

医学试题精编丛书

MEDICAL
TEST
SERIES



卫生统计学

主编 赵耐青

MEDICAL
TEST
SERIES

 复旦大学出版社
www.fudanpress.com.cn

医学试题精编丛书

卫生统计学

主 编 赵耐青

副主编 施学忠 康晓平 贺 佳 余红梅

编 委(以姓氏笔画为序)

朱 焱	贵阳医学院
毕 煜	复旦大学
杨永利	郑州大学
李 强	西安交通大学
李丽霞	广东药学院
余红梅	山西医科大学
金 欢	复旦大学
赵 杨	南京医科大学
赵耐青	复旦大学
施学忠	郑州大学
贺 佳	第二军医大学
康晓平	北京大学
彭 斌	重庆医科大学
缪之文	北京大学

复旦大学出版社

图书在版编目(CIP)数据

卫生统计学/赵耐青主编. —上海:复旦大学出版社, 2009. 7
(医学试题精编丛书)
ISBN 978-7-309-06364-6

I. 卫… II. 赵… III. 卫生统计-研究生-入学考试-习题
IV. R195.1-44

中国版本图书馆 CIP 数据核字(2008)第 172149 号

卫生统计学

赵耐青 主编

出版发行 复旦大学出版社 上海市国权路 579 号 邮编 200433
86-21-65642857(门市零售)
86-21-65100562(团体订购) 86-21-65109143(外埠邮购)
fupnet@fudanpress.com http://www.fudanpress.com

责任编辑 贺琦

出品人 贺圣遂

印 刷 常熟市华顺印刷有限公司
开 本 787 × 960 1/16
印 张 13
字 数 226 千
版 次 2009 年 7 月第一版第一次印刷

书 号 ISBN 978-7-309-06364-6/R · 1055
定 价 24.00 元

如有印装质量问题, 请向复旦大学出版社发行部调换。

版权所有 侵权必究

前 言

卫生统计学是分析和处理医学研究中随机现象的一门学科。随着医学科学的发展,医学研究的研究设计、资料收集和结果分析越来越多地运用卫生统计学知识,已经成为医学科学研究中的一个重要组成部分,卫生统计学已经是医学各专业本科生和研究生的必修课程。

许多医学生都感到卫生统计学比较难学,甚至有些医学生对卫生统计学的课程有一定的恐惧感,特别是对卫生统计学的习题感到困难更大。针对上述情况,本书的上篇以人民卫生出版社出版的《卫生统计学》内容为基本框架,先对每一章的学习要点进行归纳,列出每一章的学习重点,然后针对性地列出每章的基本习题及其解答;下篇针对读者常见的数据统计分析中的困难,以高等教育出版社出版的普通高等教育“十一五”国家级规划教材《卫生统计学》的内容为基本学习要点,列出综合性的练习题及其题解。

本书可以作为五年制预防医学专业本科生和其他医学专业的本科生和研究生学习卫生统计学的参考书。

本书的出版得到了复旦大学出版社的大力支持,在此表示衷心的感谢。本书凝聚着全国 10 所院校 14 位编委的智慧和心血,在此谨向各位同仁致以崇高敬意和深深谢意!我们也特别珍惜在本书编写过程中结下的友谊!

由于编者能力所限,不足之处,敬请广大师生提出宝贵意见。

编 者

2009 年 7 月

目 录

上 篇

第一章 绪论	1
第二章 定量资料的统计描述	9
第三章 定性资料的统计描述	15
第四章 常用概率分布	25
第五章 参数估计基础	32
第六章 假设检验	38
第七章 χ^2 检验	47
第八章 实验设计	59
第九章 方差分析	63
第十章 基于秩次的非参数检验	69
第十一章 两个变量的关联性分析	76
第十二章 简单回归分析	82
第十三章 多重线性回归与相关	90
第十四章 调查设计	99
第十五章 总体特征抽样调查的设计与分析	103
第十六章 横断面研究资料的统计分析	110
第十七章 队列研究的设计与分析	117
第十八章 病例-对照研究的设计与分析	124
第十九章 Logistic 回归分析	131
第二十章 生存分析	143
第二十一章 综合评价	152
第二十二章 Meta 分析	164

下 篇

综合习题	176
------------	-----

上 篇

第一章 绪 论

学习要点

1. 总体与样本的定义和理解。
2. 变量与资料类型的定义与理解。
3. 参数与统计量的定义与理解。

由于各本教材对变量的分类不太相同,现将变量分类综合归纳为下列框图(图 1-1),供使用不同教材的读者理解、沟通和使用(其中分类资料按类汇总后构成频数的表格资料又称为计数资料)。

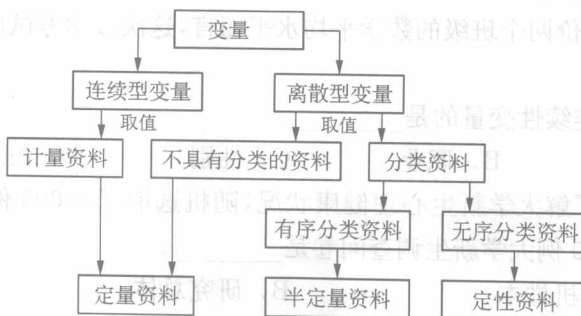


图 1-1 变量与资料的分类关系图

习 题

一、是非题

1. 统计研究的过程包括:研究设计、收集数据、整理数据、分析数据和解释统计分析结果。
2. 数量上有联系的两个变量一定有因果关系。
3. 家庭中子女数是离散型的定量资料。

4. 为了研究某校入学男生的平均身高,研究者随机抽取了1 000名男同学测量其身高,得到平均身高为168 cm,标准差为2.4 cm,可以称样本均数是一个统计量。()
5. 在推断性研究中,研究者往往对样本进行分析,但最终的目的是推断总体。()
6. 同质个体之间的变异称为个体变异。()
7. 学校往往通过考试评价学生的学习效果,但事实上对于评价学生学习效果而言,一次考试就是一次抽样,故存在抽样误差。()

二、选择题

1. 某学校有甲班和乙班,用同一张数学试卷进行考试,经过盲态改卷,甲班平均分为87分,乙班平均分为86分,请问:下列哪个说法是正确的_____。
 - A. 通过这次考试,可以认为甲班的数学平均水平高于乙班
 - B. 通过这次考试,可以认为甲班的这次数学考试平均成绩高于乙班
 - C. 对于评价两个班级的这次数学考试平均成绩而言,两个班级的数学平均分存在抽样误差
 - D. 对于评价两个班级的数学平均水平而言,这次数学考试成绩不存在抽样误差
2. 下列属于连续性变量的是_____。
 - A. 血压
 - B. 职业
 - C. 性别
 - D. 民族
3. 某高校欲了解大学新生心理健康状况,随机选取了1 000例大学新生调查,这1 000例大学新生调查问卷是_____。
 - A. 一份随机样本
 - B. 研究总体
 - C. 目标总体
 - D. 个体
4. 下列描述中,不正确的是_____。
 - A. 在一个同质总体中,主要研究指标的个体观察值之间存在差异
 - B. 在一个同质总体中,主要研究指标的个体观察值之间一般不会出现非常大的差异
 - C. 总体中的个体在同质的基础上有变异
 - D. 如果个体间有变异,那它们肯定不是来自同一总体
5. 下列描述中,不正确的是_____。
 - A. 在干预性研究中,通过随机化、重复等措施可以很好地控制混杂因素对结果的影响,使实验者很好地观察处理因素的作用

- B. 在观察性研究中,研究者只是被动地观察研究结果,不能施加干预措施
- C. 在科学研究中采用的设计方案不同,则采用的分析方法也不同
- D. 在观察性研究中没有办法控制混杂因素
6. 某研究用 x 表示儿童在一年中患感冒的次数,共收集了 1 000 人,请问:儿童在一年中患感冒次数的资料属于_____。
- A. 连续型资料 B. 有序分类资料
- C. 不具有分类的离散型资料 D. 以上均不对

三、简答题

1. 简述总体和样本的定义,并且举例说明。
2. 简述参数和统计量的定义,并且举例说明。
3. 变量的类型有几种? 请举例说明各种类型变量有什么特点?
4. 请举例说明一种类型的变量如何转换为另一种类型的变量。
5. 请简述卫生统计工作可以分为哪几个基本步骤。
6. 请简述什么是小概率事件? 什么是不可能事件? 什么是必然事件? 概率有哪些性质?
7. 在某个治疗儿童哮喘的激素喷雾剂新药的临床试验中,研究者收集了 300 名哮喘儿童患者,随机分为试验组和对照组,试验组在哮喘缓解期内采用激素喷雾剂,在哮喘发作期内采用激素喷雾剂+扩展气管药;对照组在哮喘缓解期不使用任何药物,在哮喘发作期内采用扩展气管药物。通过治疗 3 个月,评价两种治疗方案的有效性和安全性。请阐述这个研究中的总体是什么。
8. 为了研究某药物治疗骨质疏松症的疗效,某医院收集 400 名首次诊断为骨质疏松症的患者(从未治疗过),在治疗前测量每个患者的骨密度,然后每个患者服用该药物,服药满 6 个月时,测量每个患者的骨密度,考察这个药物治疗骨质疏松症的疗效。请问:在这个研究中,总体是如何定义的? 总体均数是如何定义的?
9. 某医生收治 200 名幽门螺杆菌感染(Hp+)患者,随机分成两组,每组 100 人。一组用 A 药,另一组用 B 药。经过 2 个月的治疗,A 药组中有 70 人检查的结果为 Hp 阴性,B 药组有 65 人检查的结果为 Hp 阴性,请根据现有结果评议下列说法是否正确,为什么?
 - (1) A 药组的疗效高于 B 药组。

图 (2) A 药的疗效高于 B 药。

参考答案及注释

一、是非题

1. 对。详细见教材第一章绪论。
2. 错。有联系但不一定是因果关系。如:吸烟的人容易患肺癌,但吸烟的人并不是都患肺癌,不吸烟的人也有患肺癌的,只是吸烟的人更容易患肺癌。
3. 对。家庭中的子女数不存在最大是多少,也没有分类的意义。
4. 对。样本均数是由样本构造的表达式,并且不含有任何未知参数,所以样本均数是一个统计量。
5. 对。一般而言,研究的目的是通过样本信息的统计分析推断总体,否则推断性研究将失去重要意义。
6. 对。见个体变异的定义。
7. 对。评价学生学习状况就是指学生学习成绩的总体情况。每次考试前学生不能预先知道考试成绩,并且任何两次考试的成绩往往是不同的;所以考试的成绩存在一定的随机性,所有的考试成绩的综合才是学生学习成绩的总体情况,所以可以认为学生的一次考试是一次随机抽样。

二、选择题

1. 选 B。因为 A 是指数学总体水平;B 是指样本均数,所以这是一个客观事实;C 的前提是基于评价这次数学考试成绩为目标,而不是评价学生的总体情况而言,所以这次考试不是抽样,因此也就不存在抽样误差。
2. 选 A。其他都是分类资料。
3. 选 A。根据题意可知,这是一次随机抽样,是随机抽样的调查结果,所以这是一份随机样本。
4. 选 D。对于总体中的个体变异是普遍存在的。
5. 选 D。因为是干预性研究,随机化可以控制混杂因素的影响,所以 A 的叙述是正确的;观察性研究的特征就是如 B 所叙述,所以 B 的叙述是正确的;同一研究问题,采用不同的设计,所获取的样本资料的统计性质是不同的(如:配对设计获取资料往往是不独立的),所以要根据设计和资料分布类型,采用相应的统计方法进行分析,因此 C 的叙述是正确的;在观察性研究中,可以采取合适的抽样方案(如通过配对设计)控制混杂因素的影响,所以 D 的叙述是错误的。

6. 选 C。因为一年中的感冒次数是一个离散型的资料,但不具有分类性质,所以选 C 是正确的。

三、简答题

1. 简述总体和样本的定义,并且举例说明。

总体(population):根据研究目的确定的同质观察单位的全体。更严格地说,总体是根据研究目的确定的同质观察单位的观察值全体所构成的集合。样本(sample):从研究总体中抽取的一部分满足代表性的个体观察值所构成的集合称为样本,在收集指标较多时,简单样本往往是从研究总体中抽取的一部分有代表性的个体构成的。例如,研究 1990 年某地区 15~30 岁男性的高血压患病率。假定 1990 年该地区 15~30 岁男性共有 20 万人,则主要观察指标为是否患高血压,1990 年该地区 15~30 岁 20 万名男性是否患高血压的观察值全体构成这个研究的总体,这个研究中的高血压总体患病率就是这 20 万人中患高血压的人数/20 万。总体患病率是指一个总体参数,用于刻画总体特征的一个指标。事实上,很难准确获得一个地区患高血压的人数,但可以在这 20 万人中随机抽取样本估计这个研究总体的高血压患病率。例如,随机抽取 200 人,检查这 200 人是否患高血压,这 200 个人是否患高血压的观察值构成一个样本,并且可以得到这个样本的高血压患病率=这 200 人中患高血压的人数/200。样本患病率是一个统计量,可以用样本患病率估计这个研究的总体患病率,进一步统计分析可以考察样本患病率的估计误差等。

2. 简述参数和统计量的定义,并且举例说明。

参数(parameter):描述总体分布特征的指标,由于总体一般很大,所以参数常常是未知的。例如,总体均数是总体中某个观察指标的所有观察值的平均数,是刻画该观察指标的总体平均水平的特征指标。总体参数通常有总体均数 μ , 总体标准差 σ , 总体率 π ; 统计量(statistics):根据样本构成的表达式,并且表达式中不能含有未知参数。通常统计量是刻画样本的某种特征的指标并且用于估计总体参数。例如,样本均数 $\bar{x}$, 样本标准差 S , 样本率 p 。举例如上题所述。

3. 变量的类型有哪几种? 请举例说明,各有什么特点?

变量可分为定性变量与定量变量。定性变量的取值表现为互不相容的类别或属性,可以分为无序分类与有序分类变量两种。其中无序分类变量又可以分为二分类与多分类变量。例如,性别、某病治疗结果为治愈与未愈,均为二分类变量,血型、职业等为无序多分类变量。为了便于统计处

理与计算机的录入,我们为无序变量进行编码,但只是一个代码,不参与计算;有序分类变量又称等级变量,有大小或顺序的差别。例如,产品的等级、对服务的满意度等级等。

定量变量可以分为连续型变量和离散型变量。连续型变量通常含有度量衡单位,可以比较大小。连续型变量可以在数轴上连续取值。例如,身高、体重、胸围等为连续型变量。离散型变量的取值不连续。例如,家庭中的子女数、孕妇的产次等。变量的分类见图 1-1。不同类型的变量,使用的统计描述指标与统计推断方法也不同。资料分析时,要分清是何种变量,从而选择合适的统计分析方法。例如,对于连续型资料,经常使用 t 检验或方差分析,定性资料使用 χ^2 检验,当然,还要看研究者的目的及所采用的设计方法等。

4. 请举例说明一种类型的变量如何转换为另一种类型的变量。

研究者根据自己的研究目的,可以将一种类型的变量转换为另一种类型的变量,但变量只能从高级向低级转换:连续型 $\rightarrow$ 有序 $\rightarrow$ 分类 $\rightarrow$ 二值,但不能从低级向高级转换。例如,血压值为定量变量,我们可以根据血压值分为:低血压、正常血压、轻度高血压、中度高血压和重度高血压。此时此变量为有序变量(等级变量);我们也可以将血压分为正常、不正常,此时为二分类变量。变量的类型不同,采用的统计学分析方法也不同。

5. 请简述卫生统计工作可以分为哪几个基本步骤。

卫生统计工作有以下 4 个基本步骤。

(1) 设计:一项研究能否成功设计是最关键的环节,按研究中是否对被试对象施加干预,可以分为调查设计与实验设计。调查设计旨在客观地描述研究总体,对研究对象不施加任何干预措施,而实验设计可以人为地施加干预措施。无论调查设计还是实验设计,均包括专业设计与统计学设计两个方面,其中统计学设计包括对资料搜集、整理和分析全过程的设想与安排。

(2) 搜集资料:获取准确、可靠的原始数据。统计资料一般有 3 个方面的来源:统计报表、经常性工作记录、专题调查或实验所获得的数据。

(3) 整理资料:对原始数据进行检查、核对,纠正可能的记录或录入错误,将原始数据根据研究的目的合理分组并归纳汇总,为进一步的分析作基础。

(4) 分析资料:从大量数据中揭示其内在的规律性,包括统计描述与统计推断两个方面。统计描述是指选用合适的统计指标、合适的统计表

与统计图对资料的特征进行描述;统计推断是指用样本的信息去推断总体的特征,包括参数估计与假设检验。

以上步骤是紧密联系的,任何一步有问题都会影响统计分析的最终结果。

6. 请简述什么是小概率事件? 什么是不可能事件? 什么是必然事件? 概率有哪些性质?

概率是描述某随机事件发生的可能性大小的数值。随机事件 A 发生的概率记为 $P(A)$, 随机事件的概率取值在 $0 \sim 1$ 之间, 即 $0 \leq P \leq 1$, P 值越接近 1, 说明随机事件发生的可能性越大; P 值越接近 0, 说明随机事件发生的可能性越小。当观察单位足够多时, 可以将频率做为概率的估计值。对于不可能发生的事件, 发生该事件的概率 $P = 0$, 并称为不可能事件; 对于必然发生的事件, 则发生该事件的概率 $P = 1$, 并称为必然事件。如果某随机事件发生的概率 $P \leq 0.05$, 或 $P \leq 0.01$, 表示该事件发生的可能性很小, 称其为小概率事件。其意义为对于一次随机抽样, 这样的事件一般是不会发生的。因此, 统计推断中, 对于一次随机抽样, 一般认为小概率事件是不会发生的, 这就是统计推断中作出推断结论的理论基础。

7. 在某个治疗儿童哮喘的激素喷雾剂新药的临床试验中, 研究者收集了 300 名哮喘儿童患者, 随机分为试验组和对照组, 试验组在哮喘缓解期内采用激素喷雾剂, 在哮喘发作期内采用激素喷雾剂+扩展气管药; 对照组在哮喘缓解期不使用任何药物, 在哮喘发作期内采用扩展气管药物。通过治疗 3 个月, 评价两种治疗方案的有效性和安全性。请阐述这个研究中的总体是什么。

该研究有两个总体: 试验组所对应的总体是采用试验组治疗方案的全体哮喘儿童患者, 对照组所对应的总体是采用对照组治疗方案的全体哮喘儿童患者。

8. 为了研究某药物治疗骨质疏松症的疗效, 某医院收集 400 名首次诊断为骨质疏松症的患者(从未治疗过), 在治疗前测量每个患者的骨密度, 然后每个患者服用该药物, 服药满 6 个月时, 测量每个患者的骨密度, 考察这个药物治疗骨质疏松症的疗效。请问: 在这个研究中, 总体是如何定义的? 总体均数是如何定义的?

对于治疗骨质疏松症, 通常要考察经过治疗后骨密度是否上升? 骨密度平均上升幅度是多少? 骨密度是否恢复到正常水平? 特别是当骨密度上升到正常水平时, 再考察骨密度上升幅度是没有意义的。因此, 通常要分

两种情况考察:①治疗后骨密度能达到多少水平?②如果骨密度普遍没有恢复正常水平,则还要考虑通过治疗后,骨密度平均上升幅度是多少?因此该研究应该考虑两个效应指标:通过治疗后的骨密度观察值和通过治疗后骨密度改变量,对应的总体分别为初次诊断为骨质疏松症并且接受这种药物治疗达6个月时所有患者的骨密度测量值的全体所构成的集合和初次诊断为骨质疏松症并且接受这种药物治疗达6个月时所有患者的骨密度改变量观察值的全体所构成的集合,相应的总体分别为初次诊断为骨质疏松症并且接受这种药物治疗达6个月时所有患者的骨密度测量值的平均数和初次诊断为骨质疏松症并且接受这种药物治疗达6个月时所有患者的骨密度改变量观察值的平均数。

9. 某医生收治200名幽门螺杆菌感染(Hp+)患者,随机分成两组,每组100人。一组用A药,另一组用B药。经过2个月的治疗,A药组中有70人检查的结果为Hp阴性,B药组有65人检查的结果为Hp阴性,请根据现有结果评议下列说法是否正确,为什么?

(1) A药组的疗效高于B药组。

(2) A药的疗效高于B药。

(1) 所叙述的是样本结果,这是一个客观事实,所以这种说法是正确的,但我们不提倡这样说,因为很容易对不熟悉统计学的研究者造成误导。

(2) 所叙述的是一个对总体所下的结论,仅根据题目所给出的结果是不能对总体直接下结论的,所以不能认为(2)的叙述是正确的。

第二章 定量资料的统计描述

学习要点

1. 掌握定量资料的分布图制作。
 2. 掌握对称分布和非对称分布的概念。
 3. 掌握集中趋势的统计指标(算数均数,几何均数,中位数,百分位数)及其应用对资料的要求。
 4. 掌握离散趋势的统计指标(标准差,方差,变异系数)及其应用对资料的要求。
 5. 了解众数,调和均数,极差,偏度系数,峰度系数的定义、意义及其性质。
 6. 掌握统计图表的基本应用。
- 一般而言,统计描述的策略可以为:
- (1) 如果资料近似呈对称分布,则可以用 $\bar{x} \pm S$ 进行统计描述。
 - (2) 如果资料 X 取对数后, $Y = \lg(X)$ 近似呈对称分布,则可以用 $\bar{y}(G) \pm S_Y$ 进行统计描述,其中 G 为资料 X 的几何均数, S_Y 为取对数后的资料 Y 的标准差。
 - (3) 如果资料呈非对称分布,则可以用中位数($P_{25} \sim P_{75}$)描述,其中 $P_{25} \sim P_{75}$ 称为四分位数范围(inter-quartile range, IQR)。

习 题

一、是非题

1. 绘制箱式图的 5 个基本指标为:最小值、25 百分位数、中位数、75 百分位数、最大值。
2. 算术均数不容易受极值的影响。
3. 对于定量资料,使用直接法与加权法来计算算术均数的结果是相同的。
4. 直条图与直方图都可以用来描述连续型资料。
5. 四分位数间距反映了中间 50%数据的离散程度,易受两边极大、极小值的影响。

影响。

()

6. 定性变量与定量变量都可以用频数分布表来统计描述。

()

二、选择题

1. 抽样调查 120 名 20~30 岁健康女性的胆固醇含量,对于编制等组距的频数表,组距为 8。假定第一个组段的下限为 6,请问:下列陈述中哪项是错误的_____。
A. 第一个组段的组中值和上限分别为 10 和 14
B. 第二个组段的组中值为 18
C. 第三组的上下限分别为 22 和 30
D. 23 属于第二组段
2. 欲比较身高、体重的变异程度,但需要排除量纲对变异程度的影响,最好选择的指标为_____。
A. 变异系数 B. 标准差 C. 方差 D. 四分位数间距
3. 如果某资料呈近似对称分布,使用算术平均数来描述其集中趋势,则宜选用_____来描述其离散趋势。
A. 四分位数间距 B. 标准差
C. 极差 D. 变异系数
4. 为了描述资料分布概况,绘制直方图时,直方图的纵轴可以为_____。
A. 频数 B. 频率 C. 频率密度 D. 都可以
5. 数据集 4.5, 3, 2.5, 5.5, 8, 1, 其中位数为_____。
A. 2.5 B. 3.75 C. 3 D. 4.5

三、简答题

1. 常用的描述集中趋势的指标有哪些,并简述其适用条件。
2. 常用的统计图有哪几种? 它们各自适用于哪些资料?
3. 请简述频数表的用途。
4. 常用的描述离散趋势的指标有哪些? 并简述其适用条件。
5. 算术均数一定大于中位数,这句话对吗? 请加以解释。
6. 滴度资料是否一定可以用几何均数描述集中趋势?
7. 对于同一个非负样本资料,其算术均数是否一定大于等于几何均数?
8. 请叙述 P_{25} 和 P_{75} 的统计学意义。
9. 适用于算数均数的资料能否用中位数描述集中趋势?

10. 如果资料取对数变换后呈对称分布并且样本量非常大,中位数与几何均数是否一定非常接近?为什么?
11. 仅仅从统计描述的角度上考察,能否理解为中位数可以描述任何分布类型的连续型资料的集中趋势?



参考答案及注释

一、是非题

1. 对。依据箱式图的定义。
2. 错。一个特别大的数据可以使算数均数的数值非常大。
3. 错。所谓加权法是指以频数表形式的资料,用组段的组中值近似原始数据,求其算数平均,所以其结果一般与直接法计算均数的结果有些差异,但只要各个组段的组距较小,则误差一般较小或很小。
4. 错。直条图一般用于分组资料(或不同分类)的某个统计量取值,直方图一般用于描述连续型资料的分布情况。
5. 错。四分位数间距反映了 $P_{25} \sim P_{75}$ 范围内的数据离散情况,这个范围数据占数据总量的 50%。由此可见,如果小于 P_{25} 的数据发生变化不会对四分位数间距统计量的取值有影响;同样,大于 P_{75} 的数据发生变化也不会对四分位数间距统计量的取值有影响。所以,四分位数间距不易受两边极大、极小值的影响。
6. 对。定量资料和定性资料都有频数分布的问题,所以均可以用频数分布表进行统计描述。

二、选择题

1. 选 D。根据第一组段的下限为 6,组距为 8,得到:第一组段为 $[6, 14)$ 及其组中值为 10,第二组段为 $[14, 22)$ 及其组中值为 18,第三组段为 $[22, 30)$ 及其组中值为 26,所以 D 是错误的。
2. 选 A。因为身高与体重的量纲不同,标准差、方差和四分位数间距均含有量纲,因此选择不同的身高和体重的量纲,标准差、方差和四分位数间距的取值均不同,因此无法比较。只有采用相对离散程度的指标:变异系数,可以消除量纲,才具有可比性。
3. 选 B。标准差比较好地描述了对称分布资料离开其样本均数的离散程度,其他指标虽然也能描述其一般的离散程度,但不能刻画资料与其样本均数的离散程度。

4. 选 D。因为资料分布的纵坐标可以是频率、频数和频率密度。

5. 选 B。根据题意, 样本量 $n = 6$ 。数据排序后, 中间的两个数据为 3 和 4.5, 根据中位数计算公式, 因此中位数为这两个数的平均数 $(3 + 4.5)/2 = 3.75$ 。

三、简答题

1. 常用的描述集中趋势的指标有哪些, 并简述其适用条件。

集中趋势反映各数据向其中心值靠拢或聚集的程度, 反映一组数据中心点的位置。众数适用于描述定性资料的集中趋势, 一组数据中出现次数最多的变量值, 不受数据中极端值的影响。

算术均数适用于: 对称分布, 尤其是正态分布的资料。

中位数适用于: 非正态分布的资料, 一端或两端无确切切值或分布不清的资料。

几何均数适用于: 资料取对数变换后近似呈正态分布的。

2. 常用的统计图有哪几种? 它们各自适用于哪些资料?

根据资料的类型与分析的目的选择合适的统计图对资料进行描述, 归纳如表 2-1。

表 2-1 常用统计图的类型

统计图	资料性质	分析目的
普通线图	连续型资料	用线段的升降表示某事物在时间上的变化趋势, 或某一现象随着另一现象变化的情况
半对数线图	连续型资料	用线段的升降表示事物的相对变化速度
直条图	相互独立资料	用直条长短表示相互独立的各指标的数值大小, 一般用于比较不同组别的指标大小
直方图	连续型变量的频数分布	用矩形面积表示各组段的频数(或频率)
圆图及百分条图	构成比资料	用圆的扇形面积或直条各段的长度表示事物各组成部分的构成情况
散点图	双变量连续型资料	表示两种事物变化的相关性和趋势
箱式图	连续型资料	描述数据的分布特征(包括中位数、四分位数范围、最大值和最小值)

3. 请简述频数表的用途。

当收集到大量的数据时, 杂乱无章, 很难观察到数据中所蕴含的规律, 频数表可以很好地组织数据, 通过编制频数表, 可以揭示数据的分布类型是