**第一讲随机抽样与用样本估计总体**

id:2147491762;FounderCES

题组1随机抽样

1*.*[2015湖北,2,5分][理][数学文化题]我国古代数学名著《数书九章》有“米谷粒分”题:粮仓开仓收粮,有人送来米1 534石,验得米内夹谷,抽样取米一把,数得254粒内夹谷28粒,则这批米内夹谷约为()

A.134石 B.169石 C.338石 D.1 365石

2*.*[2014湖南,3,5分]对一个容量为*N*的总体抽取容量为*n*的样本,当选取简单随机抽样、系统抽样和分层抽样三种不同方法抽取样本时,总体中每个个体被抽中的概率分别为*p*1,*p*2,*p*3,则()

A*.p*1*=p*2*<p*3 B*.p*2*=p*3*<p*1

C*.p*1*=p*3*<p*2 D*.p*1*=p*2*=p*3

3*.*[2013新课标全国Ⅰ,3,5分][理]为了解某地区的中小学生的视力情况,拟从该地区的中小学生中抽取部分学生进行调查,事先已了解到该地区小学、初中、高中三个学段学生的视力情况有较大差异,而男女生视力情况差异不大*.*在下面的抽样方法中,最合理的抽样方法是()

A.简单随机抽样 B.按性别分层抽样

C.按学段分层抽样 D.系统抽样

4*.*[2017江苏,3,5分][理]某工厂生产甲、乙、丙、丁四种不同型号的产品,产量分别为200,400,300,100件*.*为检验产品的质量,现用分层抽样的方法从以上所有的产品中抽取60件进行检验,则应从丙种型号的产品中抽取件*.*

题组2统计图表的绘制及应用

5*.*[2017全国卷Ⅲ,3,5分][理]某城市为了解游客人数的变化规律,提高旅游服务质量,收集并整理了2014年1月至2016年12月期间月接待游客量(单位:万人)的数据,绘制了下面的折线图(图14*-*1*-*1)*.*



图14*-*1*-*1

根据该折线图,下列结论错误的是()

A.月接待游客量逐月增加

B.年接待游客量逐年增加

C.各年的月接待游客量高峰期大致在7,8月

D.各年1月至6月的月接待游客量相对于7月至12月,波动性更小,变化比较平稳

6*.*[2016山东,3,5分][理] 某高校调查了200名学生每周的自习时间(单位:小时),制成了如图14*-*1*-*2所示的频率分布直方图,其中自习时间的范围是[17*.*5,30],样本数据分组为[17*.*5,20),[20,22*.*5),[22*.*5,25),[25,27*.*5),[27*.*5,30]*.*根据直方图,这200名学生中每周的自习时间不少于22*.*5小时的人数是()

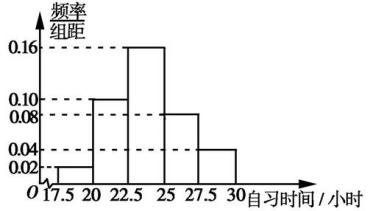


图14*-*1*-*2

A.56 B.60 C.120 D.140

7*.*[2014广东,6,5分][理]已知某地区中小学生人数和近视情况如图14*-*1*-*3所示*.*为了解该地区中小学生的近视形成原因,用分层抽样的方法抽取2*%*的学生进行调查,则样本容量和抽取的高中生近视人数分别为()

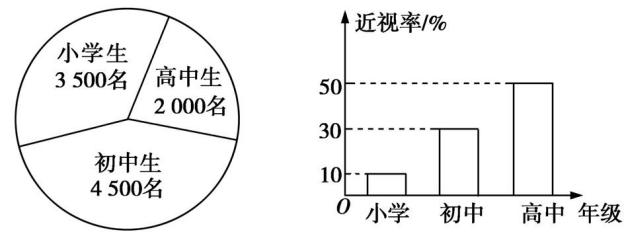


图14*-*1*-*3

A.100,10 B.200,10

C.100,20 D.200,20

8*.*[2015新课标全国Ⅱ,18,12分][理]某公司为了解用户对其产品的满意度,从A,B两地区分别随机调查了20个用户,得到用户对产品的满意度评分如下:

A地区:62738192958574645376

78869566977888827689

B地区:73836251914653736482

93486581745654766579

(Ⅰ)根据两组数据完成两地区用户满意度评分的茎叶图如图14*-*1*-*4,并通过茎叶图比较两地区满意度评分的平均值及分散程度(不要求计算出具体值,给出结论即可);

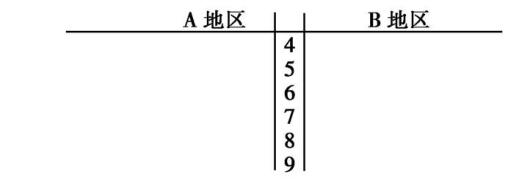


图14*-*1*-*4

(Ⅱ)根据用户满意度评分,将用户的满意度从低到高分为三个等级:

|  |  |  |  |
| --- | --- | --- | --- |
| 满意度评分 | 低于70分 | 70分到89分 | 不低于90分 |
| 满意度等级 | 不满意 | 满意 | 非常满意 |

记事件*C*:“A地区用户的满意度等级高于B地区用户的满意度等级”*.*假设两地区用户的评价结果相互独立*.*根据所给数据,以事件发生的频率作为相应事件发生的概率,求*C*的概率*.*

9*.*[2015 广东,17,12分]某城市100户居民的月平均用电量(单位:度),以[160,180),[180,200),[200,220),[220,240),[240,260),[260,280),[280,300]分组的频率分布直方图如图14*-*1*-*5所示*.*

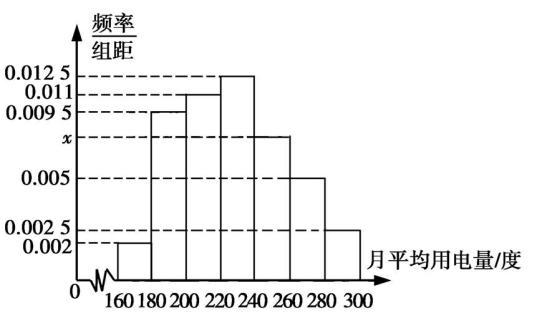


图14*-*1*-*5

(1)求直方图中*x*的值;

(2)求月平均用电量的众数和中位数;

(3)在月平均用电量为[220,240),[240,260),[260,280),[280,300]的四组用户中,用分层抽样的方法抽取11户居民,则月平均用电量在[220,240)的用户中应抽取多少户?

题组3求样本的数字特征

10*.*[2017山东,8,5分]如图14*-*1*-*6所示的茎叶图记录了甲、乙两组各5名工人某日的产量数据(单位:件)*.*若这两组数据的中位数相等,且平均值也相等,则*x*和*y*的值分别为()

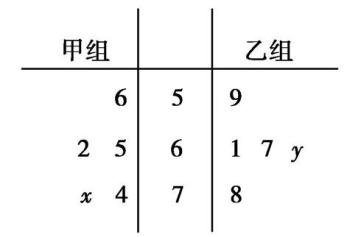


图14*-*1*-*6

A.3,5 B.5,5 C.3,7 D.5,7

11*.*[2015 安徽,6,5分][理]若样本数据*x*1,*x*2,…,*x*10的标准差为8,则数据2*x*1*-*1,2*x*2*-*1,…,2*x*10*-*1的标准差为()

A*.*8 B*.*15 C*.*16 D*.*32

12*.*[2014陕西,9,5分]某公司10位员工的月工资(单位:元)为*x*1,*x*2,…,*x*10,其均值和方差分别为和*s*2,若从下月起每位员工的月工资增加100元,则这10位员工下月工资的均值和方差分别为()

A.,*s*2*+*1002 B.*+*100,*s*2*+*1002

C.,*s*2 D.*+*100,*s*2

13*.*[2016江苏,4,5分][理]已知一组数据4*.*7,4*.*8,5*.*1,5*.*4,5*.*5,则该组数据的方差是*.*

14*.*[2015 广东,12,5分]已知样本数据*x*1,*x*2,…,*xn*的均值*=*5,则样本数据2*x*1*+*1,2*x*2*+*1,…,2*xn+*1的均值为*.*

15*.*[2015 广东,17,12分][理]某工厂36名工人的年龄数据如下表*.*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 工人编号 | 年龄 | 工人编号 | 年龄 | 工人编号 | 年龄 | 工人编号 | 年龄 |
| 1 | 40 | 10 | 36 | 19 | 27 | 28 | 34 |
| 2 | 44 | 11 | 31 | 20 | 43 | 29 | 39 |
| 3 | 40 | 12 | 38 | 21 | 41 | 30 | 43 |
| 4 | 41 | 13 | 39 | 22 | 37 | 31 | 38 |
| 5 | 33 | 14 | 43 | 23 | 34 | 32 | 42 |
| 6 | 40 | 15 | 45 | 24 | 42 | 33 | 53 |
| 7 | 45 | 16 | 39 | 25 | 37 | 34 | 37 |
| 8 | 42 | 17 | 38 | 26 | 44 | 35 | 49 |
| 9 | 43 | 18 | 36 | 27 | 42 | 36 | 39 |

(1)用系统抽样法从36名工人中抽取容量为9的样本,且在第一分段里用随机抽样法抽到的年龄数据为44,列出样本的年龄数据;

(2)计算(1)中样本的均值和方差*s*2;

(3)36名工人中年龄在*-s*与*+s*之间有多少人?所占的百分比是多少(精确到0*.*01*%*)?

id:2147491848;FounderCES

**A组基础题**

1*.*[2018陕西省部分学校摸底检测,5]某单位有老年人27人,中年人54人,青年人81人,为了调查他们的身体状况的某项指标,需从他们中间抽取一个容量为42的样本,则分别应抽取老年人、中年人、青年人的人数是()

A.7,11,18 B.6,12,18 C.6,13,17 D.7,14,21

2*.*[2018黔东南州第一次联考,3]近年呼吁高校招生改革的呼声越来越高,在赞成高校招生改革的市民中按年龄分组,得到样本频率分布直方图如图14*-*1*-*7所示,其中年龄在区间[30,40)内的有2 500人,在区间[20,30)内的有1 200人,则*m*的值为()

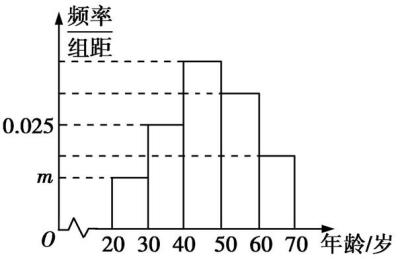


图14*-*1*-*7

A.0.013 B.0.13 C.0.012 D.0.12

3*.*[2018长春市第一次质量监测,5]已知某班级部分同学一次测验的成绩统计如图14*-*1*-*8所示,则其中位数和众数分别为()

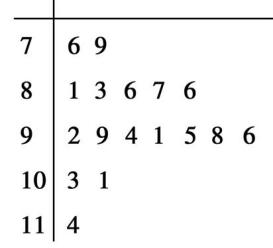
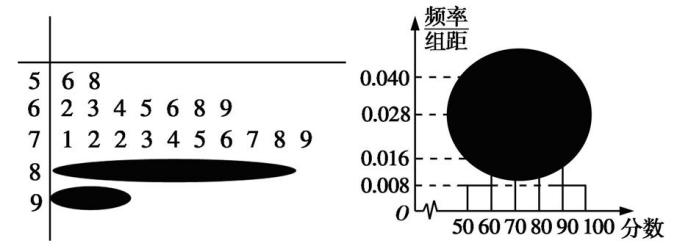


图14*-*1*-*8

A.95,94 B.92,86 C.99,86 D.95,91

4*.*[2017青海省西宁市高三一检,3]某班一次测试成绩的茎叶图和频率分布直方图的可见部分如图14*-*1*-*9(1)(2)所示,根据图中的信息可确定被抽测的人数及分数在[90,100]内的人数分别为()



(1)(2)

图14*-*1*-*9

A.20,2 B.24,4 C.25,2 D.25,4

5*.*[2018南昌市调考,13]某校高三(2)班现有64名学生,随机编号为0,1,2,…,63,依编号顺序平均分成8组,组号依次为1,2,3,…,8*.*现用系统抽样方法抽取一个容量为8的样本,若在第1组中随机抽取的号码为5,则在第6组中抽取的号码为*.*

6*.*[2018广雅中学、东华中学、河南名校高三第一次联考,14][数学文化题]《九章算术》第三章“衰分”中有如下问题:“今有甲持钱五百六十,乙持钱三百五十,丙持钱一百八十,凡三人俱出关,关税百钱,欲以钱数多少衰出之,问各几何?”其意为:“今有甲带了560钱,乙带了350钱,丙带了180钱,三人一起出关,共需要交关税100钱,依照钱的多少按比例出钱”,则丙应出钱(结果保留整数)*.*

7*.*[2018唐山市五校联考,18]某篮球队在本赛季已结束的8场比赛中,队员甲得分统计的茎叶图如图14*-*1*-*10所示*.*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
| 0 | 7 | 8 |  |  |
| 1 | 0 | 5 | 7 | 9 |
| 2 | 1 | 3 |  |  |

图14*-*1*-*10

(1)求甲在比赛中得分的均值和方差;

(2)从甲比赛得分在20分以下的6场比赛中随机抽取2场进行失误分析,求抽到2场的得分都不超过均值的概率*.*

**B组提升题**

8*.*[2018益阳市、湘潭市高三调考,18]某中学为了了解全校学生的上网情况,在全校采取随机抽样的方法抽取了80名学生(其中男女生人数恰好各占一半)进行问卷调查,并进行了统计,按性别分为两组,再将每组学生的月上网次数按[0,5),[5,10),[10,15),[15,20),[20,25]分为5组,得到如图14*-*1*-*11所示的频率分布直方图*.*

(1)写出*a*的值;

(2)求80名学生中月上网次数不少于15次的学生人数;

(3)在80名学生中,从月上网次数少于5次的学生中随机抽取2人,求至少抽取到1名男生的概率*.*

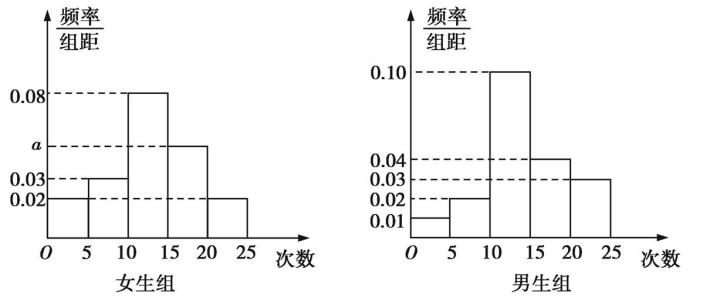


图14*-*1*-*11

9*.*[2018惠州市一调,18]某大学生在开学季准备销售一种文具盒进行试创业,在一个开学季内,每售出1盒该产品获得的利润为30元,未售出的产品,每盒亏损10元*.*该大学生通过查询资料得到开学季市场需求量的频率分布直方图,如图14*-*1*-*12所示*.*该大学生为这个开学季购进了160盒该产品,以*x*(单位:盒,100 ≤ *x* ≤ 200 )表示这个开学季内的市场需求量,*y*(单位:元)表示这个开学季内经销该产品的利润*.*

(1)根据直方图估计这个开学季内市场需求量*x*的众数和平均数;

(2)将*y*表示为*x*的函数;

(3)根据直方图估计利润*y*不少于4 000元的概率*.*

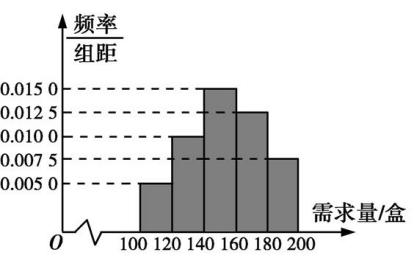


图14*-*1*-*12

10*.*[2017宁夏银川市、吴忠市部分重点中学高三联考,18]某教师为了了解本校高三学生一模考试的数学成绩,将所教两个班级的数学成绩(单位:分)绘制成如图14*-*1*-*13所示的茎叶图*.*

|  |  |  |
| --- | --- | --- |
| 甲班 |  | 乙班 |
|  | 8 | 8 |
| 9 9 8 7 6 5 5 4 3 1 1 0 | 9 | 0 1 1 2 2 2 2 4 6 7 8 8 9 |
| 9 8 8 7 7 6 5 5 4 3 3 3 3 0 | 10 | 0 1 1 1 1 2 2 5 6 6 7 9 9 |
| 9 7 5 5 5 2 2 1 0 | 11 | 0 3 4 5 5 6 7 9 |
| 9 8 6 5 3 3 1 0 0 | 12 | 0 1 2 2 3 3 5 8 9 |
| 6 2 1 1 | 13 | 0 1 8 |
| 6 0 | 14 | 3 |

图14*-*1*-*13

(1)分别求出甲、乙两个班级数学成绩的中位数、众数;

(2)若规定成绩大于等于115分为优秀,分别求出两个班级数学成绩的优秀率;

(3)在(2)的条件下,若用甲班学生数学成绩的频率估计概率,从该校高三年级中随机抽取3人,记这3人中数学成绩优秀的人数为*X*,求*X*的分布列和数学期望*.*

**答案**

id:2147495203;FounderCES

1*.*B依题意,可得这批米内夹谷约为*×*1 534≈169石,选B*.*

2*.*D根据抽样方法的概念可知,简单随机抽样、系统抽样和分层抽样这三种抽样方法中每个个体被抽到的概率都是*p=*,故*p*1*=p*2*=p*3,故选D*.*

3*.*C由于该地区的中小学生人数比较多,不能采用简单随机抽样,排除选项A;由于小学、初中、高中三个学段的学生视力差异比较大,可采取按照学段进行分层抽样,而男女生视力情况差异不大,不能按照性别进行分层抽样,排除B和D,选C.

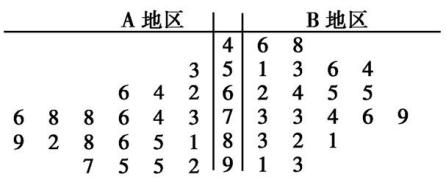
4*.*18应从丙种型号的产品中抽取60*×=*18(件)*.*

5*.*A根据题中折线图可知,2014年8月到9月、2014年10月到11月等期间月接待游客量都是减少,所以A错误*.*

6*.*D由题中的频率分布直方图可知,这200名学生每周的自习时间不少于22*.*5小时的频率为(0*.*16*+*0*.*08*+*0*.*04)*×*2*.*5*=*0*.*7,故这200名学生中每周的自习时间不少于22*.*5小时的人数为200*×*0*.*7*=*140*.*故选D*.*

7*.*D由题意知,样本容量为(3 500*+*4 500*+*2 000)*×*2*%=*200;抽取的高中生人数为2 000*×*2*%=*40,由于其近视率为50*%*,所以抽取的高中生近视的人数为40*×*50*%=*20*.*故选D*.*

8*.*(Ⅰ)两地区用户满意度评分的茎叶图如图D 14*-*1*-*2所示,



图D 14*-*1*-*2

通过茎叶图可以看出,A地区用户满意度评分的平均值高于B地区用户满意度评分的平均值;A地区用户满意度评分比较集中,B地区用户满意度评分比较分散*.*

(Ⅱ)记*CA*1表示事件“A地区用户的满意度等级为满意或非常满意”;

*CA*2表示事件“A地区用户的满意度等级为非常满意”;

*CB*1表示事件“B地区用户的满意度等级为不满意”;

*CB*2表示事件“B地区用户的满意度等级为满意”,

则*CA*1与*CB*1独立,*CA*2与*CB*2独立,*CB*1与*CB*2互斥,*C=CB*1*CA*1∪*CB*2*CA*2*.*

*P*(*C*)*=P*(*CB*1*CA*1∪*CB*2*CA*2)

*=P*(*CB*1*CA*1)*+P*(*CB*2*CA*2)

*=P*(*CB*1)*P*(*CA*1)*+P*(*CB*2)*P*(*CA*2)*.*

由所给数据得*CA*1,*CA*2,*CB*1,*CB*2发生的频率分别为,,,,故*P*(*CA*1)*==*,*P*(*CA*2)*==*,*P*(*CB*1)*==*,*P*(*CB*2)*==*,

*P*(*C*)*=×+×=*0*.*48*.*

9*.*(1)依题意,20*×*(0*.*002*+*0*.*009 5*+*0*.*011*+*0*.*012 5*+x+*0*.*005*+*0*.*002 5)*=*1,解得*x=*0*.*007 5*.*

(2)由题图可知,最高矩形的数据组为[220,240),

∴众数为*=*230*.*

*∵*[160,220)的频率之和为(0*.*002*+*0*.*009 5*+*0*.*011)*×*20*=*0*.*45,

∴依题意,设中位数为*y*,则0*.*45*+*(*y-*220)*×*0*.*012 5*=*0*.*5,解得*y=*224,

∴中位数为224*.*

(3)月平均用电量在[220,240)内的用户在四组用户中所占比例为*=*,

∴月平均用电量在[220,240)的用户中应抽取11*×=*5(户)*.*

10*.*A根据两组数据的中位数相等可得65*=*60*+y*,解得*y=*5,又它们的平均值相等,所以*=*,解得*x=*3*.*故选A*.*

11*.*C令*yi=*2*xi-*1(*i=*1,2,3,…,10),则*σ*(*y*)*=*2*σ*(*x*)*=*16*.*故选C*.*

12*.*D解法一对平均数和方差的意义深入理解可巧解*.*因为每个数据都加上了100,故平均数也增加100,而离散程度应保持不变,故选D*.*

解法二由题意知*x*1*+x*2*+*…*+xn=n*,*s*2*=*[(*x*1*-*)2*+*(*x*2*-*)2*+*…*+*(*xn-*)2],

则所求均值*=*[(*x*1*+*100)*+*(*x*2*+*100)*+*…*+*(*xn+*100)]*=*(*n+n×*100)*=+*100,

而所求方差*t*2*=*[(*x*1*+*100*-*)2*+*(*x*2*+*100*-*)2*+*…*+*(*xn+*100*-*)2]*=*[(*x*1*-*)2*+*(*x*2*-*)2*+*…*+*(*xn-*)2]*=s*2,故选D*.*

13*.*0*.*1这组数据的平均数*==*5*.*1,则方差*s*2*=*

*=*

*=*0*.*1*.*

14*.*11由*=*5,得*=*2*×+*1*=*2*+*1*=*11*.*

15*.*(1)由系统抽样的知识可知,36人分成9组,每组4人,其中第一组中抽到工人年龄为44,所以其编号为2,故所有样本数据的编号为4*n-*2,*n=*1,2,…,9*.*其数据为44,40,36,43,36,37,44,43,37*.*

(2)*==*40*.*

由方差公式知,*s*2*=×*[(44*-*40)2*+*(40*-*40)2*+*…*+*(37*-*40)2]*=.*

(3)因为*s*2*=*,所以*s=*∈(3,4),

所以36名工人中年龄在*-s*和*+s*之间的人数等于年龄在区间[37,43]内的人数,即40,40,41,…,39,共23人*.*

所以36名工人中年龄在*-s*和*+s*之间的人数所占的百分比为≈63*.*89*%.*

id:2147495217;FounderCES

**A组基础题**

1*.*D因为该单位共有27*+*54*+*81*=*162(人),样本容量为42,所以应当按*=*的比例分别从老年人、中年人、青年人中抽取样本,且分别应抽取的人数是7,14,21*.*故选D*.*

2*.*C由题意,得年龄在区间[30,40)内的频率为0*.*025*×*10*=*0*.*25,则赞成高校招生改革的市民有*=*10 000(人),因为年龄在区间[20,30)内的有1 200人,所以*m==*0*.*012*.*

3*.*B由题中茎叶图可知,此组数据由小到大排列依次为76,79,81,83,86,86,87,91,92,94,95,

96,98,99,101,103,114,共17个,故中位数为92,出现次数最多的为众数,故众数为86,故选B*.*

4*.*C解法一频率分布直方图中,[50,60)对应的频率为0*.*008*×*10*=*0*.*08,而茎叶图中[50,60)内有2(人),所以总人数为*=*25,又[90,100]的频率为0*.*08,所以分数在[90,100]内的有0*.*08*×*25*=*2(人),故选C*.*

解法二由频率分布直方图可知[50,60)和[90,100]的频率相同,所以[90,100]内的人数和[50,60)内的人数一样多,而[50,60)内有2人,所以[90,100]内也有2人,排除B,D;又[50,60)和[90,100]内各有2人,[60,70)内有7人,[70,80)内有10人,共21人,所以总人数应该大于21,排除A,选C*.*

5*.*45依题意,分组间隔为*=*8,因为在第1组中随机抽取的号码为5,所以在第6组中抽取的号码为5*+*5*×*8*=*45*.*

6*.*17因为依照钱的多少按比例出钱,所以丙应该出的钱为*×*100*=×*100≈17*.*

7*.*(1)甲在比赛中得分的均值*=×*(7*+*8*+*10*+*15*+*17*+*19*+*21*+*23)*=*15,

方差*s*2*=×*[(*-*8)2*+*(*-*7)2*+*(*-*5)2*+*02*+*22*+*42*+*62*+*82]*=*32*.*25*.*

(2)题设所述的6场比赛中甲的得分分别为:7,8,10,15,17,19*.*

从中随机抽取2场,这2场比赛的得分如下:(7,8),(7,10),(7,15),(7,17),(7,19),(8,10),(8,15),

(8,17),(8,19),(10,15),(10,17),(10,19),(15,17),(15,19),(17,19),共15种,

其中抽到2场的得分都不超过均值的情况有(7,8),(7,10),(7,15),(8,10),(8,15),(10,15),共6种,

所以所求概率*P==.*

**Ｂ组提升题**

8*.*(1)*a==*0*.*05*.*

(2)在所抽取的女生中,月上网次数不少于15次的学生人数的频率为(0*.*05*+*0*.*02)*×*5*=*0*.*35,

所以月上网次数不少于15次的女生有0*.*35*×*40*=*14(人)*.*

在所抽取的男生中,月上网次数不少于15次的学生人数的频率为(0*.*04*+*0*.*03)*×*5*=*0*.*35,

所以月上网次数不少于15次的男生有0*.*35*×*40*=*14(人)*.*

故抽取的80名学生中月上网次数不少于15次的学生有28人*.*

(3)记“在80名学生中,从月上网次数少于5次的学生中随机抽取2人,至少抽到1名男生”为事件*A*,

在抽取的女生中,月上网次数少于5次的学生人数的频率为0*.*02*×*5*=*0*.*1,人数为0*.*1*×*40*=*4,

在抽取的男生中,月上网次数少于5次的学生人数的频率为0*.*01*×*5*=*0*.*05,人数为0*.*05*×*40*=*2,

则在80名学生中,从月上网次数少于5次的学生中随机抽取2人,所有可能的结果有15种,而事件*A*包含的结果有9种,所以*P*(*A*)*==.*

9*.*(1)由题中频率分布直方图得,这个开学季内市场需求量*x*的众数是150盒,

需求量在[100,120)内的频率为0*.*005 0*×*20*=*0*.*1,

需求量在[120,140)内的频率为0*.*010 0*×*20*=*0*.*2,

需求量在[140,160)内的频率为0*.*015 0*×*20*=*0*.*3,

需求量在[160,180)内的频率为0*.*012 5*×*20*=*0*.*25,

需求量在[180,200]内的频率为0*.*007 5*×*20*=*0*.*15*.*

则平均数 *=*110*×*0*.*1*+*130*×*0*.*2*+*150*×*0*.*3*+*170*×*0*.*25*+*190*×*0*.*15*=*153(盒)*.*

(2)因为每售出1盒该产品获得的利润为30元,未售出的产品,每盒亏损10元,

所以当100≤*x<*160时,*y=*30*x-*10*×*(160*-x*)*=*40*x-*1 600;

当160≤*x*≤ 200时,*y=*160*×*30*=*4 800*.*

所以*y=*

(3)因为利润*y*不少于4 000元,所以当100≤*x<*160时,由40*x-*1 600≥4 000,解得140≤*x<*160;

当160≤*x*≤200时,*y=*4 800*>*4 000恒成立,所以40≤*x*≤200时,利润*y*不少于4 000元*.*

故由(1)知利润*y*不少于4 000元的概率*P=*1*-*0*.*1*-*0*.*2*=*0*.*7*.*

10*.*(1)由题中所给的茎叶图知,甲班50名同学的成绩由小到大排序,排在第25,26位的是108,109,数量最多的是103,故甲班数学成绩的中位数是108*.*5,众数是103;

乙班48名同学的成绩由小到大排序,排在第24,25位的是106,107,数量最多的是92和101,故乙班数学成绩的中位数是106*.*5,众数为92和101*.*

(2)由题中茎叶图中的数据可知,甲班中数学成绩优秀的人数为20,优秀率为*=*;乙班中数学成绩优秀的人数为18,优秀率为*=.*

(3)用甲班学生数学成绩的频率估计概率,则该校高三学生数学成绩的优秀率*P=.X*的所有可能取值为0,1,2,3,*X*服从二项分布,即*X~B*(3,),

*P*(*X=*0)*=×*()3*=*;

*P*(*X=*1)*=××*()2*=*;

*P*(*X=*2)*=×*()2*×=*;

*P*(*X=*3)*=×*()3*=.*

因此*X*的分布列为

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *X* | 0 | 1 | 2 | 3 |
| *P* |  |  |  |  |

*E*(*X*)*=*0*×+*1*×+*2*×+*3*×=*(或*E*(*X*)*=*3*×=*)*.*