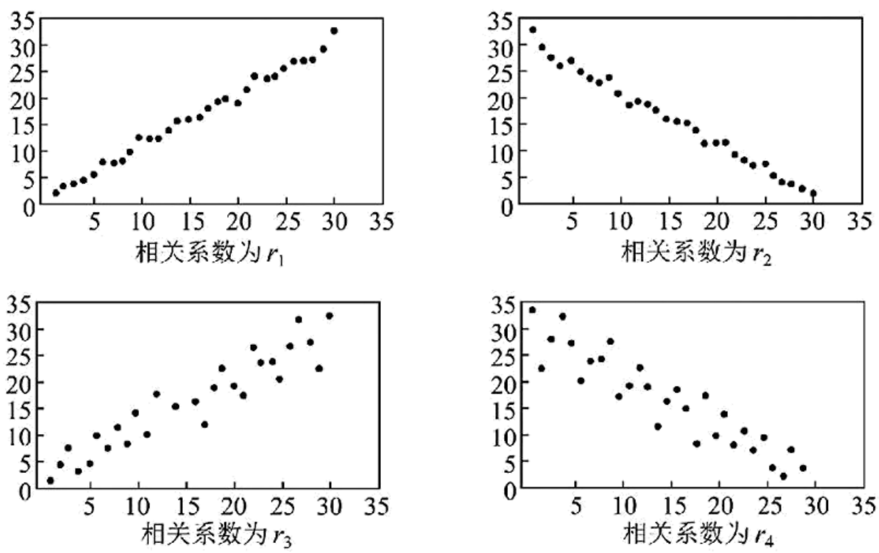
且图形中点(x,y)比(v,u)更加集中在一条直线附近，

则$|r_1|>|r_2|$，又$r_1<0$，$r_2>0$，得$r_1+r_2<0$.

故C错误，D正确.

故选：D.

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（　）



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/977026160133010020