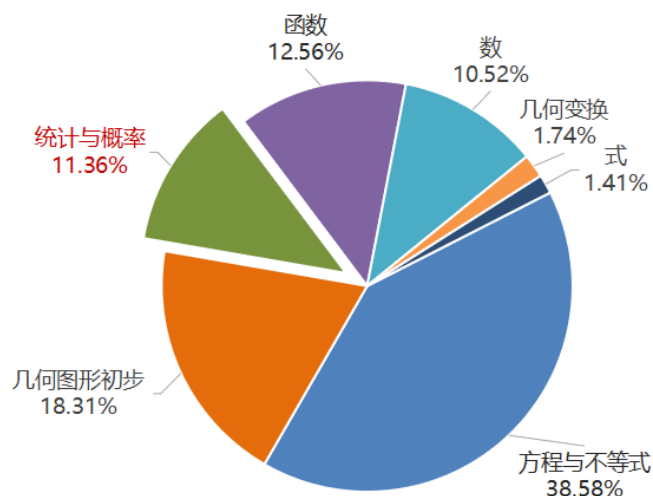


第1讲 数据的收集、整理与描述



一、数据的收集与整理

对数据的收集一般都采用问卷调查的方法。

收集数据的步骤：

- ①明确调查问题；
- ②明确调查对象；
- ③选择调查方法和调查形式；
- ④展开调查；
- ⑤统计并整理调查结果；
- ⑥分析结果并得出结论。

收集数据的方式：

直接收集和间接收集。

收集数据的方法：

统计调查是收集数据常用的方法，一般有全面调查和抽样调查两种。

除了问卷调查、访问调查等外，查阅文献资料和实验也是获得数据的有效方法。

1. 全面调查

全面调查：考察_____的调查叫做全面调查。

例如：2010年我国进行的第六次人口普查，就是一次全面调查。

全面调查的方法：问卷调查，访问调查，电话调查等。

全面调查的特点：全面调查收集的数据_____、_____，但一般花费的时间长，消耗的人力、物力大。

全面调查的适用范围：当调查的范围____，调查不具有____，数据要求_____时采用全面调查。

2. 抽样调查

抽样调查：从_____中抽取_____进行调查，然后根据调查数据推断全体对象情况的一种调查方法。

抽样调查的方法：民意调查法，实地调查法、媒体调查法等。

抽样调查的特点：抽样调查具有花费____、_____的特点，但抽取的样本是否具有代表性，直接关系到对总体估计的准确程度。

抽样调查的适用范围：当所受调查对象涉及面____、范围____、受条件限制或具有破坏性时采用抽样调查。

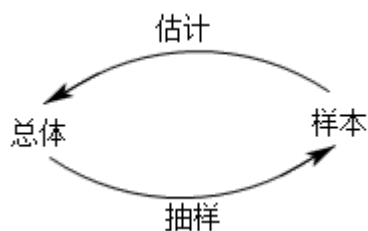
【注意】

抽样调查中的抽样必须具有_____，选取样本时应注意：

- ①选取的样本应具有代表性，不偏向总体中的某些个体；
- ②选取的样本容量要足够大；
- ③选取样本时，要避免遗漏总体中的某一群体。

🔗 总体、个体、样本与样本容量

总体	考察对象的全体叫做总体
个体	总体中每一个考察对象叫做个体
样本	从总体中所抽取的一部分个体叫做总体的一个样本
样本容量	样本中个体的数目叫做样本容量



【注意】 样本容量只是一个数字，不带单位。

🔗 用样本估计总体

样本估计总体是指通常不直接去研究总体，而是通过从总体中_____一个样本，根据样本的情况去估计总体的相应情况。

📊 例题1

- 当前，“低头族”已成为热门话题之一，小颖为了解路边行人步行边低头看手机的情况，她应采用的收集数据的方式是（ ）。
A. 对学校的同学发放问卷进行调查
B. 对在路边行走的学生随机发放问卷进行调查
C. 对在路边行走的行人随机发放问卷进行调查
D. 对在图书馆里看书的人发放问卷进行调查
- 下列调查活动中适合使用全面调查的是（ ）。
A. 某种品牌插座的使用寿命
B. 全国植树节中栽植树苗的成活率
C. 了解某班同学课外阅读经典情况
D. 调查“厉害了，我的国”大型记录电影在线收视率

练习1

3. 在数据收集处理的过程中，以下顺序正确的是（ ）.
- A. 收集数据→描述数据→整理数据→分析数据 B. 收集数据→分析数据→整理数据→描述数据
C. 收集数据→整理数据→分析数据→描述数据 D. 收集数据→整理数据→描述数据→分析数据
4. 国务院决定于**2020年11月1日**零时开展第七次全国人口普查，人口调查采用普查方式的理由是（ ）.
- A. 人口调查需要获得全面准确的信息 B. 人口调查的数目不太大
C. 人口调查具有破坏性 D. 受条件限制，无法进行抽样调查
5. 下列事件中最适合使用普查方式收集数据的是（ ）
- A. 了解全市每天丢弃的废旧电池数 B. 了解某班同学的身高情况
C. 了解**50**发炮弹的杀伤半径 D. 了解我省农民的年人均收入情况
6. 下列调查方式合适的是（ ）
- A. 为了了解市民对电影《南京》的感受，小华在某校随机采访了**8**名初三学生
B. 为了了解全校学生用于做数学作业的时间，小民同学在网上向**3**位好友做了调查
C. 为了了解全国青少年儿童的睡眠时间，统计人员采用了普查的方式
D. 为了了解“嫦娥一号”卫星零部件的状况，检测人员采用了普查的方式

例题2

7. 为了了解我市参加中考的**75000**名学生的视力情况，抽查了**1000**名学生的视力进行统计分析，下面四个判断中，正确的是（ ）.
- A. **75000**名学生是总体 B. **1000**名学生的视力是总体的一个样本
C. 每名学生是总体的一个个体 D. 上述调查是普查

练习2

8. 为了解学生每天自主学习时间，某校抽取了**50**名学生作为样本进行调查，在这个问题中，样本容量是 _____ .
9. 为了解某校初一年级**300**名学生的体重情况，从中抽取**50**名学生的体重进行统计分析. 在这个问题中，总体是指（ ）.
- A. **300**名学生 B. 被抽取的**50**名学生
C. **300**名学生的体重 D. 被抽取**50**名学生的体重

10. 中学生骑电动车上学给交通安全带来隐患，为了解某中学2500个学生家长对“中学生骑电动车上学”的态度，从中随机调查400个家长，结果有360个家长持反对态度，则下列说法正确的是（ ）.
- A. 调查方式是普查全面调查
B. 该校有360个家长对此持反对态度
C. 本次调查的样本是360个家长
D. 可以估计该校约有90%的家长对此持反对态度

|| 例题3

11. 某地区为估计该地区黄羊的只数，先捕捉20只黄羊给它们分别作上标志，然后放回，待有标志的黄羊完全混合于黄羊群后，第二次捕捉60只黄羊，发现其中2只有标志．从而估计该地区有黄羊 _____ 只．

|| 练习3

12. 生物工作者为了估计一片山林中雀鸟的数量，设计了如下方案：先捕捉100只雀鸟，给它们做上标记后放回山林；一段时间后，再从中随机捕捉500只，其中有标记的雀鸟有5只．请你帮助工作人员估计这片山林中雀鸟的数量约为 _____ 只．
13. 为了解植物园内某种花卉的生长情况，在一片约有3000株此类花卉的园地内，随机抽测了200株的高度作为样本，统计结果整理后列表如下：（每组数据可包括最低值，不包括最高值）

高度（cm）	40～45	45～50	50～55	55～60	60～65	65～70
频数	33	42	22	24	43	36

试估计该园地内此类花卉高度小于55厘米且不小于45厘米的约为 _____ 株．

二、数据描述

1. 数据整理

统计中经常用表格整理数据．表格通常由行和列组成，运用表格统计数据的优点是简单、清楚、突出数据的分布规律．

（1）整理数据时，可以用划记法记录数据，即通过“正”字的每一划（笔画）代表一个数据，统计数据的个数应等于所画“正”字的个数乘以5，再加上未画完的“正”字的笔画数．

（2）表格统计法的特点：

优点：统计表中的数据比较准确、详实，可以清楚的反应各个量的真实情况．

缺点：统计表得到的信息需要进行分析，表达不够直观．

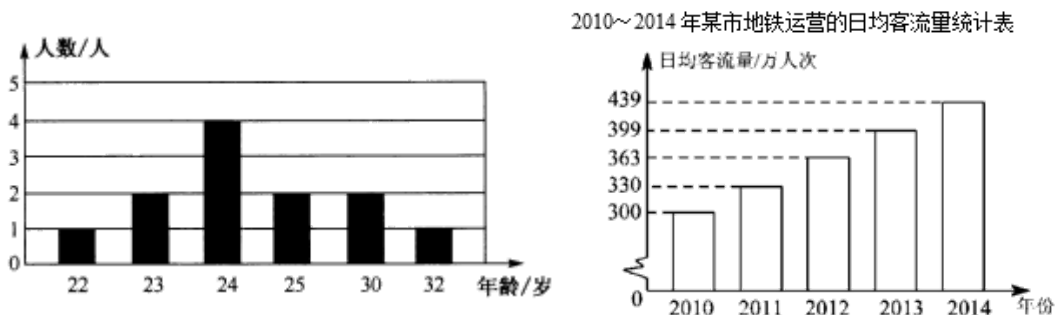
2. 数据描述

统计图能直观反映出事情的发展变化或总体与部分的关系，常见的统计图有条形图、扇形图、折线图。

3. 统计图

条形统计图

条形统计图：用线段长度表示数据，根据数据的多少画成长短不同的长方形直条，然后按照顺序把这些直条排列起来。



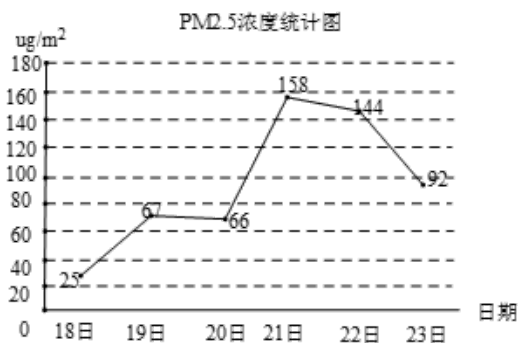
(1) 条形统计图的特点：条形图能清楚地表示出每个项目的**具体数目**，但是条形图不能表示不同时间的数目的变化和部分在总体中所占比值的大小。

(2) 制作条形图的一般步骤：

- ①根据具体情况，画出两条互相垂直的射线；
- ②在水平射线上，适当分配条形的宽度、位置及间隔；
- ③在与水平射线垂直的射线上，根据数据大小的具体情况，确定单位长度；
- ④按照数据的大小，画出长短不同的直条并标明数量。

折线统计图

折线统计图：用单位长度表示一定的数据，根据数量多少描出各点，然后把各点用线段依次连接起来。



(1) 折线统计图特点：能清楚地反映事物的**变化情况**，但不能表示各部分在总体中所占的比值。

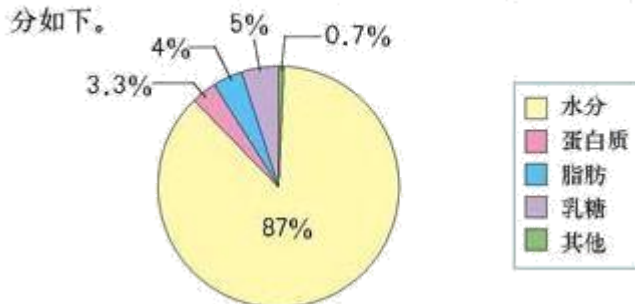
(2) 制作折线统计图的一般步骤：

- ①根据统计图的资料整理数据；
- ②画横轴、纵轴。且横轴纵轴都要有单位，按纸面的大小来确定一定单位长度表示一定的数量；
- ③根据数量的多少，在横轴、纵轴的恰当位置描出各点，然后把各点用线段顺次连接起来

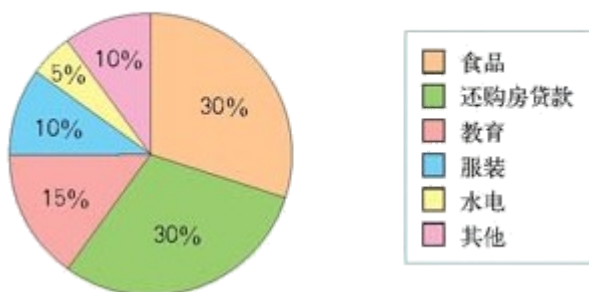
扇形统计图

扇形统计图：用整个圆表示总体，用圆内各个扇形的大小表示各部分数据。

牛奶里含有丰富的营养成分，每 100 g 牛奶所含营养成分如下。



陈东家每月各种支出计划



(1) 扇形统计图的特点：扇形图能清楚地表示出各部分在总体中所占的百分比，便于总体与部分间的直接比较，但不能清楚地表明每一个项目的具体数目。

(2) 制作扇形统计图的一般步骤

- ①先算出各部分数量占总数量的百分数；
- ②再算出各部分数量的扇形的圆心角的度数；
- ③取适当的半径画圆，在圆内画出各个扇形；
- ④在各扇形中标出数量名称和所占的百分数。

(3) 在扇形统计图中的计算：

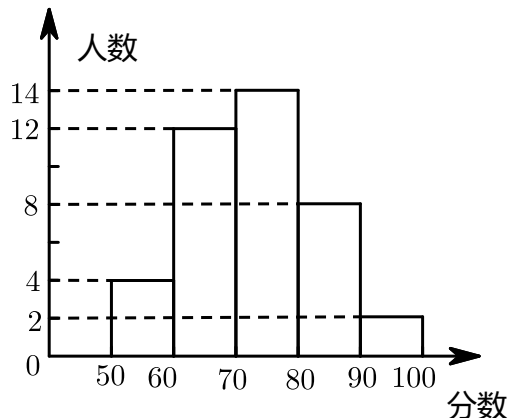
各部分所占的百分比之和为1

个体数=总体数×扇形所占的百分比

圆心角的度数=扇形所占的百分比× 360°

例题4

14. 某次考试中，某班级的数学成绩统计图如下．下列说法错误的是（ ）．



- A. 得分在70 ~ 80分之间的人数最多
 B. 该班的总人数为40
 C. 得分在90 ~ 100分之间的人数最少
 D. 及格 (≥ 60 分) 人数是26

15. 为了响应国家要求中小学生每天锻炼1小时的号召，某校开展了形式多样的“阳光体育运动”活动，小明对某班同学参加锻炼的情况进行了统计，并绘制了图1和图2．根据你对图1与图2的理解，回答下列问题：

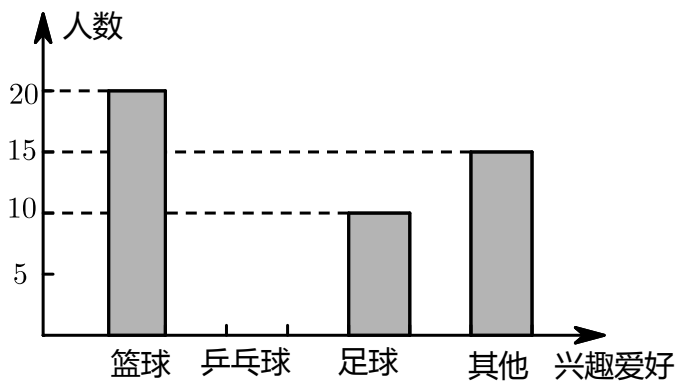


图 1

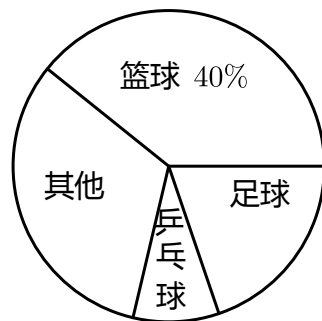


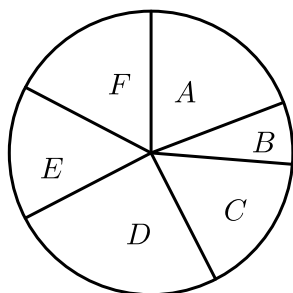
图 2

- (1) 小明调查的这个班级有多少名学生．
 (2) 请你将图1中“乒乓球”部分的图象补充完整．
 (3) 若这个学校共有1200名学生，估计参加乒乓球活动的学生有多少名．
 (4) 求出扇形统计图中表示“足球”的扇形的圆心角的度数．

练习4

16. 下列四个统计图中，用来表示不同品种的奶牛的平均产奶量最为合适的是（ ）。

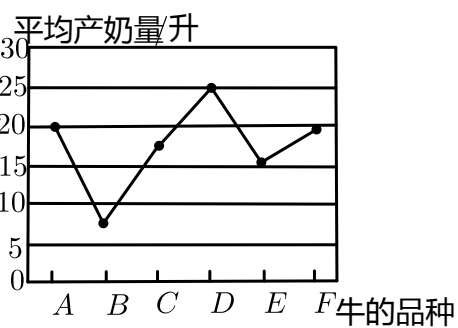
A.



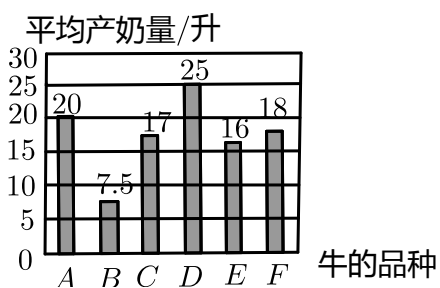
B.



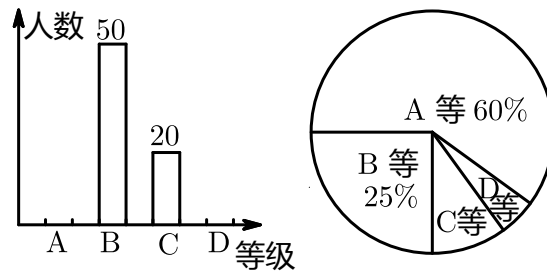
C.



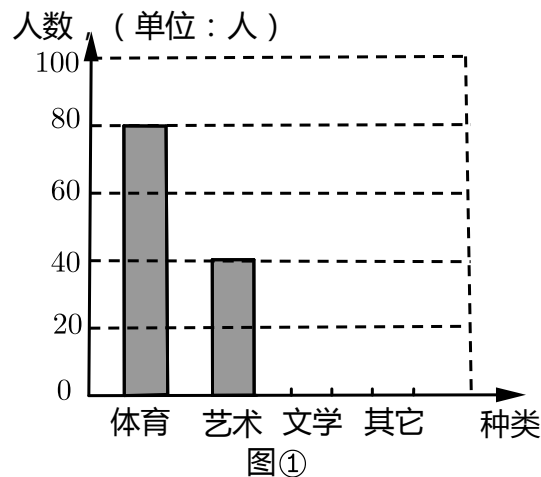
D.



17. 为积极响应南充市创建“全国卫生城市”的号召，某校1500名学生参加了卫生知识竞赛，成绩记为A、B、C、D四等．从中随机抽取了部分学生成绩进行统计，绘制成如图两幅不完整的统计图表，根据图表信息，以下说法不正确的是（ ）．



- A. 样本容量是200
B. D等所在扇形的圆心角为 15°
C. 样本中C等所占百分比是10%
D. 估计全校学生成绩为A等大约有900人
18. 为促进学生多样化发展，某校组织了课后服务活动，设置了体育类、艺术类、文学类及其它类社团（要求人人参与，每人只能选择一类）．为了解学生喜爱哪类社团活动，学校做了一次抽样调查，根据收集到的数据，绘制成如下两幅不完整的条形统计图和扇形统计图(如图①、图②)如下，请根据图中所给的信息，解答下列问题：



- (1) 此次共调查了多少人？
(2) 求艺术类在扇形统计图中所占的圆心角的度数．
(3) 将条形统计图补充完整．
(4) 如果该校有学生2200人，那么在全校学生中，喜爱文学类和其它类两个社团的学生共有多少人？

三、直方图

有关概念

频数	在记录数据时，某类数据出现的次数称为这类数据的频数，各对象的频数之和等于数据总和。
频率	频数与总次数的比值（或者百分比）称为这类数据的频率，即频率= $\frac{\text{频数}}{\text{数据总数}}$ 。 各对象的频率之和等于1。
组数	将一组数据进行适当的分组，把分成的组的个数叫做组数。
组距	把所有数据分成若干组，每个小组的两个端点之间的距离（组内数据的取值范围）称为组距。

频数分布表与频数分布直方图

频数分布表

将一组数据进行适当的分组，然后根据每一小组的频数的多少去研究数据的分布情况，对分析问题大有帮助，这样就产生了频数分布表。

频数记录一栏一般采用画“正”字的方法统计各数据段的频数；

频数一栏表示的是每个数据段内数据的个数，

频数的和即为这一组数据的总个数。

收集的63名学生身高如下：

158	158	160	168	159	159	151	158	159
168	158	154	158	154	169	158	158	158
159	167	170	153	160	160	159	159	160
149	163	163	162	172	161	153	156	162
162	163	157	162	162	161	157	157	164
155	156	165	166	156	154	166	164	165
156	157	153	165	159	157	155	164	156

对数据做出整理，频数分布表为：

频数分布表

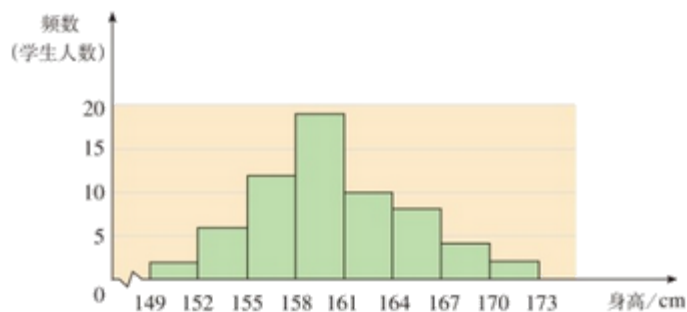
身高分组	划记	频数
$149 \leq x < 152$	┐	2
$152 \leq x < 155$	正┐	6
$155 \leq x < 158$	正正┐	12
$158 \leq x < 161$	正正正正┐	19
$161 \leq x < 164$	正正	10
$164 \leq x < 167$	正┐	8
$167 \leq x < 170$	正┐	4
$170 \leq x < 173$	┐	2

频数分布直方图

为了更直观形象的看出频数分布的情况，画出频数分布直方图。

频数分布直方图也是一种条形图。

频数分布直方图是以小长方形的面积来反映数据落在各个小组内的频数的大小。



特别的，等距分组时，各个小长方形的面积（频数）与高的比是常数（组距），因此，画等距分组的频数分布直方图时，为画图与看图方便，通常直接用小长方形的高表示频数。

频数分布直方图的制作步骤：

- ①计算出数据中的最大值和最小值的差；
- ②确定组距与组数，一般100以内的数据分成5 ~ 12组；
- ③决定分点，常使分点比数据多一位小数，并且把第一组的起点稍微减少一点；
- ④列频数分布表，用划记法对数据进行频数统计；
- ⑤画出频数分布直方图，构建一个坐标系，用横轴表示各段数据，纵轴表示频数与组距之比，这样画出的小长方形的面积就代表频数，各小组的频数之和等于样本中数据的总个数。

例题5

19. 将50个数据分成5组列出频数分布表，其中第一组的频数为6，第二组与第五组的频数之和为20，第三组的频率为0.2，则第四组的频率为（ ）。
- A. 0.28 B. 0.3 C. 0.4 D. 0.2

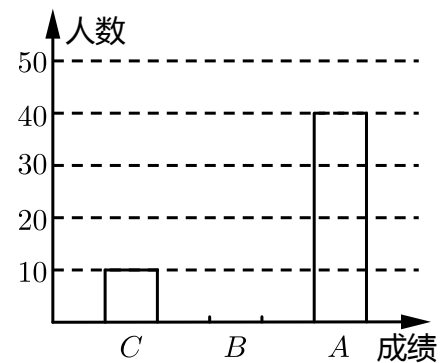
练习5

20. 某校准备组建七年级男生篮球队，有60名男生报名，体育老师对60名男生的身高进行了测量，获得60个数据，数学老师将这些数据分成5组绘制成频数分布直方图，已知从左至右的5个小长方形的高度比为1 : 3 : 5 : 4 : 2，则第五个小组的频数为（ ）。
- A. 12 B. 16 C. 20 D. 8
21. 2019年5月1日至10日我市空气质量指数（AQI）分别为77，52，46，57，58，78，75，34，47，43，将数据进行分组，落在53.5 ~ 59.5这一组的频数是 _____。
22. 一组数据中的最小值是33，最大值是103，若取组距为9，则组数为（ ）。
- A. 7 B. 8 C. 9 D. 7或8均可

例题6

23. 某校对九年级学生体育测试情况进行调查，从该校360名学生中抽取了部分学生的成绩（成绩分为A, B, C三个层次）进行分析，绘制了频数分布表与频数分布直方图，请根据图表信息解答下列问题：

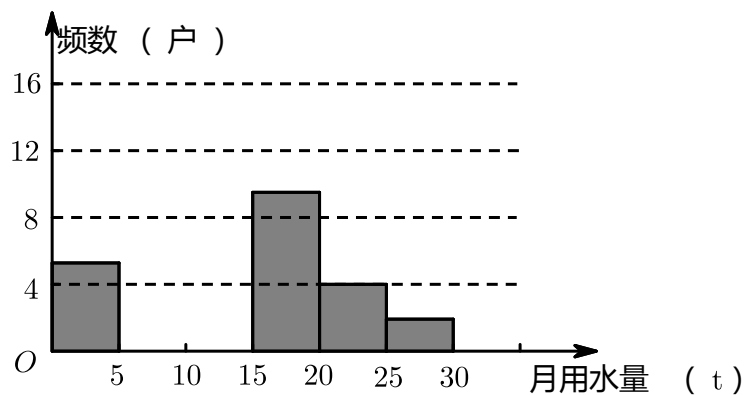
分组	频数	频率
C	10	0.1
B	a	0.5
A	40	b
合计	100	1



- (1) 表中的 $a = \underline{\hspace{2cm}}$, $b = \underline{\hspace{2cm}}$.
- (2) 补全频数分布直方图 .
- (3) 如果成绩为A等级的同学属于优秀，请你估计该校九年级的有多少人达到优秀水平？

24. 某区在实施居民用水管理前，随机调查了部分家庭（单位：户）去年的月均用水量（单位：t），并将调查数据进行整理，绘制如下不完整的统计图表：

月均用水量	频数	频率
$0 \leq x < 5$	6	12%
$5 \leq x < 10$	12	24%
$10 \leq x < 15$	_____	32%
$15 \leq x < 20$	10	20%
$20 \leq x < 25$	4	_____
$25 \leq x < 30$	2	4%
合计	_____	100%

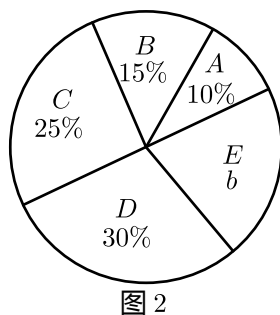
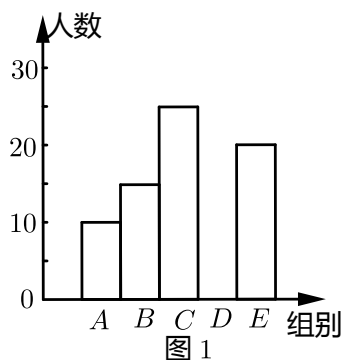


请解答以下问题：

- (1) 把上面的频数分布表和频数分布直方图补充完整．
- (2) 若该小区有2000户家庭，根据此次随机抽查的数据估计，该小区月均用水量不低于20t的家庭有多少户？
- (3) 为了鼓励节约用水，要确定一个月均用水量的标准，超出该标准的部分按1.5倍价格收费，若要使68%的家庭水费支出不受影响，那么，你觉得家庭月均用水量应定为多少？

练习6

25. 某中学举行“汉字听写”比赛，每位学生听写汉字39个，比赛结束后，随机抽查部分学生的听写结果，以下是根据抽查结果绘制的统计图的一部分，根据信息解决下列问题：

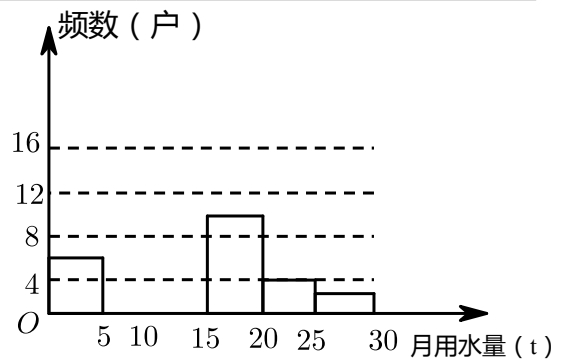


组别	正确字数 x	人数
A	$0 \leq x < 8$	10
B	$8 \leq x < 16$	15
C	$16 \leq x < 24$	25
D	$24 \leq x < 32$	a
E	$32 \leq x < 40$	20

- (1) 在统计表中， $a = \underline{\hspace{2cm}}$ ， $b = \underline{\hspace{2cm}}$ 。
- (2) 补全条形统计图，计算扇形统计图中“D组”所对应的圆心角的度数。
- (3) 若该校共有500名学生，如果听写正确的字数少于16个定为不合格，请你估计这所中学本次比赛听写不合格的学生人数。

26. 九（1）班同学为了解2011年某小区家庭月均用水情况，随机调查了该小区部分家庭，并将调查数据进行如下整理．

月均用水量 x （t）	频数（户）	频率
$0 < x \leq 5$	6	0.12
$5 < x \leq 10$		0.24
$10 < x \leq 15$	16	0.32
$15 < x \leq 20$	10	0.20
$20 < x \leq 25$	4	
$25 < x \leq 30$	2	0.04



请解答以下问题：

- （1）把上面的频数分布表和频数分布直方图补充完整．
- （2）求该小区用水量不超过15t的家庭占被调查家庭总数的百分比．
- （3）若该小区有1000户家庭，根据调查数据估计，该小区月均用水量超过20t家庭大约有多少户？