

高中数学 新课程

学习指导

3

必修

北师大版

与北师大版普通高中课程标准
实验教科书配套

河南省基础教育教学研究室 编

大象出版社

第一章

统计

从普查到抽样

抽样方法

统计图表

数据的数字特征

用样本估计总体

统计活动：结婚年龄的变化

相关性

最小二乘估计

高考同步链接

本章综合测试

第二章

算法初步

算法的基本思想

算法框图的基本结构及设计

几种基本语句

高考同步链接

本章综合测试

第三章

概率

随机事件的概率

古典概型

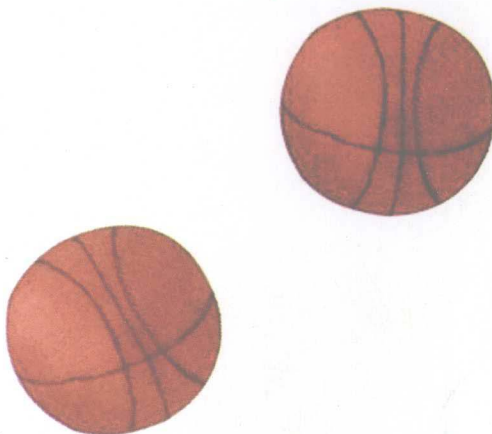
模拟方法——概率的应用

高考同步链接

本章综合测试

阶段评价测试

习题详解点拨



高中数学 新课程

学习指导

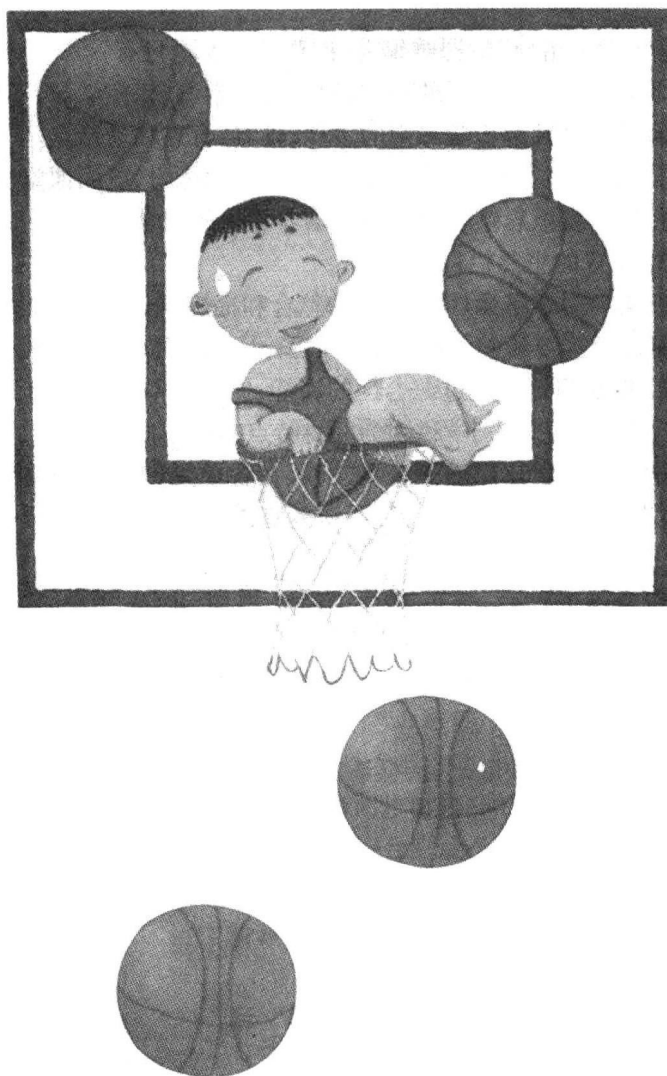
3
必修

北师大版

与北师大版普通高中课程标
准实验教科书配套

河南省基础教育教学研究室 编

大象出版社



编写说明

从2008年秋季开始,河南省全面进入普通高中新课程改革。为了新课程实验在河南省的顺利实施,为了更好地服务于高中教学,河南省基础教育教学研究室和大象出版社在深入调研、充分论证的基础上,对传统品牌教辅“高中学习指导”进行重新定位,重新组织开发了“高中新课程学习指导”丛书。这套丛书将于2008年秋季开始在全省推广使用。

遵循推进课改、利于教学的原则,树立以学生发展为本的教育理念,由省内外教研专家和高中一线名师倾力打造的“高中新课程学习指导”具有以下特色:**基础性**——体现基础教育教学改革的精神,为学生的终身发展奠定基础;**选择性**——提供个性化、多样化的学习资源,为促进学生全面而有个性的发展创造广阔的自主学习空间;**适用性**——为河南省高中学生量身定做;**创新性**——站在课改前沿,依据新课程理念,培养学生创新精神。

“高中新课程学习指导”按课时编写,设置的主要栏目有:

自主探究学习 学生是学习的主体,通过自主学习、探究学习,不断提高学习能力。

名师要点解析 名师解析学习中的重点、难点、盲点和易错点。

课堂基础自测 课堂是学习的主战场,通过基础练习,巩固课堂所学知识。

综合能力拓展 发散思维、凝聚要点,培养学生的综合能力。

每单元(章)设置的主要栏目有:

知识要点归纳 对本单元(章)知识的整合和提炼,帮助学生巩固学习要点。

高考同步链接 为学生打开高考的一面窗,让他们体验高考、感悟高考。

单元(本章)综合测试 通过综合性的训练,促进对本单元(章)知识的全面掌握。

(上述各栏目的设置,个别学科因为教材特点略有不同。)

为方便同学们对所学知识进行自我检验,在各单元(章)讲解和训练之后还设置了“**阶段评价测试**”;在全书最后附有“**习题详解点拨**”,对所有习题提供详尽的答案和解题思路。

本套丛书包括思想政治、语文、数学、英语、物理、化学、历史、地理、生物九个学科,涉及在我省实验的各种教材版本。

参加本册编写的作者是赵小强、芦国贤、李忠义同志,最后由骆传枢、张海营、刘志凤同志统稿。

对使用中发现的错谬缺漏之处,恳请广大师生批评指正。

河南省基础教育教学研究室

2008年12月

目 录

第一章 统计/1

- 1.1 从普查到抽样/1
- 1.2 抽样方法/1
- 1.3 统计图表/4
- 1.4 数据的数字特征/7
- 1.5 用样本估计总体/10
- 1.6 统计活动:结婚年龄的变化/10
- 1.7 相关性/13
- 1.8 最小二乘估计/15
- 高考同步链接/18
- 本章综合测试/20

第二章 算法初步/24

- 2.1 算法的基本思想/24
- 2.2 算法框图的基本结构及设计/28
- 2.3 几种基本语句/38
- 高考同步链接/44
- 本章综合测试/46

第三章 概率/49

- 3.1 随机事件的概率/49
- 3.2 古典概型/52
- 3.3 模拟方法——概率的应用/55
- 高考同步链接/60
- 本章综合测试/61

阶段评价测试/64

附 习题详解点拨

第一章 统计

1.1 从普查到抽样

1.2 抽样方法

自主探究学习

1. 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($n < N$),如果每次抽取时总体内的每个个体被抽到的概率_____,就把这种抽样方法叫做_____. 这种抽样方法最常用的有_____和_____.

2. 将总体分成均衡的几个部分,然后按照预先定出的规则,从每一部分抽取一个个体,得到所需要的样本,这种抽样方法叫做_____.

3. 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_____,从各层独立地抽取一定数量的个体,将各层抽取的个体合在一起作为_____,这种抽样方法是一种_____.

名师要点解析

【要点导学】

1. 简单随机抽样

简单随机抽样是一种最简单、最基本的抽样方法. 抽样中选取个体的方法有两种:放回和不放回. 我们在抽样调查中用的是不放回抽取.

(1)简单随机抽样的特点:①被抽取样本的总体个数有限. ②从总体中逐个进行抽取,使抽样便于在实践中操作. ③它是不放回抽取,这使其具有广泛的应用性. ④每一次抽样时,每个个体等可能地被抽到,保证了抽样方法的公平性.

(2)实施抽样的方法:①抽签法:方法简单,易于理解. ②随机数表法:要理解好随机数表,即表中每个位置上等可能出现 $0, 1, 2, \dots, 9$ 这十个数字的数表. 随机数表中各个位置上出现各个数字的等可能性,决定了利用随机数表进行抽样时抽取到总体

中各个个体序号的等可能性.

注:①选定开始数字,要保证所选数字的随机性. ②确定读数方向获取样本号码时,读数方向可向左、向右、向上、向下,样本号码不能重复,否则舍去.

2. 系统抽样

(1)系统抽样适用于总体中的个体数较多的情况.

(2)系统抽样与简单随机抽样之间存在着密切联系,即在将总体中的个体均分后的每一段中进行抽样时,采用的是简单随机抽样.

(3)系统抽样的操作步骤:第一步,利用随机的方式将总体中的个体编号. 第二步,将总体的编号分段,要确定分段间隔 k ,当 $\frac{N}{n}$ (N 为总体中的个体数, n 为样本容量)是整数时, $k = \frac{N}{n}$; 当 $\frac{N}{n}$ 不是整数时,通过从总体中剔除一些个体使剩下的个体数 N' 能被 n 整除,这时 $k = \frac{N'}{n}$. 第三步,在第一段用简单随机抽样确定起始个体编号 l ,再按事先确定的规则抽取样本. 通常是将 l 加上间隔 k 得到第2个编号 $(l+k)$,将 $(l+k)$ 加上 k ,得到第3个编号 $(l+2k)$,这样继续下去,直到获取整个样本.

3. 分层抽样

分层抽样适用于总体由差异明显的几部分组成的情况,注意与简单随机抽样的区别.

对这种抽样须说明的是:

(1)分层抽样要将相近一类归入一层,不同类型归于不同层,所以又称类型抽样.

(2)分层抽样所取得的样本更具代表性,更能反映总体情况.

(3)分层抽样时,各部分抽取的个体数与这一部分个体数的比等于样本容量与总体个体数的比.

(4)分层抽样的过程可分为四步:第一步,确定

样本容量与总体个体数的比;第二步,计算出各层需抽取的个体数;第三步,采用简单随机抽样或系统抽样在各层中抽取个体;第四步,将各层中抽取的个体合在一起,就是所要抽取的样本.

【经典例题】

【例1】一单位有一线职工 160 人,为了了解职工的收入情况,从中抽取一个容量为 20 的样本,按下述方法抽取:

将 160 人从 1~160 编上号,再将用白纸做成的 1~160 号的 160 个签放入箱内拌匀,然后从中抽 20 个签,与签号相同的 20 个人被选出.问:

- (1)上述方法中,总体、个体、样本分别是什么?
- (2)上述方法采用了何种抽取样本的方法?
- (3)你还可以采用什么方法?试阐述其过程.

【分析】由于总体中的个体数不算太多,因此可以采取简单随机抽样的方法,可以采用抽签法和随机数表法.

【解】(1)上述方法中,其总体是该单位 160 名一线职工的收入,其个体是该单位各个一线职工的收入,其样本为所抽取的 20 名一线职工的收入.

(2)采用了简单随机抽样方法中的抽签法.

(3)还可采用随机数表法.首先对职工进行编号,从 000 到 159.然后在随机数表中任选一个数,从选定的数开始向右读,得到一个三位数,若小于 160,将它取出;继续向右读……依次下去,直到样本的 20 个号码全部取出即可.

【点拨】简单随机抽样有操作简便易行的优点,在总体个体数不多的情况下是行之有效的.

变式训练

请举出几个用抽签法或随机数表法抽取样本的实际例子.你认为用简单随机抽样方法有什么优点和缺点?

【答案】略

【例2】人们打桥牌时,将洗好的扑克牌(52 张,即剔除大王、小王)随机确定一张为起始牌,这时开始按次序摸牌,对任何一方来说,都是从 52 张总体中抽取一个 13 张的样本.问:这种样本方法是否为简单随机抽样?

【分析】依据系统抽样的特点来判断.

【解】简单随机抽样的实质是逐个从总体中随机抽样,而这里只是随机确定了起始张,这时其他各张虽然是逐张摸牌的,其实各张在谁手里已被确定,所以不是简单随机抽样,根据其等距摸牌的特点,应

将其定位为系统抽样.

【点拨】判断一种抽样是否为系统抽样,关键是正确理解系统抽样的概念,看其是否符合系统抽样的特点,即:将总体均分为几个部分,按事先确定的规则在各部分抽取.

变式训练

从 2003 名职工中抽取一个由 20 人组成的旅游观光团,试叙述系统抽样的步骤.

【答案】略

【例3】某学校在编人员 160 人,其中行政人员有 16 人,教师有 112 人,后勤人员有 32 人,教育部门为了了解学校机构改革意见,要从中抽取一个容量为 20 的样本,试确定用何种方法抽取,并写出抽样过程.

【分析】因为本题样本总体分成三类:行政人员、教师、后勤人员,符合分层抽样的特点,故选用分层抽样方法.

【解】因为 $\frac{20}{160} = \frac{1}{8}$,所以从行政人员中抽取 $16 \times \frac{1}{8} = 2$ (人),从教师中抽取 $112 \times \frac{1}{8} = 14$ (人),从后勤人员中抽取 $32 \times \frac{1}{8} = 4$ (人).

因为行政人员和后勤人员较少,可将他们分别按 1~16 编号和 1~32 编号,然后采用抽签法分别抽取 2 人和 4 人,对教师按 000,001,……,111 编号,然后用随机数法抽取 14 人.

【点拨】(1)当已知总体是由差异明显的几部分组成时,为了使样本更充分地反映总体的情况,常采用分层抽样法.

(2)分层抽样是将总体分成几层,分层进行抽取,抽取时可采用抽签法和随机数法.

变式训练

某学校有在校高中学生 2400 人,其中高一学生 720 人,高二学生 800 人,高三学生 880 人,如果想通过抽取其中 120 人来调查学生的课外阅读情况,考虑到学生的年级不同,课外阅读情况有明显的差异,而同一年级内课外阅读情况差异较小,问:应采用怎样的抽样方法?每个年级中应依次抽取多少人?

【答案】略

课堂基础自测

1. 统计的基本思想是

【 】

- A. 用样本估计总体的变化
- B. 用样本估计总体
- C. 用总体估计样本
- D. 样本平均数与总体平均数相等

2. 某次考试有 70000 名学生参加,为了了解这 70000 名考生的数学成绩,从中抽取 1000 名考生的数学成绩进行统计分析,在这个问题中,有以下四种说法:①1000 名考生是总体的一个样本;②1000 名考生数学成绩的平均数是总体平均数;③70000 名考生是总体;④样本容量是 1000. 其中正确的说法有 【 】

- A. 1 种
- B. 2 种
- C. 3 种
- D. 4 种

3. 关于简单随机抽样的特点,有以下几种说法,其中不正确的是 【 】

- A. 要求总体的个数有限
- B. 从总体中逐个抽取
- C. 这是一种不放回抽样
- D. 每个个体被抽到的机会不一样,与先后有关

4. 为了解 1200 名学生对学校食堂的意见,打算从中抽取一个样本容量为 30 的样本,考虑采用系统抽样,则最合适的分段间隔 k 为 【 】

- A. 40
- B. 30
- C. 20
- D. 12

5. 从 2006 名学生志愿者中选取 50 名组成一个志愿者团,若采用下面的方法选取:先用简单随机抽样从 2006 人中剔除 6 人,余下的 2000 人再按系统抽样的方法进行,则每人入选的机会 【 】

- A. 不全相等
- B. 均不相等
- C. 都相等
- D. 无法确定

6. 甲校有 3600 名学生,乙校有 5400 名学生,丙校有 1800 名学生,为统计三校学生某方面的情况,计划采用分层抽样法,抽取一个容量为 90 人的样本,应在这三校分别抽取学生 【 】

- A. 30 人,30 人,30 人
- B. 30 人,45 人,15 人
- C. 20 人,30 人,10 人
- D. 30 人,50 人,10 人

7. 从某鱼池中随机捕得 120 条鱼,做了记号之后,再放回池中,经过适当的时间后,再从池中随机捕得 100 条鱼,计算其中有记号的鱼为 10 条,试估计鱼池中有条数为 【 】

8. 一个总体中共有 100 个个体,随机编号 00, 01, 02, 03, ..., 99, 依编号顺序平均分成 10 个小组,组号依次为 01, 02, 03, ..., 10. 现用系统抽样的方法抽取一个容量为 10 的样本,规定如果在第 1 组随机

抽取的号码为 m ,那么在第 k 组中抽取的号码个位数字与 $m+k$ 的个位数字相同. 若 $m=6$,则在第 7 组中抽取的号码是 【 】

9. 某地区有农民、工人、知识分子家庭共计 2004 户,其中农民家庭 1600 户、工人家庭 303 户. 现要从中抽出容量为 40 的样本,则在整个抽样过程中,可以用到下列抽样方法中的 【 】. (将你认为正确的序号都写上)

- ①简单随机抽样
- ②系统抽样
- ③分层抽样

综合能力拓展

10. (2006 · 山东)某学校共有师生 2400 人,现用分层抽样的方法,从所有师生中抽取一个容量为 160 的样本. 已知从学生中抽取的人数为 150,那么该学校的教师人数是 【 】

11. 某超市有普通水果和无公害水果若干千克,现按 5% 的比例分层抽样,抽取了 15 千克普通水果、45 千克无公害水果进行分析,则该超市共有水果 【 】 千克.

12. 某校有高一学生 400 人,高二学生 302 人,高三学生 250 人,现在按年级分层抽样,从所有学生中抽取一个容量为 190 的样本. 求:应该剔除多少人? 每个年级分别应抽取多少人?

13. 某服装厂平均每小时大约生产服装 362 件,要求质检员每小时抽取 40 件服装检验其质量状况,请你设计一个调查方案.

1.3 统计图表

自主探究学习

1. 统计图表就是_____和_____数据的重要工具.

2. 在初中阶段我们已经学习过的统计图有_____,_____和_____,本节我们会学习到茎叶图,用茎叶图表示数据的两个突出的优点是:_____与_____.

名师要点解析

【要点导学】

我们在初中学习了条形统计图、折线统计图、扇形统计图,并能解决简单的实际问题.在此基础上,高中阶段还将进一步学习茎叶图,体会它们各自的特点;通过实例体会分布的意义和作用,在表示样本数据的过程中,学会列频率分布表,画频率分布直方图、频率折线图、茎叶图;能根据问题的需要选择合适的统计图表,并能用自己的方式表示.

本节的重点是条形统计图、折线统计图、扇形统计图和茎叶图的绘制;难点是通过实例体会频率分布直方图、频率折线图、茎叶图的各自特征,从而恰当地选择上述方法分析样本的分布;要能够把问题用数量化的形式表示出来,并利用数学工具来解决,但要注意一种图表并不能全面地反映总体数据,要在具体问题中根据情况有针对性地选择一些合适的

统计图表.

【经典例题】

【例 1】某市对上、下班交通情况作抽样调查,上、下班时间各抽取了 12 辆机动车,其行驶时速如下(单位:km/h):

上班时间:30,33,18,27,32,40,26,28,35,20,21,28;

下班时间:27,19,32,29,36,29,30,22,25,16,17,30.

用茎叶图表示上面的样本数据,并求出样本数据的中位数.

【解】根据题意绘出该市上、下班交通情况的茎叶图,如图 1.3-1 所示:

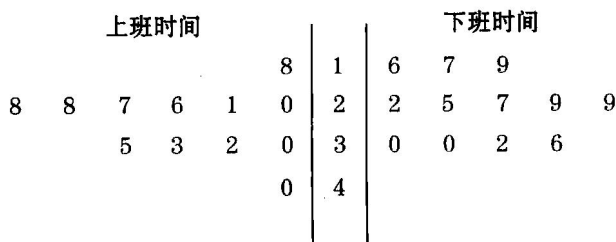


图 1.3-1

由图可见,上班时行驶时速的中位数是 28,下班时行驶时速的中位数是 28.

【点拨】茎叶图保留了原始数据,所有的数据信息都可以很容易地从表中获得.

【例 2】甲、乙两人数学成绩的茎叶图如图 1.3-2 所示(单位:分).

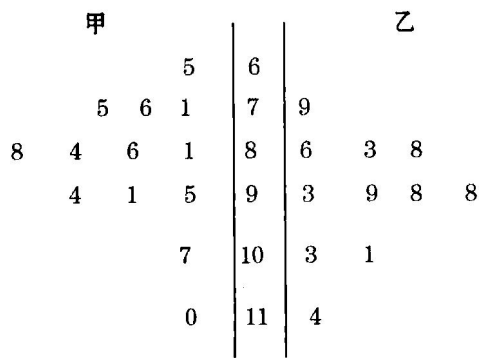


图 1.3-2

(1) 求出这两名同学的数学成绩的平均数;

(2) 比较这两名同学的数学成绩,谈谈你的看法.

【解】(1) 由表中数据,用计算器求得 $\bar{x}_{\text{甲}} = 87$ 分, $\bar{x}_{\text{乙}} = 95$ 分;

(2) 由 $\bar{x}_{\text{甲}} = 87$ 分, $\bar{x}_{\text{乙}} = 95$ 分,故甲的数学学习

平均水平不如乙. 从茎叶图上可以看出:乙同学的得分情况是大致对称的,中位数是 99 分;甲同学的得分情况除一个特殊得分以外,也大致对称,中位数是 89 分. 因此,乙同学发挥得比较稳定,总体得分情况比甲同学好.

【点拨】首先由茎叶图读出数据,再利用计算器求值,依据结果进行比较,并与茎叶图比较统计作用.

课堂基础自测

1. 在如图 1.3-3 所示的茎叶图表示的数据中,众数和中位数分别为 【 】

- A. 23 与 26
B. 31 与 26
C. 23 与 30
D. 26 与 30

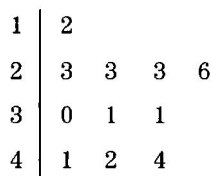


图 1.3-3

2. 下列关于茎叶图的描述:①茎叶图保留了原始数据,便于比较;②茎叶图方便随时记录,便于表示;③数据越多或数据位数越多时,茎叶图越方便. 其中,正确的是 【 】

- A. ①②③
B. ①②
C. ①③
D. ②③

3. 有一个容量为 100 的样本,数据的分组及各组的样本数如下: [12.5, 15.5), 6; [15.5, 18.5), 16; [18.5, 21.5), 18; [21.5, 24.5), 22; [24.5, 27.5), 20; [27.5, 30.5), 10; [30.5, 33.5), 8. 要研究总体数据分布情况应选用的统计图是 【 】

- A. 扇形图
B. 折线图
C. 茎叶图
D. 频率分布直方图

4. 某地 2006 年 3 月的天气情况如图 1.3-4 所示,则这月晴天的天数是 【 】

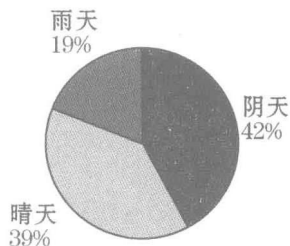


图 1.3-4

- A. 12
B. 13
C. 6
D. 10

5. 甲、乙两班举行电脑汉字输入速度比赛,参加的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	平均字数
甲	55	149	191	135
乙	55	151	110	135

根据上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乙班优秀的人数多于甲班优秀的人数(每分钟输入汉字的个数大于 150 为优秀);③甲班成绩的波动性比乙班成绩的波动性大. 其中,正确的是 【 】

- A. ①②③
B. ①②
C. ①③
D. ②③

6. 比较图 1.3-5 所示两幅频数分布直方图中的数据,平均值较大的为 【 】

- A. 图 a
B. 图 b
C. 一样大
D. 不能确定

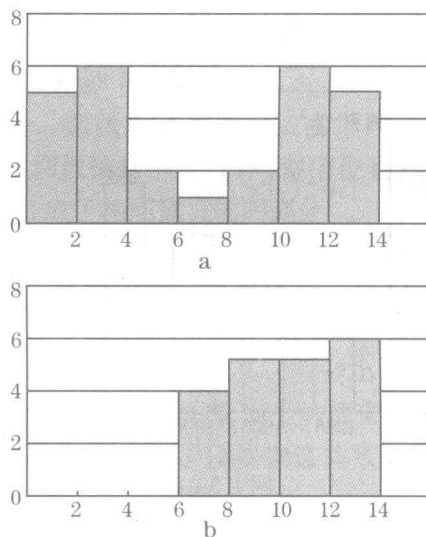


图 1.3-5

7. 如图 1.3-6a、b 两组数据中,波动比较大的为 【 】

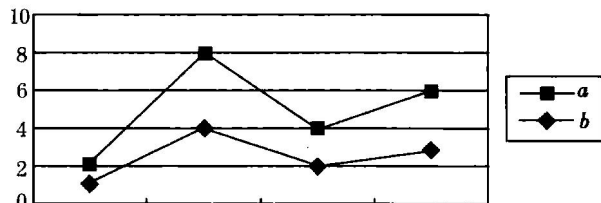


图 1.3-6

- A. a 组
B. b 组
C. 一样大
D. 无法确定

8. 甲、乙两人在相同条件下各射靶 10 次,每次射靶的成绩情况如图 1.3-7 所示. 则甲、乙命中 9

环以上(包括9环)的次数分别为 【 】

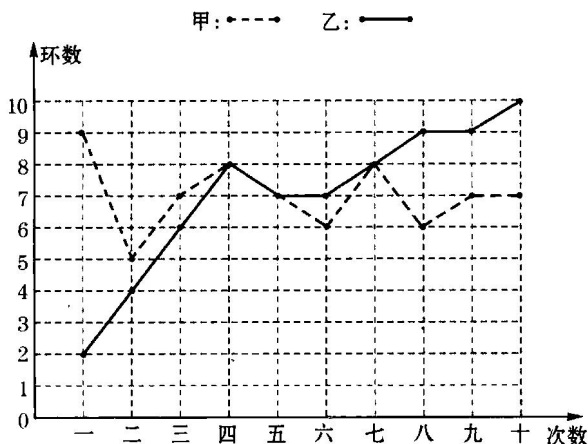


图 1.3-7

- A. 0, 1 B. 1, 3
C. 2, 2 D. 1, 2

9. 据新华社 2002 年 3 月 12 日电, 1985 ~ 2000 年我国农村人均居住面积如图 1.3-8 所示. 其中, 从 _____ 年到 _____ 年这五年间增长最快.

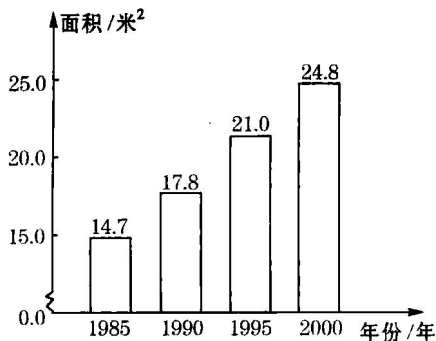


图 1.3-8

10. 200 辆汽车通过某一路段时, 时速频率分布直方图如图 1.3-9 所示, 则时速在 $[50, 60]$ 的汽车大约有 _____ 辆.

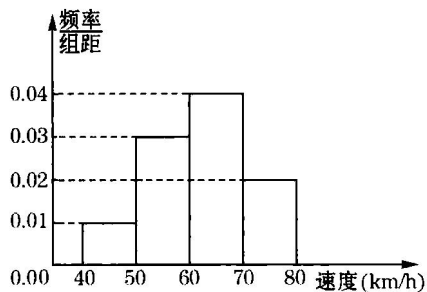


图 1.3-9

综合能力拓展

11. 观察图 1.3-10, 回答下列问题:

某种学生快餐 (300g) 营养成分统计图

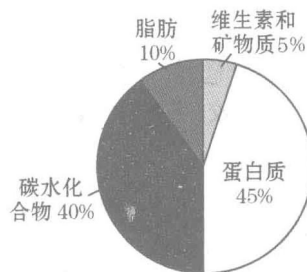


图 1.3-10

- (1) 这种快餐中, 脂肪占百分之几? 有多少克?
- (2) 这种快餐中, 哪一种营养成分最多? 哪一种最少? 最多的营养成分是最少的多少倍?
- (3) 表示蛋白质的扇形的圆心角是多少度? 你是怎样计算的?

12. 英语老师布置了 10 道选择题, 小丽将全班同学的解题情况绘成了条形统计图(如图 1.3-11 所示). 根据图表求平均每个学生做对的题数.

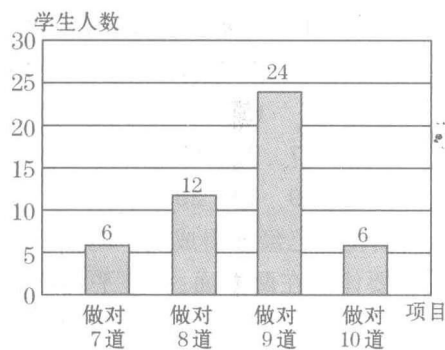


图 1.3-11

13. 在某电脑杂志的一篇文章中, 每个句子的字数如下:

10, 28, 31, 17, 23, 27, 18, 15, 26, 24, 20, 19, 36, 27, 14, 25, 15, 22, 11, 24, 27, 17.

在某报纸的一篇文章中, 每个句子的字数如下:

27, 39, 33, 24, 28, 19, 32, 41, 33, 27, 35, 12, 36, 41, 27, 13, 22, 23, 18, 46, 32, 22.

(1) 用茎叶图表示这两组数据;

(2) 将这两组数据进行比较分析, 你能得出什么结论?

1.4 数据的数字特征

自主探究学习

样本数据的数字特征有 _____ 等, _____ 是三种最常用的数字特征. _____ 就是样本数据中出现次数最多的那个值. _____ 把样本数据分成了数目相同的两部分, 其中一部分都比这个数小, 另一部分都比这个数大. 我们可以把 _____ 比喻成为样本数据的“重心”. _____ 是样本数据到平均值的一种平均距离.

名师要点解析

【要点导学】

1. 众数、中位数与平均数

(1) 一组数据中出现次数最多的数据叫做众数. 如果变量是分类的, 用众数是很有必要的. 例如班委会要作出一项决定, 考察全班同学对它赞成与否就可以用众数.

(2) 将一组数据从小到大依次排列, 把中间数据(或中间两数据的平均数)叫做中位数. 中位数把样本数据分成了相同数目的两部分.

(3) 样本数据的算术平均数, 即 $\bar{x} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$.

注: 由于众数仅能刻画某一数据出现的次数较多, 中位数对极端值不敏感, 而平均数又受极端值左右, 因此这些因素制约了仅依赖这些数字特征来估计总体数字特征的准确性.

2. 标准差与方差

数据的离散值程度可以用极差、方差或标准差来描述. 极差反映了一组数据变化的幅度; 样本方差描述了一组数据围绕平均数波动的大小; 样本方差的算术根表示样本的标准差, 它也描述了数据对平均数的离散程度. 一般来说, 用样本去估计总体时, 样本越具有代表性, 容量越大, 这时对总体的估计也就越精确.

【经典例题】

【例 1】为了保护学生的视力, 教室内的日光灯在使用一段时间后必须更换. 已知某校使用的 100 支日光灯在必须换掉前的使用天数如下表:

使用 天数	151 ~ 180	181 ~ 210	211 ~ 240	241 ~ 270	271 ~ 300	301 ~ 330	331 ~ 360	361 ~ 390
灯泡 数	1	11	18	20	25	16	7	2

(1) 试估计这种日光灯的平均使用寿命.

(2) 若定期更换, 可选择多长时间统一更换合适?

【分析】总体的平均数与标准差往往很难求, 甚至是不可求的, 通常的做法就是用样本的平均数与标准差去估计总体的平均数与标准差. 只要样本的代表性好, 这种做法就是合理的.

【解】(1) 各组组中值分别为 165, 195, 225, 255, 285, 315, 345, 375, 由此算得平均数约为 $165 \times 1\% + 195 \times 11\% + 225 \times 18\% + 255 \times 20\% + 285 \times 25\% + 315 \times 16\% + 345 \times 7\% + 375 \times 2\% = 267.9 \approx 268$ (天).

(2) 将组中值对于此平均数求方差:

$$\frac{1}{100} \times [1 \times (165 - 268)^2 + 11 \times (195 - 268)^2 + 18 \times (225 - 268)^2 + 20 \times (255 - 268)^2 + 25 \times (285 - 268)^2 + 16 \times (315 - 268)^2 + 7 \times (345 - 268)^2 + 2 \times (375 - 268)^2] = 2128.6 (\text{天}^2),$$

故标准差为 $\sqrt{2128.6} \approx 46$ (天).

答: 估计这种日光灯的平均使用寿命约为 268 天, 标准差约为 46 天, 故可在 222 天到 314 天统一更换较合适.

【点拨】(1) 在刻画样本数据的分散程度上, 方差与标准差是一样的, 但在解决实际问题时, 一般多采用标准差;

(2) 平均数和标准差是工业生产中检测产品质量的重要指标, 当样本的平均数或标准差超过了规定界限的时候, 说明这些产品的质量可能距生产要求有较大的偏离, 应该进行检查, 找出原因, 从而及时解决问题.

【例 2】一位研究化肥的科学家将一片土地划分成 100 个 50m^2 的小块, 并在 50 个小块上施用新化肥, 留下 50 个条件大致相当的小块不施用新化肥, 施用新化肥的 50 块土地的小麦产量如下 (单位: kg):

15 29 22 15 3 30 22 16 5 2
22 13 20 25 42 25 20 38 12 29
14 21 26 13 21 27 13 21 11 18
10 18 24 24 36 34 23 18 10 9

17 23 33 8 16 23 31 16 23 40
没有施用新化肥的 50 块土地上的产量如下 (单位: kg):

23 16 16 17 22 3 10 10 8 14
16 5 24 16 32 23 15 18 9 21
4 24 5 24 15 2 15 25 17 29
33 39 16 17 2 15 17 17 26 13
26 11 18 19 12 20 27 12 28 22

你认为新化肥已经取得成功了吗?

【解】主要体现在两方面: 一是平均产量是否提高; 二是产量是否相对稳定. 这就需分别考察两组数据的平均数与标准差. 利用科学计算器进行运算:

第一组数据平均数 $\bar{x} = 20.52$, 标准差 $s = 9.2212$;

第二组数据平均数 $\bar{x}' = 17.36$, 标准差 $s' = 8.2407$.

可见, 施用新化肥后平均产量有了提高, 且产量相对稳定 (虽然出现了几个极端值), 所以新化肥已经取得了初步成效, 还应在稳定性上加强, 新化肥是比较成功的.

课堂基础自测

1. 一组数据的方差是 s^2 , 将这组数据中的每个数据都乘 2, 所得到的一组新数据的方差是 【 】

A. $\frac{1}{2}s^2$ B. s^2 C. $2s^2$ D. $4s^2$

2. 人们常用来反映数据 $x_1, x_2, \dots, x_n$ 的离散程度的量是 【 】

A. 中位数 B. 众数
C. 标准差 D. 平均值

3. 已知 n 个数据 $x_1, x_2, \dots, x_n$, 那么 $\frac{1}{n} [(x_1^2 + x_2^2 + \dots + x_n^2) - n\bar{x}^2]$ 是指 【 】

A. $\bar{x}$ B. $\bar{x}^2$ C. s D. s^2

4. 设有两组数据 $x_1, x_2, \dots, x_n$ 与 $y_1, y_2, \dots, y_n$, 它们的平均数分别是 $\bar{x}$ 和 $\bar{y}$, 则新的一组数据 $2x_1 - 3y_1 + 1, 2x_2 - 3y_2 + 1, \dots, 2x_n - 3y_n + 1$ 的平均数是 【 】

A. $2\bar{x} - 3\bar{y}$ B. $2\bar{x} - 3\bar{y} + 1$
C. $4\bar{x} - 9\bar{y}$ D. $4\bar{x} - 9\bar{y} + 1$

5. $\bar{x}$ 是 $x_1, x_2, \dots, x_{100}$ 的平均数, a 是 $x_1, x_2, \dots, x_{40}$ 的平均数, b 是 $x_{41}, x_{42}, \dots, x_{100}$ 的平均数, 则下列各式正确的是 【 】

$$\text{A. } \bar{x} = \frac{40a+60b}{100} \quad \text{B. } \bar{x} = \frac{60a+40b}{100}$$

$$\text{C. } \bar{x} = a+b \quad \text{D. } \bar{x} = \frac{a+b}{2}$$

6. 数据 $x_1, x_2, \dots, x_n$ 的平均数为 $\bar{x}$, 标准差为 s , 方差为 s^2 , 则数据 $3x_1+5, 3x_2+5, \dots, 3x_n+5$ 的平均数和方差分别是 【 】

$$\text{A. } \bar{x} \text{ 和 } s^2$$

$$\text{B. } 3\bar{x} \text{ 和 } 3s^2$$

$$\text{C. } 3\bar{x}+5 \text{ 和 } 9s^2$$

$$\text{D. } 3\bar{x}+5 \text{ 和 } 9s^2+30s+25$$

7. 甲、乙两名射击运动员参加某大型运动会的预选赛, 他们分别射击了 5 次, 成绩如下表(单位: 环):

甲	10	8	9	9	9
乙	10	10	7	9	9

如果甲、乙两人中只有 1 人入选, 则入选的应是 _____.

8. 甲、乙两个总体各抽取一个样本, 甲的样本均值为 15, 乙的样本均值为 17, 甲的样本方差为 3, 乙的样本方差为 2, _____ 的总体波动小.

综合能力拓展

9. 已知数据 $x_1, x_2, \dots, x_{10}$ 的平均数 $\bar{x} = 20$, 方差 $s^2 = 0.015$. 求:

(1) $3x_1, 3x_2, \dots, 3x_{10}$ 的平均数和方差;

(2) $4x_1-2, 4x_2-2, \dots, 4x_{10}-2$ 的平均数和方差.

10. 为了了解高三年级(1)(2)班的数学学习情况, 从两个班各抽出 10 名学生进行数学水平测试, 成绩如下(单位: 分):

(1)班: 76, 90, 84, 86, 81, 87, 86, 82, 85, 83;

(2)班: 82, 84, 85, 89, 79, 80, 91, 89, 79, 74.

比较两组数据的方差, 并估计(1)(2)两个班哪个班学生的数学成绩比较整齐.

11. 个体户王某经营一家餐馆, 下面是餐馆所有工作人员在某个月份的工资:

王某	厨师甲	厨师乙	杂工	招待甲	招待乙	会计
9000 元	1350 元	1200 元	960 元	1050 元	960 元	1230 元

(1) 计算平均工资;

(2) 计算出的平均工资能否反映帮工人员这个月的收入的一般水平?

(3) 去掉王某的工资后, 再计算平均工资;

(4) (3) 中的平均工资能代表帮工人员的收入吗?

(5) 根据以上计算, 从统计的观点看, 你对(1)(3)的结果有什么看法?

1.5 用样本估计总体

1.6 统计活动：结婚年龄的变化

自主探究学习

1. 列频率分布表,画频率分布直方图的步骤:

(1) 计算_____,即计算最大值与最小值的差;

(2) 决定_____ (组数与样本容量有关,一般来说,样本容量越大,分组就越多,样本容量不超过 120 时,按数据的多少,常分成 5~12 组);

(3) 将数据_____;

(4) 列_____;

(5) 画_____.

2. 用样本的频率分布去估计总体的分布时,如果样本容量不断增大,分组的组距不断缩小,则相应的频率折线图会越来越接近于一条_____,统计中称这条光滑曲线为_____.

名师要点解析

【要点导学】

1. 总体分布反映了在各个范围内取值的可能性,由于总体分布通常不易知道,我们往往是用样本的频率分布去估计总体分布.一般说来,样本的容量越大,这种估计就越精确.

2. 在画频率分布直方图的过程中,要做到分组合理,确定恰当的组距,严格按步骤画出频率分布直方图.

3. 用样本估计总体,是研究统计问题的一个基本思想方法,而对于总体分布,我们总是用样本的频率分布对它进行估计,因此,初中学过的样本的频率分布知识是研究总体分布的基础,需要做好复习.

4. 求一组数据的频率分布,可以按下面的步骤进行:

(1) 频率分布表列出的是在各个不同区间内取值的频率,频率分布直方图是用小长方形面积的大小来表示在各个区间内取值的频率.直角坐标系中的纵轴表示频率与组距的比值,即小长方形的面积

$$= \text{组距} \times \frac{\text{频率}}{\text{组距}} = \text{频率}.$$

(2) 各组频率的和等于 1,即所有长方形面积的和等于 1.

(3) 频率分布表在数量表示上比较确切,但不

够直观、形象,不利于分析数据分布的总体态势.

(4) 从频率分布直方图可以清楚地看出数据分布的总体态势,但是从直方图本身得不出原始的数据内容.

【经典例题】

【例 1】观察下面表格:

(1) 完成表中的频率分布表;

(2) 根据表格,画出频率分布直方图.

分组	频数	频率
[20.7, 20.8)	4	
[20.8, 20.9)	8	
[20.9, 21.0)	13	
[21.0, 21.1)	16	
[21.1, 21.2)	23	
[21.2, 21.3)	20	
[21.3, 21.4)	10	
[21.4, 21.5)	4	
[21.5, 21.6)	2	
合计	100	

【解】(1) 频率依次为:0.04, 0.08, 0.13, 0.16, 0.23, 0.20, 0.10, 0.04, 0.02;

(2) 频率分布直方图如图 1.5-1 所示.

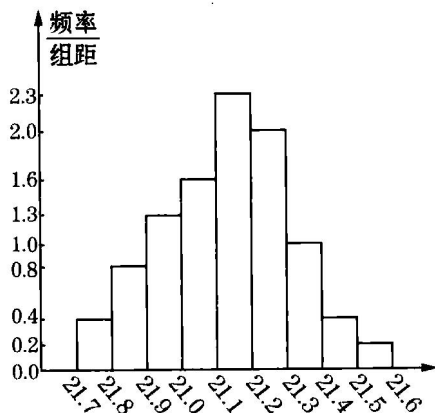


图 1.5-1

【例 2】一个社会调查机构就某地居民的月收入调查了 10000 人,并根据所得数据画了样本的频率分布直方图(如图 1.5-2).为了分析居民的收入与年龄、学历、职业等方面的关系,要从这 10000 人中再用分层抽样方法抽出 100 人作进一步调查,则在 [2500, 3000) (元)月收入段应抽出_____人.

【解】该题考查统计内容的知识.首先处理频率

分布直方图,月收入出现在 $[2500,3000)$ (元)内居民的概率是 $500 \times 0.0005 = 0.25$,这就是分层抽样的抽样比,故从100个居民样本中月收入在 $[2500,3000)$ (元)内抽取的人数为 $100 \times 0.25 = 25$ (人).

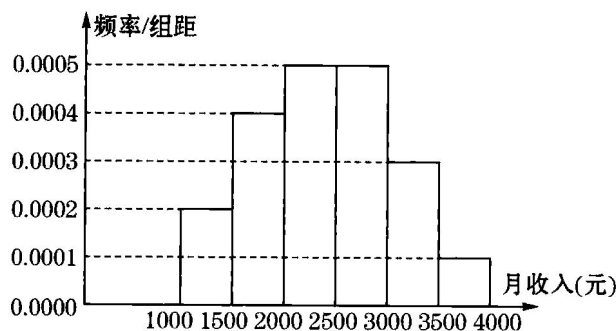


图 1.5-2

课堂基础自测

1. 关于频率分布直方图中的有关数据,下列说法正确的是 【 】

- A. 直方图的高表示取某数的频率
- B. 直方图的高表示该组上的个体在样本中出现的频率
- C. 直方图的高表示该组上的个体数与组距的比值
- D. 直方图的高表示该组上的个体在样本中出现的频率与组距的比值

2. 一个容量为32的样本,已知该组样本的频率为0.125,则该组样本的频数为 【 】

- A. 2 B. 4 C. 6 D. 8

3. 在频率分布直方图中共有11个小长方形,若中间一个小长方形面积等于所有各长方形面积和的 $\frac{1}{4}$,样本容量是160,则中间一组的频数是 【 】

- A. 32 B. 0.2
C. 40 D. 0.25

4. 已知样本:

10, 8, 6, 10, 13, 8, 10, 12, 11, 7,
8, 9, 11, 9, 12, 9, 10, 11, 12, 12.

那么频率为0.3的范围是 【 】

- A. 5.5 ~ 7.5
B. 7.5 ~ 9.5
C. 9.5 ~ 11.5
D. 11.5 ~ 13.5

5. 容量为100的样本数据,按从小到大的顺序分为8组,如下表:

组号	1	2	3	4	5	6	7	8
频数	10	13	14	14	15	13	12	9

第3组的频数和频率分别是 【 】

- A. 14 和 0.14
B. 0.14 和 14
C. $\frac{1}{14}$ 和 0.14
D. $\frac{1}{3}$ 和 $\frac{1}{14}$

6. 在用样本频率估计总体分布的过程中,下列说法正确的是 【 】

- A. 总体容量越大,估计越精确
B. 总体容量越小,估计越精确
C. 样本容量越大,估计越精确
D. 样本容量越小,估计越精确

7. 统计某校400名学生的数学会考成绩,得到样本频率分布直方图如图1.5-3所示,规定不低于60分为及格,不低于80分为优秀,则及格率是_____,优秀人数是_____人.

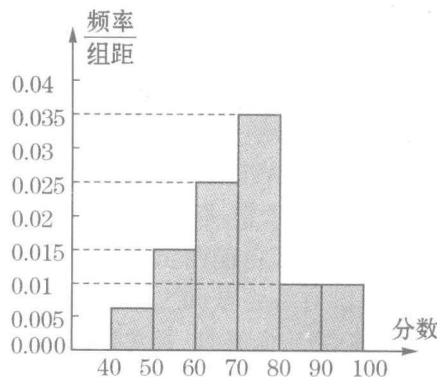


图 1.5-3

8. 将一个容量为 n 的样本分成若干组,已知某组的频数和频率分别为30和0.25,则 $n =$ _____.

9. 已知10个数据:63, 65, 67, 69, 66, 64, 66, 64, 65, 68. 对这些数据编制频率分布表,其中 $(64.5, 66.5)$ 的频率是_____.

综合能力拓展

10. 某赛季甲、乙两名篮球运动员每场比赛得分情况如下:

甲的得分:12, 15, 24, 25, 31, 31, 36, 36, 37, 39, 44, 49, 50;

乙的得分:8, 13, 14, 16, 23, 26, 28, 33, 38, 39, 51.

请设计适当的茎叶图表示上述数据,并利用茎叶图对两名运动员的成绩进行比较.

(1) 求前三个小组的频率分别是多少;

(2) 抽取的顾客人数是多少?

(3) 尺码落在区间 $(37.5, 43.5)$ 内的可能性是多少?

11. 有一容量为 50 的样本,数据的分组及各组的频率数如下: $[10, 15)$, 4; $[15, 20)$, 5; $[20, 25)$, 10; $[25, 30)$, 11; $[30, 35)$, 9; $[35, 40)$, 8; $[40, 45)$, 3.

(1) 列出样本的频率分布表;

(2) 画出频率分布直方图.

13. 对某班 50 人进行智力测验,其得分如下(单位:分):

73, 48, 64, 82, 86, 71, 48, 64, 41, 86, 79, 71, 68, 82, 84, 68, 64, 62, 68, 81, 57, 90, 56, 52, 74, 78, 47, 66, 55, 64, 56, 88, 69, 40, 73, 97, 68, 56, 67, 59, 70, 52, 79, 44, 55, 69, 62, 58, 32, 58.

(1) 列出样本的频率分布表;

(2) 画出频率分布直方图.

12. 为了了解商场某日旅游鞋的销售情况,抽取了部分顾客的鞋的尺码,将所得数据整理后,画出频率分布直方图,如图 1.5-4 所示. 已知从左至右前 3 个小组的频率之比为 1:2:3,第 4 小组与第 5 小组的频率分别为 0.175 和 0.075,第 2 小组的频数为 10.

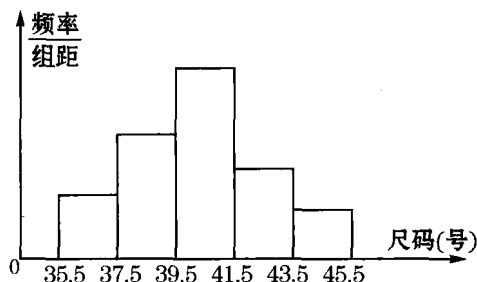


图 1.5-4

1.7 相关性

自主探究学习

1. 当散点图中点散布在从左下角到右上角的区域时称为_____,点散布在从左上角到右下角的区域时称为_____.

2. 相关关系:自变量取值_____时,因变量的取值带有_____的两个变量之间的关系叫做相关关系. 相关关系与函数关系的相同点是两变量之间存在_____.

3. 具有相关关系的两个变量组成一组数据,将各组数据在直角坐标系中描点,这种图形叫_____.

名师要点解析

【要点导学】

本节首先从实例考查变量之间的关系,认识在现实世界中存在不能用函数模型描述的变量关系,从而体会研究变量之间的相关关系的重要性;然后通过绘制散点图,从散点图上判断这两者之间是否存在相关关系;能够近似地描述这种线性关系,画出直线,利用它们之间的近似关系作一个估计;通过解决实际问题,体会统计与确定性思维的差异;通过分析具有不同相关系数的数据的散点图,进一步加深对相关系数的直观理解.

本节的重点是通过收集现实问题中两个有关联变量的数据作出散点图,并利用散点图直观认识变量间的相关关系;难点是用不同的方法来拟合两个变量之间的线性关系;关键是在探究的过程中寻求较好的拟合方法,并注意相关关系和函数关系的区别.

【经典例题】

【例1】班主任为了对本班学生的考试成绩进行分析,决定从全班25名女同学、15名男同学中随机抽取一个容量为8的样本进行分析.若这8位同学的数学、物理成绩对应如下表(单位:分):

学生编号	1	2	3	4	5	6	7	8
数学成绩 x	60	65	70	75	80	85	90	95
物理成绩 y	72	77	80	84	88	90	93	95

根据上表数据用变量 y 与 x 的相关关系画出样本的散点图,并说明物理成绩 y 与数学成绩 x 之间是正相关还是负相关.

【解】画样本散点图如图 1.7-1:

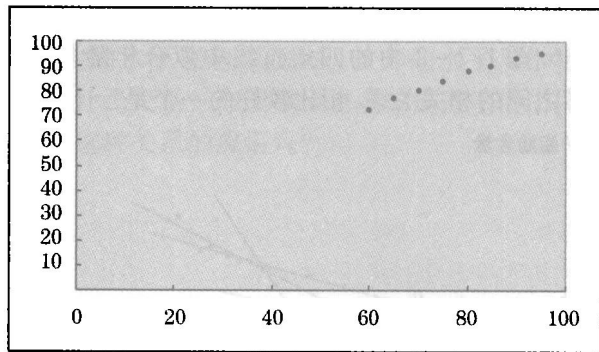


图 1.7-1

由图可知:物理成绩 y 与数学成绩 x 之间是正相关关系.

【例2】两个变量成线性相关关系时,散点图有什么特点?

【解】两个变量成线性相关关系时,散点图上所有的点看上去都在一条直线附近波动.此时,可以用一条直线来拟合.

课堂基础自测

1. “吸烟有害健康”,那么吸烟与健康之间存在的关系为 _____ 【 】

- A. 函数关系 B. 相关关系
C. 无相关关系 D. 不确定

2. 下列对相关关系的理解:①变量之间只有函数关系,不存在相关关系;②两个变量之间产生相关的原因是许多不确定的随机因素的影响;③需要通过样本来判断变量之间是否存在相关关系;④相关关系是一种因果关系,具有确定性.其中,正确的是 _____ 【 】

- A. ①④ B. ③④ C. ②③ D. ②④

3. 下列两个变量之间的关系不是函数关系的是 _____ 【 】

- A. 角度和它的正弦值
B. 圆的半径和该圆的面积
C. 正 n 边形的边数和内角和
D. 人的年龄和身高

4. 炼钢时钢水的碳含量与冶炼时间有 _____ 【 】

- A. 确定性关系 B. 相关关系
C. 函数关系 D. 无任何关系

5. 下面哪些变量是相关关系 _____ 【 】

- A. 出租车费与行驶里程
B. 房屋面积与房屋价格
C. 身高与体重