理科概统 第九节 抽样与频率分布

1、某单位有职工 750 人，其中青年职工 350 人，中年职工 250 人，老年职工 150 人，为了了解该单位职工的健康情况，用分层抽样的方法从中抽取样本．若样本中的青年职工为 7 人，则样本容量为________

解：$\dfrac{\boldsymbol{n}}{750}=\dfrac{7}{350},\boldsymbol{n}=15$，注：$a:b:c=A:B:C$，$\dfrac{a}{A}=\dfrac{b}{B}=\dfrac{c}{B}=\dfrac{n}{N}$

2、将参加夏令营的 600 名学生编号为：001，002，……600，采用系统抽样方法抽取一个容量为 50 的样本，且随机抽得的号码为 003．这 600 名学生分住在三个营区，从 001 到 300 在第Ⅰ营区，从 301 到 495 住在第Ⅱ营区，从 496 到 600 在第Ⅲ营区，三个营区被抽中的人数依次为________

解：$\boldsymbol{n}=50,\boldsymbol{d}=\dfrac{600}{50}=12$，

设在 001 到 300 中有 $\boldsymbol{n}_1$ 个，则 $3+12(\boldsymbol{n}_1-1)\le 300,1+4(\boldsymbol{n}_1-1)\le 100,\boldsymbol{n}_1\le\dfrac{99}{4}+1=25\dfrac{3}{4}$，

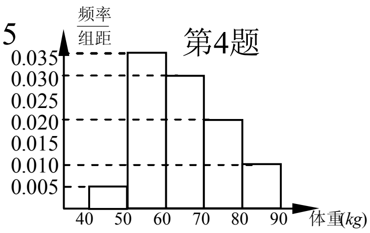
设在 001 到 300 中，首项 3，末项 291，于是 $\boldsymbol{n}_1=\dfrac{291-3}{12}+1=25,$

从 301 到 495 中，首项 304，末项 495，于是 $\boldsymbol{n}_2=\dfrac{495-303}{12}+1=17,\boldsymbol{n}_3=50-25-17=8$

3、 8 个班参加合唱比赛的得分茎叶图，
则这组数据的中位数和平均数分别是(A)
A、91.5 和 91.5 B、91.5 和 92 C、91 和 91.5D、92 和 92

茎	叶
8	**9 7**
9	**3 1 6 4 0 2**

解：中位数=(91+92)/2=91.5 $\overline{\boldsymbol{x}}=90+\dfrac{1}{8}(-1-3+3+1+6+4+0+2)=91.5$



第4题

4．一样本频率分布直方图(如图)
(1) 众数≈____(2) 中位数≈____ (3) 平均体重数≈____
解：(1) 众数≈55

(2) 中位数≈60+$60+10\times\dfrac{1}{3}=63.33$

(3) $\overline{\boldsymbol{x}}=45\times0.05+55\times0.35+65\times0.3+75\times0.2+85\times0.1$
$=64.5$

	频率
[40,50)	0.05
[50,60)	0.35
[60,70)	0.30
[70,80)	0.2
[80,90]	0.1

补 100 名体重在[60,70), [70,80), [80,90]三组内用分层抽样的方法取 12 人则选取的 12 人中体重在[60,80)内的男生要抽取的人数为____

解 1：

	频率	频数
[60,70)	0.3	30
[70,80)	0.2	20
[80,90]	0.1	10

$\dfrac{\boldsymbol{n}}{50}=\dfrac{12}{60},\boldsymbol{n}=10$

解 2：[60,70), [70,80), [80,90]的比例 $3:2:1$， [60, 80)有 5 份，于是 $12\times\dfrac{5}{6}=10$

5、为了了解某工厂开展群众体育活动的情况，拟采用分层抽样的方法从 A，B,C 三个区中抽取 7 个工厂进行调查，已知 A,B，C 区中分别有 18，27，18 个工厂（Ⅰ）求从 A,B,C 区中分别抽取的工厂个数；（Ⅱ）若从抽取的 7 个工厂中随机抽取 2 个进行调查结果的对比，用列举法计算这 2 个工厂中至少有 1 个来自 A 区的概率。

解：（1）18:27:18=2:3:2, $7\times\dfrac{2}{7}=2,7\times\dfrac{3}{7}=3,7\times\dfrac{2}{7}=2$

从 A,B,C 区中分别抽取的工厂个数分别是 2，3，2

（2）设 $\boldsymbol{M}=\{$这2个工厂中至少有1个来自A区$\}$

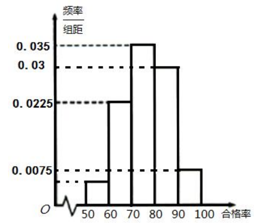
设从 A，B,C 抽得工厂分别记为 $\boldsymbol{a}_1,\boldsymbol{a}_2$； $\boldsymbol{b}_1,\boldsymbol{b}_2,\boldsymbol{b}_3$； $\boldsymbol{c}_1,\boldsymbol{c}_2$

试验的基本事件有：

$\{a_1,a_2\},\{a_1,b_1\},\{a_1,b_2\},\{a_1,b_3\},\{a_1,c_1\},\{a_1,c_2\},$
$\{a_2,b_1\},\{a_2,b_2\},\{a_2,b_3\},\{a_2,c_1\},\{a_2,c_2\},$
$\{b_1,b_2\},\{b_1,b_3\},\{b_1,c_1\},\{b_1,c_2\},$
$\{b_2,b_3\},\{b_2,c_1\},\{b_2,c_2\},$
$\{b_3,c_1\},\{b_3,c_2\},$
$\{c_1,c_2\}$,共21种

事件的基本事件有：

$\{a_1,a_2\},\{a_1,b_1\},\{a_1,b_2\},\{a_1,b_3\},\{a_1,c_1\},\{a_1,c_2\},$
$\{a_2,b_1\},\{a_2,b_2\},\{a_2,b_3\},\{a_2,c_1\},\{a_2,c_2\}$共11种, $P(A)=\dfrac{11}{21}$



6、某工厂共有工人 40 人，在一次产品大检查中每人的**产品合格率**（百分比）绘制成频率分布直方图，如图所示.

(Ⅰ) 求合格率在［50，60）内的工人人数；

(Ⅱ)为了了解工人在本次大检查中产品不合格的情况，从合格率在[50，70）内的工人中随机选取 3 人的合格率进行分析，用 X 表示所选工人合格率在[60，70）内的人数，求 X 的分布列和数学期望.

解：(Ⅰ) 产品合格率在［50，60）内的频率为：

1－（0.035＋0.03＋0.0225＋0.0075）×10＝0.05，　40×0.05＝2

所以产品合格率在［50，60）内的人数共有 2 人.

(Ⅱ) 产品合格率在[60，70）内的人数有 40×0.0225×10＝9，

所以产品合格率在[50，70）内的人数共有 11 人.

依题意，X 的可能取值是 1，2，3.

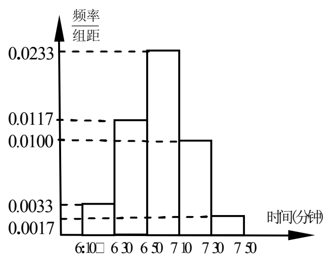
$P(X=1)=\dfrac{C_2^2C_9^1}{C_{11}^3}=\dfrac{3}{55}$；$P(X=2)=\dfrac{C_2^1C_9^2}{C_{11}^3}=\dfrac{24}{55}$；$P(X=3)=\mathrm{P(A)}=\dfrac{28}{55}$.

则 X 分布列为：

X	1	2	3
P	$\dfrac{3}{55}$	$\dfrac{24}{55}$	$\dfrac{28}{55}$

所以 $EX=1\times\dfrac{3}{55}+2\times\dfrac{24}{55}+3\times\dfrac{28}{55}=\dfrac{27}{11}$.

7、小明订了一份日报，暑假期间他收集了每天报纸送达的时间的数据，并绘制成频率分布直方图如图所示



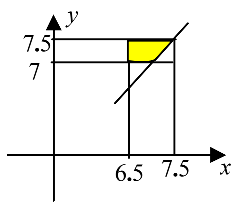
(Ⅰ) 请你根据图中的数据信息，写出众数 x_0=____

(Ⅱ)小明的父亲离家去上班的时间 y 在上午 7：00-7:30 之间，为此小明要求送报人每在 x_0 时间前后半小时内把报纸送达（每个时间点送达的可能性相等）

(1)求小明的父亲在去上班前能取到取到报纸(称为事件 A)的概率

(2) 求小明的父亲一周 5 天（假日除外）能取到取到报纸的的天数 X 的数学期望

解：(Ⅰ) x_0=7：00



(Ⅱ) (1)设报纸送达时刻为 x 点,

试验的基本事件为区域 $\boldsymbol{\Omega}=\{(x,y)\,|\,6.5\le x\le 7.5,7\le y\le 7.5\}$

事件 A 的基本事件为区域 $\{(x,y)\,|\,6.5\le x\le 7.5,7\le y\le 7.5,x\le y\}$

作出上述区域如图

$S_{\Omega}=1\times\frac{1}{2}=\frac{1}{2}$，$S_{A}=\frac{1}{2}-\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}=\frac{3}{8}$

$$P(A)=\frac{S_A}{S_{\Omega}}=\frac{\frac{3}{8}}{\frac{1}{2}}=\frac{3}{4}$$

8．某班有甲、乙两个研究性学习小组，甲组有 10 个学生，乙组有 5 个学生，他们的小组个人综合成绩如茎叶图所示。

（1）问甲、乙哪个小组的平均成绩更高？

（2）现采用分层抽样方法从甲、乙两组中共抽取 3 个学生的成绩，求至少一人成绩优秀的概率；　　（成绩在 80 分以上为优秀）

（3）抽样办法同（2），记ξ表示抽取的 3 个学生成绩优秀的数目，求ξ的分布列及数学期望。

甲组		乙组
3 7 8	6	6
2	7	6 9
2 3 5	8	9
1 4 5	9	5

解：(1) $\overline{x}_{甲}=\frac{1}{10}(63+67+68+72+82+83+85++91+92+95)=80$

$\overline{x}_{乙}=\frac{1}{10}(66+76+79+89+95)=81$

因 $\overline{x}_{甲}<\overline{x}_{乙}$，故乙组平均成绩更高

(2)用分层抽样方法，在甲组中抽取 2 个学生，乙中抽取 1 个学生。设至少一人成绩优秀的事件为 A,则成绩都不优秀的事件为 $\overline{A}$，

$P(\overline{A})=\frac{C_4^2C_3^1}{C_{10}^2C_5^1}=\frac{2}{25}$，$P(A)=1-P(\overline{A})=1-\frac{2}{25}=\frac{23}{25}$

(3) ξ 的可能取值为 0,1,2,3, ξ 服从超几何分布，

$P(\xi=0)=\frac{2}{25}$ ，$P(\xi=1)=\frac{C_6^1C_4^1C_3^1+C_4^1C_2^1}{C_{10}^2C_5^1}=\frac{28}{75}$ ，$P(\xi=2)=\frac{C_6^1C_4^1C_2^1+C_6^2C_3^1}{C_{10}^2C_5^1}=\frac{31}{75}$

$P(\xi=3)=\frac{C_6^2C_2^1}{C_{10}^2C_5^1}=\frac{10}{75}$,$\xi$ 的分布列

ξ	0	1	3	4
P	$\frac{2}{25}$	$\frac{28}{75}$	$\frac{31}{75}$	$\frac{10}{75}$

所以，$E\xi=1\times\frac{28}{75}+2\times\frac{31}{75}+3\times\frac{10}{75}=\frac{8}{5}=1.6$

9、2012 年 3 月 2 日，国家环保部发布了新修订的《环境空气质量标准》.其中规定：居民区中的 PM2.5 年平均浓度不得超过 35 微克/立方米，PM2.5 的 24 小时平均浓度不得超过 75 微克/立方米. 某城市环保部门随机抽取了一居民区去年 40 天的 PM2.5 的 24 小时平均浓度的监测数据，数据统计如下：

组别	PM2.5（微克/立方米）	频数（天）	频率
第一组	(0,15]	4	0.1
第二组	(15,30]	12	0.3
第三组	(30,45]	8	0.2
第四组	(45,60]	8	0.2
第三组	(60,75]	4	0.1
第四组	(75,90)	4	0.1

(Ⅰ)写出该样本的众数和中位数（不必写出计算过程）；

（Ⅱ）求该样本的平均数，并根据样本估计总体的思想，从 PM2.5 的年平均浓度考虑，判断该居民区的环境是否需要改进？说明理由；

（Ⅲ）将频率视为概率，对于去年的某 2 天，记这 2 天中该居民区 PM2.5 的 24 小时平

均浓度符合环境空气质量标准的天数为ξ，求ξ的分布列及数学期望E（ξ）.

解：(Ⅰ) 众数为 22.5 微克/立方米, 中位数为 37.5 微克/立方米.

（Ⅱ）去年该居民区 PM2.5 年平均浓度为

$7.5\times0.1+22.5\times0.3+37.5\times0.2+52.5\times0.2+67.5\times0.1+82.5\times0.1=40.5$（微克/立方米）.

因为$40.5>35$，所以去年该居民区 PM2.5 年平均浓度不符合环境空气质量标准，故该居民区的环境需要改进.

(Ⅲ)记事件 A 为“一天 PM2.5 的 24 小时平均浓度符合环境空气质量标准”，则$P(A)=\frac{9}{10}$.

随机变量ξ的可能取值为 0,1,2.且$\xi\approx B(2,\frac{9}{10})$.

所以$P(\xi=k)=C_2^k(\frac{9}{10})^k(1-\frac{9}{10})^{2-k}(k=0,1,2)$,

所以变量ξ的分布列为

ξ	0	1	2
p	$\frac{1}{100}$	$\frac{18}{100}$	$\frac{81}{100}$

$E\xi=0\times\frac{1}{100}+1\times\frac{18}{100}+2\times\frac{81}{100}=1.8$（天）,或$E\xi=nP=2\times\frac{9}{10}=1.8$（天）.