

暑假专题——数据的收集与处理

知识要点：

1. 调查方式

(1) 普查 (具有获取数据相对准确的优点)

(2) 抽样调查 (具有节省时间, 人力、物力和财力的特点)

2. 抽样调查要注意样本的代表性和广泛性

3. 频数：考察中每个对象出现的次数。

4. 频率 = $\frac{\text{频数}}{\text{总数}}$

5. 频数分布直方图 (频率分布直方图) 能非常直观的表示数据的分布情况。

6. 极差：最大值和最小值的差。

7. 方差 (标准差) 反映数据的波动情况 (稳定性, 离散程度等)

$$S^2 = \frac{1}{n} [(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2]$$

【典型例题】

例 1. 填空题：

(1) 为了了解你们班同学所穿鞋子的尺码, 可对全班同学采用哪种调查方式?

_____。

答案：普查

(2) 为了了解你所在地区老年人的健康状况, 你采用哪种调查方式?

_____。

答案：抽样调查

(3) 了解我国初中学生的视力情况, 你认为采用什么调查方式更合适?

_____。

答案：抽样调查

(4) 一所中学的数学研究小组的 10 位教师的年龄分别是：

55, 40, 40, 48, 36, 36, 37, 25, 33, 40。这 10 名教师年龄的平均数是_____, 众数是_____, 中位数是_____, 最大值是_____, 最小值是_____, 极差是_____。

答案：39, 40, 38.5, 55, 25, 30

分析： $\bar{x} = \frac{1}{10}(55 + 40 + 40 + 48 + 36 + 36 + 37 + 25 + 33 + 40) = 39$ (岁)

排序为：25, 33, 36, 36, 37, 40, 40, 40, 48, 55

$\frac{37 + 40}{2} = 38.5$ (岁)

$$55 - 25 = 30 \text{ (岁)}$$

(5) 甲、乙两名射击运动员各连续射靶 8 次，命中的环数如下：

甲：8，9，10，9，8，7，9，10；

乙：9，8，8，9，10，9，8，9

如果你是教练，你认为谁的成绩更稳定？_____。

答案：乙

分析：甲的极差 $10 - 7 = 3$

乙的极差 $10 - 8 = 2$

例 2. 选择题：

(1) 甲、乙两支仪仗队队员的身高如下：

甲队：178，177，179，179，178，178，177，178，179；

乙队：178，179，176，178，180，180，178，176，178。

要考察哪支仪仗队更为整齐，我们可以使用两支仪仗队队员身高的（ ）

A. 平均数 B. 众数 C. 中位数 D. 方差

答案：D

(2) 下面是今年与前年在大致相同条件下饲养的 10 头猪的体长数据（单位：厘米）：

前年：112，110，110，117，113，122，125，124，119，127；

今年：111，122，115，123，114，115，118，114，116，115。

则饲养的 10 头猪的体长比较一致的是（ ）

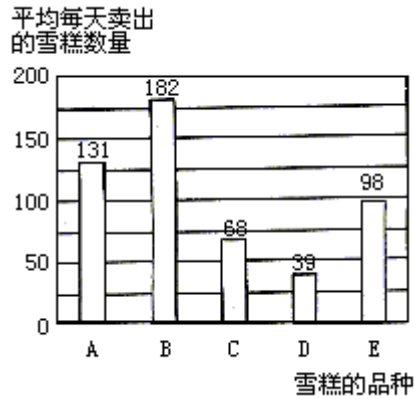
A. 前年 B. 今年 C. 一样 D. 无法判断

答案：B

分析：前年的极差： $127 - 110 = 17$

今年的极差： $123 - 111 = 12$

(3) 小颖统计了最近一个星期王奶奶平均每天能卖出的 A、B、C、D、E 五个牌子雪糕的数量，并绘制出频数分布直方图，则 A 种雪糕的频数与频率分别为（ ）



- A. $131, \frac{131}{200}$ B. $131, \frac{131}{150}$
 C. $131, \frac{131}{518}$ D. $131, 131$

答案：C

分析： $131 + 182 + 68 + 39 + 98 = 518$

例3. 解答题：

(1) 已知 x_1, x_2, \dots, x_n 的平均数为 \bar{x} ，方差为 S^2 ，你能分析下列几组数据的平均数与方差吗？并总结你的发现！

- ① $x_1 + 2, x_2 + 2, \dots, x_n + 2$
 ② $x_1 - 1, x_2 - 1, \dots, x_n - 1$
 ③ $3x_1, 3x_2, \dots, 3x_n$
 ④ $\frac{1}{2}x_1, \frac{1}{2}x_2, \dots, \frac{1}{2}x_n$

解析： $\bar{x}_{\text{①}} = \frac{1}{n}(x_1 + 2 + x_2 + 2 + \dots + x_n + 2)$

$$= \frac{1}{n}(x_1 + x_2 + \dots + x_n + 2n)$$

$$= \frac{1}{n}(x_1 + x_2 + \dots + x_n) + 2$$

$$= \bar{x} + 2$$

$$S_{\text{①}}^2 = \frac{1}{n} \left\{ [(x_1 + 2) - (\bar{x} + 2)]^2 + [(x_2 + 2) - (\bar{x} + 2)]^2 + \dots + [(x_n + 2) - (\bar{x} + 2)]^2 \right\}$$

$$= \frac{1}{n} [(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2]$$

$$= S^2$$

∴ 数据①的平均数是 $\bar{x} + 2$ ，方差仍是 S^2

② 同理可验证数据②的平均数是 $\bar{x} - 1$ ，方差仍是 S^2

$$\text{③ } \bar{x}_{\text{③}} = \frac{1}{n}(3x_1 + 3x_2 + \dots + 3x_n)$$

$$\begin{aligned}
&= 3 \times \frac{1}{n} (x_1 + x_2 + \dots + x_n) \\
&= 3\bar{x} \\
S_{\text{㉓}}^2 &= \frac{1}{n} \left[(3x_1 - 3\bar{x})^2 + (3x_2 - 3\bar{x})^2 + \dots + (3x_n - 3\bar{x})^2 \right] \\
&= \frac{1}{n} \left[9(x_1 - \bar{x})^2 + 9(x_2 - \bar{x})^2 + \dots + 9(x_n - \bar{x})^2 \right] \\
&= 9 \times \frac{1}{n} \left[(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2 \right] \\
&= 9S^2
\end{aligned}$$

∴ 数据③的平均数是 $3\bar{x}$ ，方差是 $9S^2$

④ 同理可验证数据④的平均数是 $\frac{1}{2}\bar{x}$ ，方差是 $\frac{1}{4}S^2$

由上可得：平均数随数据的变化而变化，与每个数据的变化完全相同。

方差不一定随数据的变化而变化，当数据进行加减变化时方差不变，当数据进行乘除变化时，方差成平方倍变化。

(2) 为了考察甲、乙两种小麦的长势，分别从中抽取了 10 株苗，测得苗高如下（单位：厘米）：

甲：12，13，14，15，10，16，13，11，15，11

乙：11，16，17，14，13，19，6，8，10，16

① 分别计算两种小麦的平均苗高；

② 哪种小麦的 10 株苗高比较整齐？

解析：① $\bar{x}_{\text{甲}} = 13$ ， $\bar{x}_{\text{乙}} = 13$

$$S_{\text{甲}}^2 = \frac{1}{10} \left[(-1)^2 + 0 + 1^2 + 2^2 + (-3)^2 + 3^2 + 0^2 + (-2)^2 + 2^2 + (-2)^2 \right]$$

$$= 3.6$$

$$S_{\text{乙}}^2 = \frac{1}{10} \left[(-2)^2 + 3^2 + 4^2 + 1^2 + 0^2 + 6^2 + (-7)^2 + (-5)^2 + (-3)^2 + 3^2 \right]$$

$$= 15.8$$

$$\because S_{\text{甲}}^2 < S_{\text{乙}}^2$$

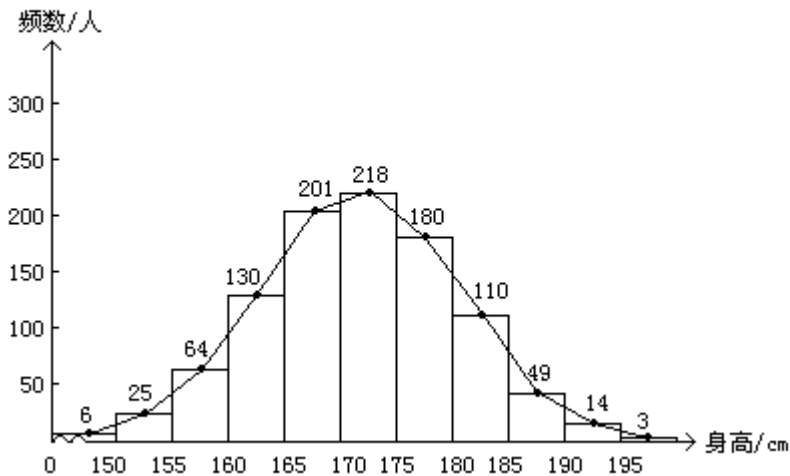
∴ 甲整齐

(3) 某部门统计了上海 1000 名高三男生的身高，得到如下数据，将这些数据绘制成频数分布直方图和折线图。

身高/cm	频数
<150	6
150~155	25
155~160	64
160~165	130
165~170	201
170~175	218
175~180	180
180~185	110
185~190	49
190~195	14
>195	3

解析：

上海 1000 名高三男生身高频数分布直方图及折线图



【模拟试题】

时间：45 分钟，满分 100 分

一. 填空题（每空 3 分，共 36 分）

1. 为了了解某校小学生的体能情况，对该校一个年级的部分学生进行一分钟跳绳次数测试，这个问题中，总体是_____，个体是_____，样本是_____。

2. 某班 50 名学生在一次数学考试中，分数在 90~100 分的频率是 0.16，则该班在这个分数段的人数是_____。

3. 一组数 1, 2, 3, 4, 5 的方差是_____。

4. 在一次科技知识竞赛中,一组学生成绩统计如下表:

分数	50	60	70	80	90	100
人数	2	5	10	13	14	6

则这组学生成绩的中位数是_____。

5. 极差是刻画数据_____的一个统计量。

6. 某班第二组男生参加体育测试，引体向上成绩（单位：个）如下：

6, 9, 11, 13, 11, 7, 10, 8, 12, 则这组成绩的众数是_____，中位数是_____。

7. 某校九年级一班在体育考试中，全班所有学生得分的情况如下表所示：

分数段	18分以下	18~20分	21~23分	24~26分	27~29分	30分
人数	2	3	12	20	18	10

那么该班共有_____人，得分在27~30分之间人数的频率是_____，从上表中，你能获取的信息是_____（写出一个即可）。

二. 选择题（每题2分，共10分）

8. 今年我市共有8万名初中毕业生参加升学考试，为了了解这8万名考生的数学成绩，从中抽取了2000名考生的数学成绩进行统计分析，以下说法中正确的是（ ）

- A. 8万名考生是总体
- B. 每名考生的数学成绩是个体
- C. 2000名考生是总体的一个样本
- D. 以上都不对

9. 下列说法正确的是（ ）

- A. 数据3, 4, 4, 5的众数是2
- B. 数据2, 3, 5, 4, 6的中位数是5
- C. 数据5, 3, 7, 8, 2的平均数是5
- D. 频率分布直方图中，各小长方形的面积和等于数据总数

10. 甲、乙两人在同样的条件下练习射击，每人打5发子弹，命中环数如下：

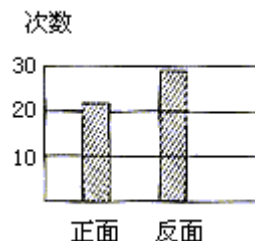
甲：6, 8, 9, 9, 8；

乙：10, 7, 7, 7, 9。

则两人射击成绩稳定程度关系是（ ）

- A. 甲比乙稳定
- B. 乙比甲稳定
- C. 甲、乙稳定程度相同
- D. 无法比较

11. 某同学抛掷硬币50次，得到的结果制作统计图如图所示，则这50次抛硬币中，正面朝上的频率是（ ）



- A. 0.44
- B. 0.56
- C. 0.22
- D. 0.28

12. 甲、乙两班举行电脑汉字输入速度的比赛，参赛学生每分钟输入汉字的个数经统计计算后填入下表：

班级	参加人数	中位数	方差	平均数
甲	55	149	19	135
乙	55	151	11	135

某同学根据此表分析得出如下结论：

- (1) 甲、乙两班学生成绩的平均水平相同；
- (2) 乙班优秀人数多于甲班优秀人数（每分钟输入汉字 ≥ 150 个为优秀）；
- (3) 甲班成绩的波动情况比乙班成绩的波动大。

上述结论中正确的是（ ）

- A. (1) (2) (3)
- B. (1) (2)
- C. (1) (3)
- D. (2) (3)

三. 解答题（第 13，14 题每题 8 分，第 15，16 题每题 9 分，第 17，18 题每题 10 分，共 54 分）

13. 下列调查各属于哪种调查方式？把答案写在后面的括号内。

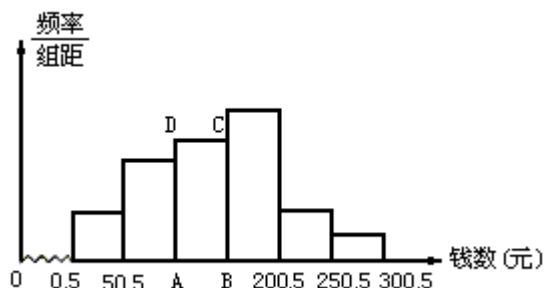
- (1) 为了了解八年级学生的视力情况，在该年级中抽取了 100 名学生进行视力检查测试；（ ）
- (2) 为了调查学校的男女生比例，调查统计了各班男、女生人数；（ ）
- (3) 为了考察同一型号的一批炮弹的杀伤半径，从中任意抽取 210 枚进行调查分析。（ ）

14. 下列抽样调查中，结果能否较准确地反映总体的情况，为什么？

- (1) 某商场为了了解 10 月份的营业情况，从 10 月 2 日开始连续调查了 5 天的营业情况；
- (2) 某公司为了了解自己产品的普及率，在市区某火车站对 100 名流动人员进行调查分析。

15. 未成年人思想道德建设越来越受到社会的关注。某青少年研究所随机调查了该市某校 100 名学生寒假中所花零花钱的钱数（钱数取整元数），以便引导学生树立正确的消费观。根据调查数据制成了频率分布表和频率分布直方图（如图）。

分组	频数	频率
0.5 ~ 50.5		0.1
50.5 ~ _____	20	0.2
100.5 ~ 150.5		
_____ ~ 200.5	30	0.3
200.5 ~ 250.5	10	0.1
250.5 ~ 300.5	5	0.05
合 计	100	



- (1) 补全频率分布表；
- (2) 在频率分布直方图中，长方形 ABCD 的面积是_____，这次调查的样本容量是_____；
- (3) 研究所认为，应对消费 150 元以上的学生提出勤俭节约的建议。试估计应对该校 1000 名学生中约多少名学生提出这项建议？

16. 为了预测 4 月份电费开支，小红在月初连续 8 天（每天早上）抄下了电表上的读数（如下表）。若每度电 0.5 元，估计小红家 4 月份（按 30 天计）的电费是多少元？

日期	1	2	3	4	5	6	7	8
电表读数	21	24	28	33	39	42	46	49

17. 一台机床生产直径是 40 mm 的零件，为检验机床质量，从产品中抽出 10 件进行测量，每件产品与标准零件的偏差如下（单位：毫米）：

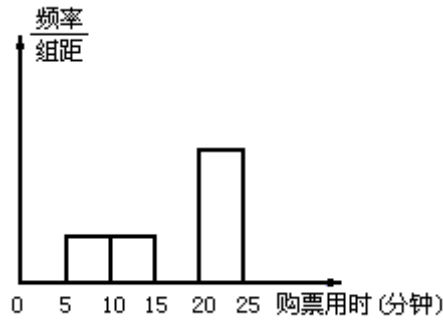
0, - 0.2, 0.1, 0.2, - 0.1, 0, 0.2, - 0.2, 0.2, - 0.2

- (1) 计算这 10 件零件直径长度的方差。
- (2) 若 10 件零件直径长度数据的标准差不能超过 0.1mm，否则该机床不合格，试判断该机床是否合格。

18. 某车站春运期间为改进服务，随机抽样调查了 100 名游客从开始在购票窗口排队到购到车票所用的时间 t （以下简称购票用时，单位为分钟）。下面是这次调查统计分析得到的频率分布表和频率分布直方图。解答下列问题：

- (1) 这次抽样的样本容量是多少？
- (2) 在表中填写出缺失的数据并补全频率分布直方图；
- (3) 旅客购票用时的平均数可能落在哪一小组？
- (4) 若每增加一个购票窗口可以使平均购票用时降低 5 分钟，要使平均购票用时不超过 10 分钟，那么请你估计最少需增加几个窗口？

分组		频数	频率
一组	$0 \leq t < 5$	0	0
二组	$5 \leq t < 10$	10	0.10
三组	$10 \leq t < 15$	10	
四组	$15 \leq t < 20$		0.50
五组	$20 \leq t < 25$	30	0.30
合计		100	1.00



【试题答案】

一. 填空题。

1. 该校全体小学生一分钟跳绳次数；该校一个学生一分钟跳绳次数；该校一个年级的部分学生一分钟跳绳次数

2. 8

3. 2

4. 80

5. 离散程度

6. 11；10

7. 65； $\frac{28}{65}$ ；略

二. 选择题。

8. B

9. C

10. A

11. A

12. A

三. 解答题。

13. (1) 抽样调查 (2) 普查 (3) 抽样调查

14. (1) 不能，因为10月2~6号仍处在国庆放假期间，得到的数据不具代表性

(2) 不能，流动人员远远少于“固定”人员，得到的数据不具代表性

15. (1) 略 (2) 0.25；100 (3) 450

16. 60元

17. (1) 0.026

(2) $\sqrt{0.026} = 0.161 > 0.1$ ，故该机床不合格

18. (1) 样本容量是100

(2) 50 0.10 略

(3) 第4小组

(4) 设需要增加 x 个窗口，则 $20 - 5x \leq 10$

$$\therefore x \geq 2$$

\therefore 至少增加2个窗口