



卫生部“十二五”规划教材配套教材

全国高等医药教材建设研究会规划教材配套教材

全国高等学校配套教材

供预防医学类专业用

# 卫生统计学

## 学习指导与习题集

第2版

主 编 方积乾 副主编 徐勇勇 陈峰



人民卫生出版社  
PEOPLE'S MEDICAL PUBLISHING HOUSE

卫生部“十二五”规划教材配套教材

全国高等医药教材建设研究会规划教材配套教材

全国高等学校配套教材

供预防医学类专业用

# 卫生统计学学习指导与习题集

第 2 版

主 编 方积乾

副主编 徐勇勇 陈 峰

编 者 (按姓氏笔画排序)

马 骏	天津医科大学	张菊英	四川大学
王 彤	山西医科大学	陈 峰	南京医科大学
王 静	安徽医科大学	易 东	第三军医大学
王心旺	广州医科大学	郝元涛	中山大学
王学梅	内蒙古医科大学	钟晓妮	重庆医科大学
方 亚	厦门大学	邵艳晖	广东药学院
方积乾	中山大学	施学忠	郑州大学
石武祥	广西桂林医学院	贺 佳	第二军医大学
刘美娜	哈尔滨医科大学	钱 聪	中国医科大学
宇传华	武汉大学	徐勇勇	第四军医大学
李晓梅	昆明医科大学	郭秀花	首都医科大学
杨士保	中南大学	郭忠琴	宁夏医科大学
吴少敏	中山大学	郭淑霞	石河子大学
余红梅	山西医科大学	陶育纯	吉林大学
张晋昕	中山大学	康晓平	北京大学

秘 书 吴少敏 中山大学

人民卫生出版社

图书在版编目(CIP)数据

卫生统计学学习指导与习题集 / 方积乾主编. —2 版. —北京: 人民卫生出版社, 2013

ISBN 978-7-117-18500-4

I. ①卫… II. ①方… III. ①卫生统计—医学院校—教学参考资料 IV. ①R195.1

中国版本图书馆 CIP 数据核字(2013)第 320112 号

人卫社官网	<a href="http://www.pmph.com">www.pmph.com</a>	出版物查询, 在线购书
人卫医学网	<a href="http://www.ipmph.com">www.ipmph.com</a>	医学考试辅导, 医学数据库服务, 医学教育资源, 大众健康资讯

版权所有, 侵权必究!

卫生统计学学习指导与习题集

第 2 版

主 编: 方积乾

出版发行: 人民卫生出版社(中继线 010-59780011)

地 址: 北京市朝阳区潘家园南里 19 号

邮 编: 100021

E - mail: [pmph@pmph.com](mailto:pmph@pmph.com)

购书热线: 010-59787592 010-59787584 010-65264830

印 刷: 北京人卫印刷厂

经 销: 新华书店

开 本: 787×1092 1/16 印张: 23

字 数: 560 千字

版 次: 2008 年 4 月第 1 版 2014 年 3 月第 2 版

2014 年 3 月第 2 版第 1 次印刷(总第 7 次印刷)

标准书号: ISBN 978-7-117-18500-4/R·18501

定价(含光盘): 39.00 元

打击盗版举报电话: 010-59787491 E-mail: [WQ@pmph.com](mailto:WQ@pmph.com)

(凡属印装质量问题请与本社市场营销中心联系退换)

本书是与卫生部“十二五”规划教材《卫生统计学》(第7版)配套的教学参考书,由《卫生统计学》(第7版)的作者同步编写。为便于学生主动地学习《卫生统计学》(第7版),深入理解基本理论、基本知识,扎实掌握基本技能,我们紧扣学生的需求,在第1版《卫生统计学学习指导与习题集》的基础上进行了如下修订:

1. 增加了“内容精要”并注明重点所在。从需要解决的问题出发,而不是从教材章节内容出发,以学生的视角加以归纳,帮助学生进行基于问题的学习(problem-based-learning, PBL)。

2. 将第1版以“教”为中心的“教学要求”(了解、熟悉、掌握)改为以“学”为中心的“学习目标”,并在“学习目标”中注明内容的“难点”。如通过第一章的学习,知道统计学和医学统计学的发展史,能简要说明统计学与公共卫生的关系,能阐述总体与样本的关系等,帮助学生自主学习和自我评价。

3. “案例讨论参考答案”与教科书每章的案例讨论相呼应,并提供了适当的提示,引导学生深入思考。建议读者在学习时不要立即放弃自己的想法,在深入讨论之后再参考这部分的内容。

4. “计算机实验程序及结果解释”较第1版有较大改动。SAS程序库、SAS文件的名称、SAS数据集名称和SAS程序的内容与格式(注释、变量说明、结果等)全部标准化。实验结果包括SAS数据的平面格式和计算结果的主要统计量,使学生明白:人需要给电脑提供什么信息,电脑才能给人提供相应的结果。

5. “思考与练习参考答案”与教科书每章后面的“思考与练习”一一对应。我们提供参考答案,旨在帮助读者自行检查与反思,也相应减轻了教师批阅作业负担。

6. 在第1版的基础上,我们对“补充练习题”部分的内容进行了修改和充实(包括选择题、是非题和案例讨论题),并提供了参考答案,旨在帮助读者进行自我测试。

为保证本书的质量,副主编徐勇勇教授为本书制订了统一的、标准化的格式和规范用语;副主编陈峰教授及其同事逐章审阅了正文;秘书吴少敏老师带领研究生们逐一运行了程序。在此,谨代表编委会向他们的辛勤劳动致以衷心的感谢。

尽管我们在编写的过程中付出了艰苦的努力,但书中仍有可能存在疏漏之处。我们将在使用过程中不断发现问题,使之进一步完善。

方积乾

2013年3月

# 目录

第一章 绪论 .....	1
【内容精要】.....	1
【学习目标】.....	2
【补充练习题】.....	2
【补充练习题参考答案】.....	4
第二章 定量变量的统计描述 .....	5
【内容精要】.....	5
【学习目标】.....	8
【案例讨论参考答案】.....	8
【计算机实验程序及结果解释】.....	9
【思考与练习参考答案】.....	15
【补充练习题】.....	20
【补充练习题参考答案】.....	23
第三章 定性变量的统计描述 .....	24
【内容精要】.....	24
【学习目标】.....	26
【案例讨论参考答案】.....	26
【计算机实验程序及结果解释】.....	27
【思考与练习参考答案】.....	36
【补充练习题】.....	38
【补充练习题参考答案】.....	41
第四章 常用概率分布 .....	42
【内容精要】.....	42
【学习目标】.....	45
【案例讨论参考答案】.....	45
【计算机实验程序及结果解释】.....	45
【思考与练习参考答案】.....	57
【补充练习题】.....	58
【补充练习题参考答案】.....	61

第五章 参数估计基础 .....	62
【内容精要】.....	62
【学习目标】.....	63
【案例讨论参考答案】.....	64
【计算机实验程序及结果解释】.....	64
【思考与练习参考答案】.....	75
【补充练习题】.....	76
【补充练习题参考答案】.....	78
第六章 假设检验基础 .....	79
【内容精要】.....	79
【学习目标】.....	81
【案例讨论参考答案】.....	81
【计算机实验程序及结果解释】.....	81
【思考与练习参考答案】.....	90
【补充练习题】.....	92
【补充练习题参考答案】.....	94
第七章 方差分析基础 .....	95
【内容精要】.....	95
【学习目标】.....	98
【案例讨论参考答案】.....	98
【计算机实验程序及结果解释】.....	99
【思考与练习参考答案】.....	109
【补充练习题】.....	113
【补充练习题参考答案】.....	115
第八章 $\chi^2$ 检验 .....	116
【内容精要】.....	116
【学习目标】.....	118
【案例讨论参考答案】.....	119
【计算机实验程序及结果解释】.....	119
【思考与练习参考答案】.....	124
【补充练习题】.....	125
【补充练习题参考答案】.....	126
第九章 基于秩次的非参数检验 .....	127
【内容精要】.....	127

【学习目标】	129
【案例讨论参考答案】	130
【计算机实验程序及结果解释】	131
【思考与练习参考答案】	143
【补充练习题】	148
【补充练习题参考答案】	151
<b>第十章 两变量关联性分析</b>	<b>152</b>
【内容精要】	152
【学习目标】	153
【案例讨论参考答案】	154
【计算机实验程序及结果解释】	154
【思考与练习参考答案】	163
【补充练习题】	164
【补充练习题参考答案】	167
<b>第十一章 简单线性回归</b>	<b>169</b>
【内容精要】	169
【学习目标】	172
【案例讨论参考答案】	172
【计算机实验程序及结果解释】	174
【思考与练习参考答案】	189
【补充练习题】	197
【补充练习题参考答案】	198
<b>第十二章 多重线性回归与相关</b>	<b>199</b>
【内容精要】	199
【学习目标】	201
【案例讨论参考答案】	201
【计算机实验程序及结果解释】	202
【思考与练习参考答案】	212
【补充练习题】	214
【补充练习题参考答案】	217
<b>第十三章 医学研究的统计学设计</b>	<b>218</b>
【内容精要】	218
【学习目标】	221
【案例讨论参考答案】	221
【计算机实验程序及结果解释】	222

【思考与练习参考答案】	227
【补充练习题】	229
【补充练习题参考答案】	231
<b>第十四章 实验研究的设计与分析</b>	233
【内容精要】	233
【学习目标】	234
【案例讨论参考答案】	234
【计算机实验程序及结果解释】	234
【思考与练习参考答案】	249
【补充练习题】	255
【补充练习题参考答案】	256
<b>第十五章 临床试验设计与分析</b>	257
【内容精要】	257
【学习目标】	258
【案例讨论参考答案】	258
【计算机实验程序及结果解释】	259
【思考与练习参考答案】	263
【补充练习题】	265
【补充练习题参考答案】	266
<b>第十六章 观察性研究的实施与分析</b>	267
【内容精要】	267
【学习目标】	271
【案例讨论参考答案】	271
【计算机实验程序及结果解释】	272
【思考与练习参考答案】	286
【补充练习题】	288
【补充练习题参考答案】	292
<b>第十七章 寿命表</b>	293
【内容精要】	293
【学习目标】	295
【案例讨论参考答案】	295
【计算机实验程序及结果解释】	295
【思考与练习参考答案】	305
【补充练习题】	309
【补充练习题参考答案】	310



<b>第十八章 logistic 回归分析</b> .....	311
【内容精要】 .....	311
【学习目标】 .....	312
【案例讨论参考答案】 .....	312
【计算机实验程序及结果解释】 .....	312
【思考与练习参考答案】 .....	315
【补充练习题】 .....	317
【补充练习题参考答案】 .....	317
<b>第十九章 生存分析</b> .....	318
【内容精要】 .....	318
【学习目标】 .....	319
【案例讨论参考答案】 .....	319
【计算机实验程序及结果解释】 .....	320
【思考与练习参考答案】 .....	332
【补充练习题】 .....	337
【补充练习题参考答案】 .....	340
<b>第二十章 meta 分析</b> .....	341
【内容精要】 .....	341
【学习目标】 .....	341
【案例讨论参考答案】 .....	342
【计算机实验程序及结果解释】 .....	343
【思考与练习参考答案】 .....	352
【补充练习题】 .....	354
【补充练习题参考答案】 .....	355

# 第一章

## 绪 论

### 【内容精要】

#### 一、统计学与公共卫生的关系

公共卫生的主要研究对象为人群,统计学技术适用于描述群体健康状况及疾病流行规律,评价卫生政策与卫生干预措施的效果。统计抽样技术适用于居民健康调查、疾病流行状况调查、卫生机构工作绩效调查,掌握人群的健康水平和卫生服务需求。统计描述适用于用群体统计指标反映疾病和卫生资源的分布特征。统计推断适用于通过样本数据推论居民健康水平、健康危险因素、卫生干预措施、治疗效果评价的总体特征。

#### 二、总体与样本(重点)

总体(population)是根据研究目的确定的同质研究对象的全体;按研究对象来源又有目标总体和研究总体。样本(sample)是指从研究总体中抽取的一部分有代表性的个体;获取样本的过程称为抽样(sampling)。抽样研究的目的是用样本数据推断总体的特征。

#### 三、同质与变异(重点)

同质(homogeneity)是指同一总体中个体的主要性质相同。变异(variation)是指同质的个体之间存在的差异。统计学的任务是在变异的背景上描述总体内的同质性,揭示不同总体间的异质性(heterogeneity)。

#### 四、变量的类型(重点)

变量  $\left\{ \begin{array}{l} \text{定性变量(qualitative variable)} \\ \text{定量变量(quantitative variable)} \end{array} \right. \left\{ \begin{array}{l} \text{分类变量(categorical variable)} \\ \text{有序变量(ordinal variable)} \end{array} \right.$

定量变量可分为离散型变量(discrete variable)和连续型变量(continuous variable)。

变量类型可以进行转化:定量→有序→分类→二值。应该注意的是,变量类型只能由信息量多的向信息量少的类型转化。

#### 五、参数与统计量(重点)

参数(parameter)是指反映总体特征的数量。由样本计算出来的反映样本特征的数量

称为样本统计量(statistic)。统计的任务是依据样本统计量来推断总体参数。

## 六、设计与分析

统计设计(design)是医药卫生科研设计不可或缺的部分。统计设计包括抽样(分组)方法、统计指标、实验设计、数据收集、样本量估计、统计分析方法等内容。统计设计和统计分析是不可分割的两项内容,设计决定分析的方法。

## 七、因果与联系

探究因果关系(causality)首先考察是否存在联系(association)。但存在联系未必有因果关系,因为存在大量的混杂因素。大多数观察性研究,单靠统计学分析只能考察变量之间的联系,难以证明因果关系。

### 【学习目标】

1. 知道统计学和医学统计学的发展史,说出对统计学和医学统计学做出杰出贡献的历史人物的姓名、国籍和主要贡献。
2. 简要说明统计学与公共卫生的关系。
3. 阐述总体与样本的关系。
4. 举例说明人群健康水平抽样研究及其推论的总体特征(难点)。
5. 阐述总体的同质性和异质性,简要说明如何构建同质总体(以某地7岁以下儿童的生长发育评价为例)。
6. 阐述变量的类型,简要说明定性变量和定量变量(以某地7岁以下儿童的生长发育评价为例)。
7. 阐述参数与统计量,简要说明如何用统计量估计总体参数(以某地7岁以下儿童的身高测量值为例)。
8. 简要说明干预性研究与观察性研究。
9. 简要说明统计设计与统计分析两者的关系。
10. 举例说明探究因果关系的研究(难点)。
11. 举例说明描述因素间关联性的研究。
12. 知道统计软件的作用和常用统计软件的名称。
13. 知道学习本课程的目标与方法。

### 【补充练习题】

#### 一、选择题

##### (一) A1型题

每题有A、B、C、D、E五个备选答案,请从中选择一个最佳答案。

1. 下面的变量中,属于分类变量的是\_\_\_\_\_。
  - A. 脉搏
  - B. 血型
  - C. 肺活量
  - D. 红细胞计数
  - E. 血压

2. 下面的变量中,属于定量变量的是\_\_\_\_\_。
- A. 性别                      B. 体重                      C. 血型  
D. 职业                      E. 民族
3. 某人记录了 50 名病人体重的测定结果,经整理表达为:小于 50kg 的 13 人,介于 50kg 和 70kg 间的 20 人,大于 70kg 的 17 人,这时,体重转化为\_\_\_\_\_。
- A. 定量变量                  B. 分类变量                  C. 有序变量  
D. 二值变量                  E. 名义变量
4. 必要时,体重还可以进一步表达为\_\_\_\_\_。
- A. 定量变量                  B. 分类变量                  C. 有序变量  
D. 二值变量                  E. 名义变量
5. 若要通过样本进行统计推断,样本应是\_\_\_\_\_。
- A. 总体中典型的一部分                      B. 总体中任一部分  
C. 总体中随机抽取的一部分                  D. 总体中选取的有意义的一部分  
E. 总体中信息明确的一部分
6. 统计量是\_\_\_\_\_。
- A. 由总体数据得到的量  
B. 反映总体统计特征的量  
C. 根据总体中的全部数据计算出的统计指标  
D. 用参数估计出来的  
E. 由样本数据计算出的统计指标
7. 因果关系\_\_\_\_\_。
- A. 就是变量间数量上的联系  
B. 可以用统计方法证明  
C. 必定表现为数量间的联系  
D. 可以通过单独考察两个变量间关系得出  
E. 可以通过变量间数量上的联系来证明

## (二) A2 型题

每题以一个小案例的形式出现,其下面有 A、B、C、D、E 五个备选答案,请从中选择一个最佳答案。

1. 教材中提及,美国人 1954 年实施了旨在评价索尔克(Salk)疫苗预防小儿麻痹或死于脊髓灰质炎效果的临床试验。有 180 万儿童参与,约有 1/4 参与者得到了随机化。这 180 万儿童是\_\_\_\_\_。
- A. 目标总体                  B. 研究总体                  C. 1 份样本  
D. 1 份随机样本              E. 180 万份样本
2. 上述试验最终认为索尔克疫苗是有效的。请问此结论是针对\_\_\_\_\_而言。
- A. 180 万儿童                      B. 每个儿童  
C. 所有使用索尔克疫苗的儿童                  D. 所有儿童和成人  
E. 180 万儿童中随机化的 1/4

## 二、是非题

1. 定量变量、分类变量和有序变量可以相互转换。 ( )
2. 假变量可以参与计算,所以假变量是定量变量。 ( )
3. 离散变量在数值很大时,单位为“千”或“万”时可以取小数值,此时可近似地视为连续型变量。 ( )
4. 同质的个体间不存在变异。 ( )
5. 如果个体间有变异,则它们一定不是来自同一总体。 ( )

### 【补充练习题参考答案】

#### 一、选择题

(一) 1. B 2. B 3. C 4. B 5. C 6. E 7. C

(二) 1. C 2. C

#### 二、是非题

1. × 2. × 3. √ 4. × 5. ×

(方积乾)

## 第二章

# 定量变量的统计描述

### 【内容精要】

#### 一、离散型定量变量和连续型定量变量的频率分布

离散型定量变量的取值是不连续的。直接清点各变量值及相同变量值出现的频数，并计算相应的频率，即为频率分布表。离散型定量变量的频率分布图可用直条图表达，以各等宽矩形直条的高度表示各组频率的大小。

连续型定量变量的取值是连续的。将数据适当分组，清点各组的频数，并计算相应的频率，即为频率分布表。连续型定量变量的频率分布图可用直方图表达，其纵坐标为频率密度，即频率/组距，直方图的面积之和等于1。

#### 二、频率分布表(图)的用途

频率分布表(图)可以揭示变量的分布类型，如对称分布或偏峰分布；也可以描述变量的平均水平和变异程度；便于发现某些特大和特小的可疑值；便于进一步计算和统计分析。

#### 三、描述平均水平的统计指标(重点)

对于连续型定量变量，平均数是应用最广泛、最重要的一类指标，它常用于描述变量数值的集中位置，反映其平均水平。常用的平均数有3种：算术均数、几何均数和中位数。

##### 1. 算术均数

(1) 意义：算术均数简称均数，常用 $\mu$ 表示总体均数， $\bar{X}$ 表示样本均数。

(2) 适用条件：适用于服从对称分布的变量，尤其是服从正态或近似正态分布的变量。

(3) 计算

A. 直接法(基于原始数据)

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n} = \frac{\sum X}{n}$$

其中 $n$ 为样本量， $X_1, X_2, \cdots, X_n$ 为观察值。

B. 频数表法(也叫加权法，基于频数表数据)

$$\bar{X} = \frac{\sum fx_0}{\sum f} = \frac{\sum fx_0}{n}$$

其中， $f$ 为组段的频数， $x_0$ 为各组段的中值， $x_0 = (\text{组段上限} + \text{组段下限})/2$ 。

## 2. 几何均数

(1) 意义: 几何均数以符号  $G$  表示, 常用来反映一组含多个数量级数据的集中位置。

(2) 适用条件: 适用于原始观察值分布不对称, 但经对数转换后呈对称分布的变量, 如对数正态分布变量。观察值间常呈倍数关系, 或变化范围跨越多个数量级。

## (3) 计算

## A. 直接法(基于原始数据)

$$G = \sqrt[n]{X_1 X_2 \cdots X_n} \quad \text{或} \quad G = \lg^{-1} \left[ \frac{\sum \lg X}{n} \right]$$

## B. 频数表法(也称加权法, 基于频数表数据)

$$G = \lg^{-1} \left[ \frac{\sum f \lg X}{\sum f} \right] = \lg^{-1} \left[ \frac{\sum f \lg X}{n} \right]$$

## 3. 中位数

(1) 意义: 中位数常用符号  $M$  表示, 是指将原始观察值从小到大排序后, 位次居中的那个数。理论上有一半的观察值低于中位数, 一半的观察值高于中位数。

(2) 适用条件: 中位数适用于服从各种分布的变量, 特别是偏峰分布的变量、分布末端无确定值的变量等。

## (3) 计算

A. 直接法(基于原始数据): 将  $n$  例观察值从小到大排列, 第  $i$  个数据用  $X_i^*$  表示。

$n$  为奇数时,

$$M = X_{\frac{n+1}{2}}^*$$

$n$  为偶数时,

$$M = \frac{1}{2} (X_{\frac{n}{2}}^* + X_{\frac{n}{2}+1}^*)$$

B. 百分位数法(基于频数表数据): 百分位数常用符号  $P_x$  表示, 是排序后的全部实测值的某百等份分割值, 即在全部实测值中有  $x\%$  个体观察值比它小, 有  $1-x\%$  个体观察值比它大。中位数就是一个特定的百分位数, 即  $M = P_{50}$ 。

$$P_x = L + \frac{i}{f_x} \left( \frac{x \cdot n}{100} - F_L \right)$$

其中,  $L$  为欲求的  $P_x$  所在组段的下限,  $i$  为该组段的组距,  $f_x$  为该组段内的频数,  $n$  为总频数,  $F_L$  为该组段之前的累积频数。

## 四、描述变异程度的统计指标(重点)

为比较全面地把握资料的分布特征, 不仅需要了解数据的集中位置, 而且需要了解数据的变异程度。常用的描述变异程度的统计量包括极差、四分位数间距、方差、标准差和变

异系数。

### 1. 极差

(1) 意义: 极差 ( $R$ ) 也称全距。样本量接近的同类资料比较时, 极差越大意味着数据间变异越大。极差只考虑数据的最大值和最小值, 所以用极差反映数据的变异程度常常比较粗略和不稳定。

(2) 适用条件: 变量分布类型不限。

(3) 计算:  $R = \text{最大值} - \text{最小值}$

### 2. 四分位数间距

(1) 意义: 四分位数间距 ( $Q$ ) 表示百分位数  $P_{75}$  和百分位数  $P_{25}$  之差。理论上  $P_{25}$  与  $P_{75}$  之间恰好包括总体中间 50% 的个体观察值。同类资料比较,  $Q$  越大意味着数据间变异越大。对同一资料, 用四分位数间距反映数据的变异程度比极差稳定。

(2) 适用条件: 四分位数间距可用于服从各种分布的变量, 特别是偏峰分布变量, 常把中位数和四分位数间距结合起来描述变量的平均水平和变异程度。

(3) 计算:  $Q = P_{75} - P_{25}$

其中,  $P_{75}$  和  $P_{25}$  的求法参见前述百分位数求法。

### 3. 方差

(1) 意义: 方差考虑了全部观察值的变异程度。总体方差用  $\sigma^2$  表示, 定义为总离均差平方和  $\sum(X - \mu)^2$  在总体的所有个体中的平均; 样本方差用  $S^2$  表示, 是总体方差的无偏估计。同类变量比较时, 方差越大意味着数据间变异度越大。

(2) 适用条件: 见标准差。

(3) 计算:  $S^2 = \frac{\sum(X - \bar{X})^2}{n-1}$

其中,  $n-1$  称为自由度。

### 4. 标准差

(1) 意义: 方差的平方根称为标准差。总体标准差用  $\sigma$  表示, 样本标准差用  $s$  表示。标准差的量纲与原变量一致, 故实际应用中常使用标准差。同类变量比较时, 标准差越大意味着观察值间变异程度越大。

(2) 适用条件: 方差和标准差都适用于服从对称分布的变量, 特别对正态分布或近似正态分布的变量, 常把均数和标准差结合起来, 全面描述变量的平均水平和变异程度。

(3) 计算:  $S = \sqrt{\frac{\sum(X - \bar{X})^2}{n-1}}$

### 5. 变异系数

(1) 意义: 变异系数用  $CV$  表示, 为标准差与算术均数之比, 是一个不带量纲的相对数。

(2) 适用条件: 适用于量纲不同或均数差别较大的变量间变异程度的比较。

(3) 计算:  $CV = \frac{S}{\bar{X}} \times 100\%$

## 五、描述分布形态的统计指标

偏度系数和峰度系数可描述变量的分布形态。理论上总体偏度系数为 0 时, 分布是对



称的；取正值时的分布为正偏峰；取负值时的分布为负偏峰。理论上正态分布的总体峰度系数为0；取负值时，其分布较正态分布的峰平阔；取正值时，其分布较正态分布的峰尖峭。

## 六、统计图和统计表(重点)

统计图和统计表是呈现统计分析结果的重要工具。统计表一般由表号及标题、标目、线条、数字和备注组成。其中表号及标题位于表上方正中，需简明扼要；标目包括横标目和纵标目，分别说明表中横行和纵列的含义；线条常用“三线表”；数字要以个位对齐。绘制统计表时要注意内容清晰、层次分明、列表规范。

统计图是用点的位置、线段的升降、直条的长短、面积的大小等表达统计数据的一种形式。绘制统计图要根据资料性质和分析目的选择最合适的图形。常用于定量变量的统计图包括频率直方图、累积频率分布图、箱式图和直条图等。

频率直方图用于描述连续型定量变量的频率分布。

累积频率分布图用于描述定量变量的累积频率分布。

箱式图用最小值、下四分位数、中位数、上四分位数、最大值描述定量变量的平均水平和变异程度。

直条图用等宽直条的高度表示相互独立的各项指标数量大小，可描述离散型定量变量和定性变量的频率分布。

### 【学习目标】

1. 知道频率分布表和频率分布图的制作方法。
2. 阐述并举例说明描述定量变量平均水平的常用指标及其意义、计算方法和适用条件。
3. 阐述并举例说明描述定量变量变异程度的常用指标及其意义、计算方法和适用条件。
4. 举例说明百分位数的意义、计算方法和应用。
5. 知道常用的统计图与统计表及绘制方法。
6. 运用统计指标、统计表和统计图综合描述定量变量的平均水平和变异程度(难点)。
7. 知道分布形态的偏度系数与峰度系数的计算方法和意义。

### 【案例讨论参考答案】

**案例 2-1** 该资料为某企业 238 名无汞作业接触史工人的发汞含量( $\mu\text{mol/kg}$ )检测结果，已整理成频率分布表(见教材表 2-9)。统计描述时首先应考察变量的分布规律，通过频率分布表和相应的直方图(见教材图 2-3)可以看出，此 238 人发汞含量的频率分布呈正偏峰分布，即观察值绝大多数集中在发汞含量较小的组段。对偏峰分布，选用算术均数和标准差进行统计描述是不恰当的，应选用中位数和下、上四分位数描述该企业工人发汞含量的平均水平和变异程度。计算如下：

$$P_{50} = L + \frac{i}{f_{50}} \left( \frac{50 \times n}{100} - F_L \right) = 5.5 + \frac{2}{60} \left( \frac{50 \times 238}{100} - 86 \right) = 6.6 (\mu\text{mol/kg})$$

$$P_{25} = L + \frac{i}{f_{25}} \left( \frac{25 \times n}{100} - F_L \right) = 3.5 + \frac{2}{66} \left( \frac{25 \times 238}{100} - 20 \right) = 4.7 (\mu\text{mol/kg})$$