

21世纪高等学校经济管理系列教材

# 统计学学习指导 及能力提升

●●● TONGJIXUE XUEXI ZHIDAO JI NENGLI TISHENG

王磊 曾霞 主编



中国财经出版传媒集团  
中国财政经济出版社

21世纪高等学校经济管理系列教材

# 统计学学习指导 及能力提升

●●●TONGJIXUE XUEXI ZHIDAO JI NENGLI TISHENG

王 磊 曾 霞 主 编



中国财经出版传媒集团  
中国财政经济出版社

· 北 京 ·

图书在版编目 (CIP) 数据

统计学学习指导及能力提升 / 王磊, 曾霞主编.

北京: 中国财政经济出版社, 2025. 1. -- ISBN 978 - 7

- 5223 - 3662 - 6

I. C8

中国国家版本馆 CIP 数据核字第 2025KV9366 号

责任编辑: 张怡然

责任校对: 胡永立

封面设计: 易 帅

责任印制: 张 健

统计学学习指导及能力提升

TONGJIXUE XUEXI ZHIDAO JI NENGLI TISHENG

中国财政经济出版社 出版

URL: <http://www.cfeph.cn>

E-mail: [cfeph@cfeph.cn](mailto:cfeph@cfeph.cn)

(版权所有 翻印必究)

社址: 北京市海淀区阜成路甲 28 号 邮政编码: 100142

营销中心电话: 010 - 88191522

天猫网店: 中国财政经济出版社旗舰店

网址: <https://zgczjcbbs.tmall.com>

三河市宏图印务有限公司印刷 各地新华书店经销

成品尺寸: 185mm × 260mm 16 开 11.25 印张 240 000 字

2025 年 1 月第 1 版 2025 年 1 月河北第 1 次印刷

定价: 30.00 元

ISBN 978 - 7 - 5223 - 3662 - 6

(图书出现印装问题, 本社负责调换, 电话: 010 - 88190548)

本社图书质量投诉电话: 010 - 88190744

打击盗版举报热线: 010 - 88191661 QQ: 2242791300

# 本书编委会

主 编

王 磊 曾 霞

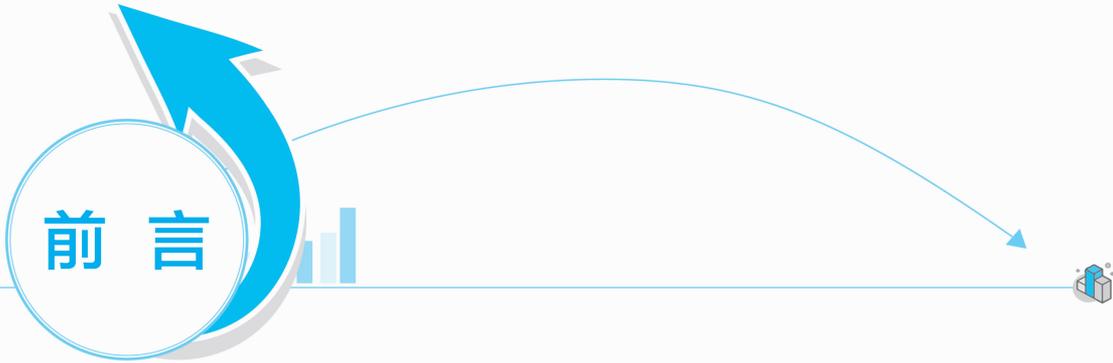
编 委

(按姓氏笔画排序)

朱冬辉 李文君 张 秀

赵 琼 耿智琳 蔡艳丽





# 前言

教育是国之大计、党之大计。在我国高等教育改革不断深化的今天，编写一套符合新时代要求、具有鲜明时代特色的统计学习题课教程，是有必要的。其不仅是统计学教学的重要组成部分，也对巩固学生统计学知识、提升统计分析能力具有重要意义。

为帮助读者理解基本概念、掌握基本知识、学会基本解题方法与解题技巧，进而提高数据分析能力，我们根据多年的教学经验编写了本书。本书秉持“以学生为中心，以实用为导向”的原则，紧密结合我国社会经济发展和新时代奋斗目标，充分体现统计学在培养高素质人才中的重要作用。以下是本书的主要特点：

一是紧密结合教材，系统梳理统计学知识点。本书针对《统计学》教材中的重要概念、理论和方法，设计了大量具有代表性的习题，帮助学生巩固所学知识，提高解题能力。

二是融入思想政治教育，全面落实立德树人根本任务。本书在编写过程中，充分挖掘统计学中的思政元素，通过习题引导学生树立正确的世界观、人生观和价值观。

三是提供丰富教学资源，助力教师教学和学生自学。本书配备了详细的解答和解析，方便教师备课和学生自学。同时，借助现代信息技术，为学生提供在线学习资源，实现线上线下相结合的教学模式。

本书与配套教材一致，包括12章内容，每章包括6个方面的内容：本章学习目标、思维导图、基本知识点、基础练习、答案解析和统计故事，最后附有两套综合测试题和参考答案。

本章学习目标：对每一章的内容明确提出了若干不同层次的要求，有助于正确把握各章重点，合理安排学习计划。

思维导图：以图形的方式展示各章的相关知识点，清晰地组织和表达知识点之间的逻辑关系，有利于读者快速查找关键信息、准确把握重点。

基本知识点：清晰简要地梳理了各章基本概念、基本公式、相互关系等知识点，有利于系统地复习巩固基本知识。

基础练习：根据每章的学习目标要求，本着“少而精”和“学以致用”的原则，运用单项选择题、多项选择题、判断题、简答题、计算题和案例讨论题六种题型加深



知识学习。

答案解析：对每章各种类型的练习题都给出了答案，并且针对部分计算题给出了详细解析的视频，有助于学生掌握规范的解答要求，正确把握解题思路。

统计故事：对我国统计的发展历程、辛普森悖论等经典统计学知识等进行讲解，让学生了解我国统计工作的历史沿革和取得的成就，从而增强学生的文化自信和国家认同感，并且激发学生的学习兴趣，使抽象的统计学原理变得具体、形象，从而提高教学效果。

本书由王磊、曾霞主编，参加编写的人员有：朱冬辉，第1章和第2章；耿智琳，第3章和第4章；曾霞，第5章、第6章和第7章；王磊，第8章和第9章；李文君，第10章和模拟试卷；赵琼，第11章和第12章。另外，张秀、蔡艳丽、李俊飞、彭雨婷等参与了复核检查工作。

由于时间仓促及编者水平有限，书中错漏和不妥之处在所难免，恳请专家和读者批评指正，对此我们表示衷心的感谢！

主编：王磊

2025年1月

# 目 录

第 1 章 绪论 .....	1
第 2 章 统计数据的搜集 .....	13
第 3 章 数据的整理与展示 .....	26
第 4 章 数据分布特征的度量 .....	38
第 5 章 抽样分布理论 .....	50
第 6 章 参数估计 .....	58
第 7 章 假设检验 .....	74
第 8 章 分类数据分析 .....	88
第 9 章 方差分析 .....	100
第 10 章 相关分析与回归分析 .....	117
第 11 章 时间序列分析 .....	135
第 12 章 指数 .....	146
模拟试卷一 .....	157
模拟试卷二 .....	164
参考文献 .....	171

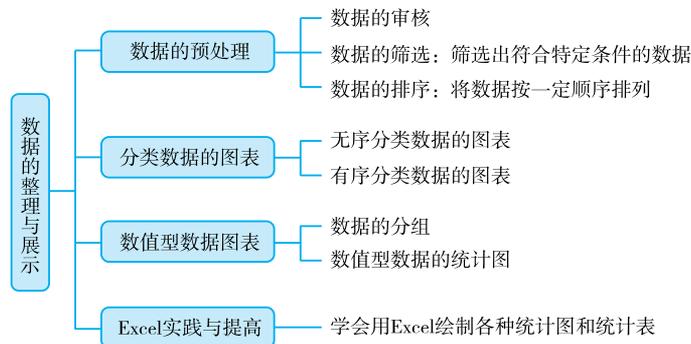


# 第3章 数据的整理与展示

## 一、本章学习目标

1. 理解数据预处理的重要性，了解数据预处理的内容。
2. 掌握分类数据的统计图和统计表特征及使用场景。
3. 掌握数值型数据的统计图和统计表特征及使用场景。
4. 学会使用 Excel 制作我国宏观经济数据的图表，展示中国特色社会主义建设的成就，展示经济发展成果和人民生活的变化。

## 二、思维导图



思维导图  
高清大图

## 三、基本知识点

章节	学习要点
3.1 数据的预处理	数据的审核：完整性、准确性、适用性、时效性
	数据的筛选、数据的排序
3.2 分类数据的图表	无序分类数据的图表：频数分布表、条形图、饼图和环形图
	有序分类数据的图表：累积频数分布表、累积频数分布图
3.3 数值型数据图表	数据的分组：单项式分组、组距分组
	数值型数据的图表：直方图、茎叶图、箱线图、散点图、雷达图、折线图
3.4 Excel 实践与提高	通过 Excel 实现统计图表的绘制



本章共有四个部分。

第一部分介绍数据的预处理。数据的预处理包括两部分内容：①数据的审核，具体可分为完整性审核、准确性审核、适用性审核、时效性审核。完整性审核主要是检查应调查的单位或个体是否缺失，所有的调查项目或指标是否填写齐全等。准确性审核主要是检查数据是否有错误，是否存在异常值等。数据准确性审核的方法主要有逻辑检查和计算检查。适用性审核主要针对二手数据，使用者应弄清楚数据的来源、目标总体的界定、数据的解释以及有关的背景资料等，确定这些数据是否与当前分析研究相匹配，是否需要重新加工整理等。时效性审核主要针对二手数据，注意数据的搜集时间是否在有效期限内。②数据的筛选和排序，其中数据筛选是指筛选出研究需要的、具有某些特征的数据。数据排序是将数据按一定顺序排列，以便研究者发现数据的特征或趋势。

第二部分介绍分类数据适合的统计图表。值得注意的是，适用于低层次计量尺度数据的统计图表同样适用于高层次计量尺度数据。但是，适用于高层次计量尺度数据的统计图表不一定适用于低计量尺度数据。

适用于无序分类数据的统计图表包括：①频数分布表。把各个类别及其频数用表格形式列出来，称为频数分布表。用 Excel 生成频数分布表的方法中，最常用的是使用数据透视表。统计表一般包括表头、行标题、列标题、数据、横线、注释等部分。②条形图，也称柱状图，用来展示分类变量数据，每一类别的数据用一个条形来表示。条形图包括简单条形图和复式条形图。③饼图和环形图。饼图侧重研究一个总体或样本结构性问题，每一类别的数据用一个扇形来表示。环形图可以比较多个总体或样本之间的结构性问题。

适用于有序分类数据的统计图表包括：①累积频数分布表。将有序数据中各类别的频数按顺序逐级累加，并把各个类别及其累加频数用表格形式列出来，称为累积频数分布表。具体可分为向上累积频数分布表和向下累积频数分布表。②累积频数分布图。

第三部分介绍数值型数据适合的统计图表。对于数值型数据，在绘制统计图表之前，往往需要进行统计分组。数据分组是指根据研究目的，按照某个标准把原始数据分成若干组，主要作用是反映数据的分布特征，可以分为单变量值分组和组距分组。单变量值分组是把一个变量值作为一个组，适合于变量值变化幅度不大的离散变量。组距分组是将数据分成若干个区间，一个区间作为一个组，适合于变量值变化幅度较大的离散变量或连续变量。组距分组中，一个组的最小值称为下限，一个组的最大值称为上限。一个组的组中值 = (上限 + 下限) ÷ 2，通常组中值作为该组数据的代表值。一个组的组距等于该组的上限与下限的差。通常，开口组以相邻组的组距作为其组距。组距分组需要遵循“不重不漏”原则及“计入下限组”原则。组距分组的步骤：确定组数；确定组距；计算每一组的频数，并绘制频数分布表。适用于数值型数据的统计图包括：直方图、茎叶图、箱线图、散点图、雷达图、折线图。

第四部分介绍如何用 Excel 实现绘制常用统计图表，要求学会通过 Excel 操作实现多种统计图表的绘制。



## 四、基础练习

### (一) 单项选择题

- 落在某一类别中的数据个数称为 ( )。  
A. 频数            B. 频率            C. 频数分布表    D. 累积频数
- 将数据分成若干组, 一个变量值作为一组的分组方式称作 ( )。  
A. 单项式分组    B. 组距分组        C. 等距分组        D. 异距分组
- 将数据分成若干组, 一个区间作为一组的分组方式称作 ( )。  
A. 单项式分组    B. 组距分组        C. 等距分组        D. 异距分组
- 组距分组中, 一个组的组中值是 ( )。  
A. 一个组的上限与下限之差    B. 一个组的上限与下限之间的中点值  
C. 一个组的最小值                D. 一个组的最大值
- 将某班级学生月平均生活费依次分为 1 500 元以下、1 500 ~ 2 000 元、2 000 ~ 2 500 元、2 500 元以上四组。第一组的组中值可近似为 ( )。  
A. 750            B. 1 000            C. 1 250            D. 1 500
- 将某班级学生月平均生活费依次分为 1 500 元以下、1 500 ~ 2 000 元、2 000 ~ 2 500 元、2 500 元以上四组。最后一组的组中值可近似为 ( )。  
A. 2 500            B. 2 750            C. 3 000            D. 3 250
- 当需要展示分类数据时, 不可以选用 ( )。  
A. 条形图            B. 饼图            C. 环形图            D. 直方图
- 下列哪个图形可以保留原始数据信息 ( )。  
A. 直方图            B. 条形图            C. 箱线图            D. 茎叶图
- 当侧重说明一个总体或样本的结构性问题时, 应该选用 ( )。  
A. 条形图            B. 饼图            C. 环形图            D. 直方图
- 当侧重比较两个或多个总体或样本的结构性问题时, 应该选用 ( )。  
A. 条形图            B. 饼图            C. 环形图            D. 直方图
- 欲表示某地某年某种流行病患者的年龄分布, 下列 ( ) 最合适。  
A. 条形图            B. 饼图            C. 线图            D. 直方图
- 大批量数值型数据的分布特征通常可以由 ( ) 反映。  
A. 条形图            B. 饼图            C. 直方图            D. 茎叶图
- 小批量数值型数据的分布特征通常可以由 ( ) 反映。  
A. 条形图            B. 饼图            C. 直方图            D. 茎叶图
- 比较多组数值型数据分布特征可以选用 ( )。  
A. 条形图            B. 直方图            C. 茎叶图            D. 箱线图
- 学习时长 (单位: 分钟) 与学习成绩 (百分制) 之间关系可以选用 ( ) 描述。  
A. 蛛网图            B. 直方图            C. 箱形图            D. 散点图



16. 多样本多变量之间关系可以选用 ( ) 描述。  
 A. 蛛网图      B. 直方图      C. 箱形图      D. 散点图
17. 时间序列数据的变化趋势可以选用 ( ) 描述。  
 A. 蛛网图      B. 直方图      C. 折线图      D. 散点图
18. 下列制作统计图的注意事项正确的是 ( )。  
 A. 标题写在图的上方      B. 横轴和纵轴应有标题  
 C. 横轴和纵轴一般不注明单位      D. 纵轴尺度起点必须是零
19. 欲说明 10 月以来某地区二手房租金每天的变化, 下列 ( ) 最合适。  
 A. 条形图      B. 饼图      C. 折线图      D. 直方图
20. 箱线图中, ( ) 的大小反映该组数据平均水平的高低。  
 A. 最值      B. 中位数      C. 算术平均值      D. 四分位数
21. 箱线图中, 箱体的高度由 ( ) 决定, 在一定程度上反映数据的波动程度。  
 A. 最值      B. 中位数      C. 算术平均值      D. 四分位数
22. 制作统计表时需注意 ( )。  
 A. 标题写在表的下方      B. 表中只可以有顶线和底线  
 C. 表中无数字可用“—”表示      D. 行标题和列标题的位置不能互换

### (二) 多项选择题

1. 数据的预处理包括 ( )。  
 A. 数据审核      B. 数据筛选      C. 数据排序      D. 数据展示  
 E. 数据分析
2. 数据的审核包括 ( )。  
 A. 完整性审核      B. 准确性审核      C. 适用性审核      D. 时效性审核  
 E. 数据量审核
3. 分类数据包括 ( )。  
 A. 无序分类数据      B. 有序分类数据      C. 数值型数据      D. 截面数据  
 E. 时序数据
4. 统计表一般包括 ( )。  
 A. 表头      B. 行列标题      C. 数据      D. 横线  
 E. 注释
5. 统计表表头一般包括 ( )。  
 A. 表的编号      B. 总标题  
 C. 行列标题      D. 表中数据计量单位  
 E. 注释
6. 统计表总标题应满足“3W”原则, “3W”是指表中数据的 ( )。  
 A. 时间      B. 地点      C. 内容      D. 来源  
 E. 获取方法
7. 组距数据中, 组中值是 ( )。



- A. 该组上限和下限的均值
  - B. 开口组没有组中值
  - C. 可以根据相邻组的组距计算开口组的组中值
  - D. 该组数据的一个代表值
  - E. 当各组数据在本组内呈现均匀分布或在组中值两侧呈对称分布时，组中值的代表性好；否则，代表性较差
8. 采用组距分组时，需要遵循（ ）原则。
- A. 不重不漏
  - B. 计入上限组
  - C. 计入下限组
  - D. 上限不在内
  - E. 下限不在内
9. 组距分组数据中，（ ）因素影响各组频数分布。
- A. 数据的个数
  - B. 组距
  - C. 组数
  - D. 变量值的大小
  - E. 组中值
10. 数据分组采用等距分组或异距分组取决于（ ）。
- A. 研究目的
  - B. 数据分布特征
  - C. 组数
  - D. 变量值的大小
  - E. 组中值
11. （ ）可以反映数据的分布情况。
- A. 直方图
  - B. 茎叶图
  - C. 箱线图
  - D. 折线图
  - E. 条形图
12. （ ）可以反映原始数值型数据的分布情况。
- A. 直方图
  - B. 茎叶图
  - C. 箱线图
  - D. 折线图
  - E. 条形图
13. 箱线图一般由一组数据的（ ）绘制而成。
- A. 最值
  - B. 中位数
  - C. 算术平均数
  - D. 众数
  - E. 四分位数

### (三) 判断题

- 1. 数据的完整性审核主要是检查应调查的单位或个体是否缺失，所有的调查项目或指标是否填写齐全等。（ ）
- 2. 当数据量较大时，如果某条记录存在信息缺失，尤其存在较多关键信息的缺失，则可以将该条记录删除。（ ）
- 3. 当数据量较少或缺失数据的记录较多时，则可以将该条记录删除。（ ）
- 4. 当数据量较少或缺失数据的记录较多时，则需要填补缺失数据。（ ）
- 5. 检查数据是否有错误属于数据准确性审核。（ ）
- 6. 检查数据是否存在异常值属于数据适用性审核。（ ）
- 7. 检查数据是否存在异常值属于数据准确性审核。（ ）
- 8. 不管异常值是否属于记录错误，都应该给予纠正或删除。（ ）
- 9. 如果判断出异常值是正确的数值，则应该给予保留并分析造成异常的原因。（ ）
- 10. 二手数据的优点是收集速度快、成本低，因此二手数据不需要做过多的预处理。



理。( )

11. 二手数据的适用性审核是指在使用二手数据时需要弄清楚数据的来源、目标总体的界定、数据的解释以及有关的背景资料等信息, 确定这些数据是否与当前研究相匹配, 是否需要重新加工整理等事项, 不能盲目使用。( )

12. 数据透视表和列联表都属于频数分布表。( )

13. 统计表的表头位于统计表下方居中位置。( )

14. 统计表的左右两边必须封口。( )

15. 箱线图和茎叶图都可以反映原始数值型数据的分布。( )

16. 与箱线图相比, 茎叶图更多用于多组数据分布特征的比较。( )

17. 单项式分组适合于所有的离散变量数据。( )

18. 单项式分组只适合于离散变量, 且变量值变化幅度不大。( )

19. 变量值变化幅度较大的离散变量或连续变量数据适合采用组距分组方法。( )

20. 数据分组时, 组数越少, 丧失的信息越少。( )

#### (四) 简答题

1. 组距分组的步骤是什么?

2. 简述直方图和条形图的异同。

3. 简述直方图和茎叶图的异同。

#### (五) 操作题

1. 为调查城镇居民参与户外运动的情况, 随机抽取容量为 48 人的一个样本。每星期户外运动参与频率分别为: 1 次、2 次、3 次。调查结果如表 3-1 所示。

表 3-1

每星期户外运动参与频率调查情况

单位: 次

2	2	2	1	3	2	1	1	3	3	2	2	1	1	2	2
3	1	2	1	1	3	2	1	3	3	2	3	3	1	3	1
2	1	2	2	2	1	2	1	1	2	2	1	3	1	3	1

(1) 用 Excel 制作频数分布表。

(2) 绘制条形图, 反映参与频率的分布。

2. 为确定某零件的使用寿命, 随机抽取容量为 100 个的一个样本进行测试。测试结果如表 3-2 所示。

表 3-2

某零件使用寿命调查情况

单位: 小时

622	571	570	619	580	709	705	578	684	594
639	627	710	698	595	639	672	686	580	589
640	666	631	646	696	598	655	605	593	581
675	671	589	698	699	585	615	665	636	692
631	651	597	631	575	633	661	680	623	624



续表

618	638	582	669	674	603	603	640	603	644
578	627	659	578	599	585	614	709	626	639
643	577	586	593	587	588	601	615	612	609
604	692	607	625	629	699	664	622	607	680
599	671	585	629	627	588	624	582	671	629

(1) 以 10 为组距进行分组，制作频数分布表。

(2) 绘制直方图，反映使用寿命的分布。

3. 为确定身高（单位：厘米）和体重（单位：千克）之间的关系，随机抽取容量为 25 人的一个样本进行调查（见表 3-3）。请根据调查结果绘制散点图，反映身高和体重之间的关系。

表 3-3 身高与体重的调查数据

身高（厘米）	178.1	190.8	172.3	173.8	164.2	159.5	155	162.3	155.3
体重（千克）	53.9	65.5	50.5	72.5	56.6	58.5	64.9	63.2	65.1
身高（厘米）	166.9	169.9	180.9	179.2	189	155.5	155.1	156.3	179.1
体重（千克）	72.4	69.3	72	73.9	89.3	41.1	43.3	66	67.4
身高（厘米）	171.6	176.3	189.2	157.5	181.9	157.9	154.6	—	—
体重（千克）	67.7	80.8	89.5	57.1	79.4	54.9	55	—	—

4. 根据 2010—2018 年湖北省居民消费水平数据（见表 3-4，数据来源：湖北省统计局网站），绘制折线图。

表 3-4 2010—2018 年湖北省居民消费水平数据

年份（年）	农村居民消费水平（元）	城镇居民消费水平（元）
2010	4 758	13 576
2011	5 653	15 935
2012	6 705	17 296
2013	7 755	19 156
2014	8 788	21 355
2015	9 542	23 561
2016	10 860	25 703
2017	13 090	28 121
2018	16 296	31 018



## 五、答案解析

### (一) 单项选择题

1. A 2. A 3. B 4. B 5. C 6. B 7. D 8. D 9. B 10. C 11. D 12. C 13. D  
14. D 15. D 16. A 17. C 18. B 19. C 20. B 21. D 22. C

### (二) 多项选择题

1. ABC 2. ABCD 3. AB 4. ABCDE 5. ABD 6. ABC 7. ACDE 8. ACD  
9. ABC 10. AB 11. ABCE 12. BC 13. ABE

### (三) 判断题

1. 对 2. 对 3. 错 4. 对 5. 对 6. 错 7. 对 8. 错 9. 对 10. 错 11. 对  
12. 对 13. 错 14. 错 15. 对 16. 错 17. 错 18. 对 19. 对 20. 错

### (四) 简答题

1. 答：组距分组的步骤具体如下：

第一步：确定组数。数据分组的作用是反映数据特征，所以组数太少或太多都不能有效显示数据的分布特征和规律。一般而言，组数不应少于5组，也不应多于15组。实际问题中，需要根据数据的数量和特征来确定组数。

第二步：确定组距。组距可以由组数来计算，具体来说， $\text{组距} = (\text{全部数据的最大值} - \text{最小值}) \div \text{组数}$ 。同时，为了计算方便，组距通常取5或10的倍数，且第一组的下限小于最小值，最后一组的上限大于最大值。

第三步：计算每一组的频数，并绘制频数分布表。

2. 答：相同点：条形的高度或长度表示各组数据频数的多少。

不同点：直方图中条形的宽度表示各组的组距，条形图中条形的宽度没有实际含义；直方图的各条形连续排列没有间隔，条形图的各条形之间有间隔；直方图用来展示数值型数据，条形图用来展示分类数据。

3. 答：相同点：两种统计图都可以反映数值型数据的分布情况。

不同点：直方图用于反映数值型分组数据的分布，茎叶图用于反映数值型原始数据的分布；直方图不能保留原始数据信息，茎叶图能够保留原始数据信息；直方图适用于大批量的数据，茎叶图适用于小批量的数据。

### (五) 操作题

1. (1) 制作频数分布表如表3-5所示。

表3-5

频数分布表

计数项：户外运动参与频率	汇总
1	18
2	18
3	12
总计	48



(2) 绘制条形图如图 3-1 所示。

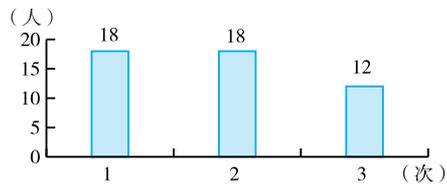


图 3-1 条形图



3.1 习题讲解

2. (1) 制作频数分布表如表 3-6 所示。

表 3-6 频数分布表

使用寿命 (小时)	数量 (个)	使用寿命 (小时)	数量 (个)
570 ~ 580	7	640 ~ 650	5
580 ~ 590	14	650 ~ 660	3
590 ~ 600	8	660 ~ 670	5
600 ~ 610	9	670 ~ 680	6
610 ~ 620	6	680 ~ 690	4
620 ~ 630	13	690 ~ 700	7
630 ~ 640	9	700 ~ 710	4
合计		100	

(2) 直方图如图 3-2 所示。

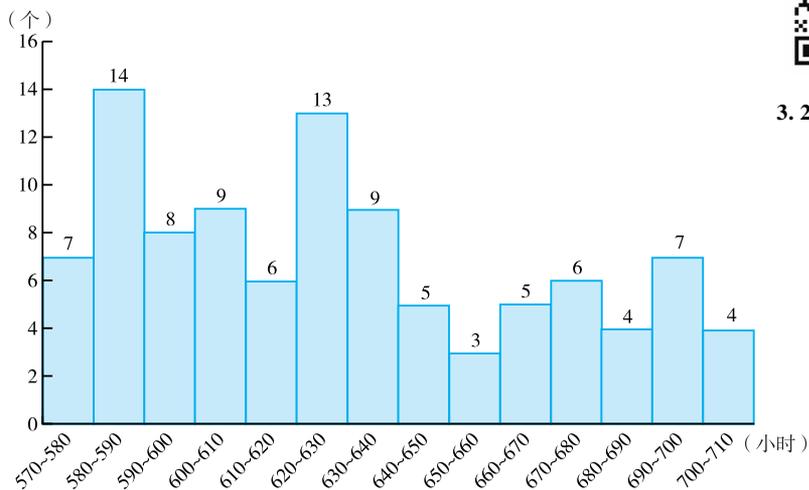


图 3-2 直方图



3.2 习题讲解



3. 散点图如图 3-3 所示。

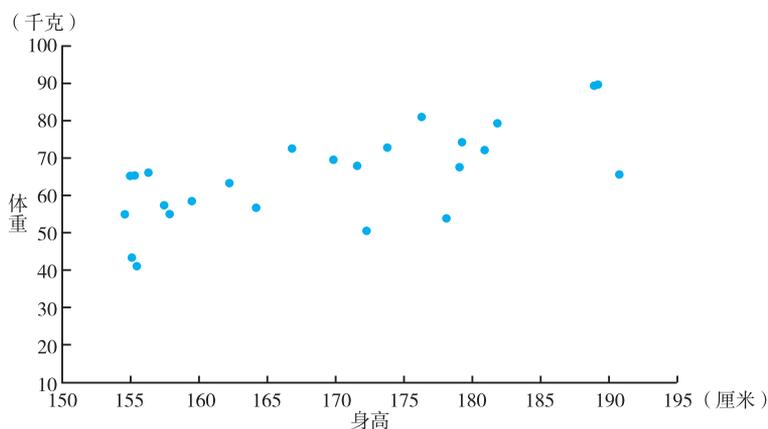


图 3-3 散点图



3.3 习题讲解

4. 折线图如图 3-4 所示。

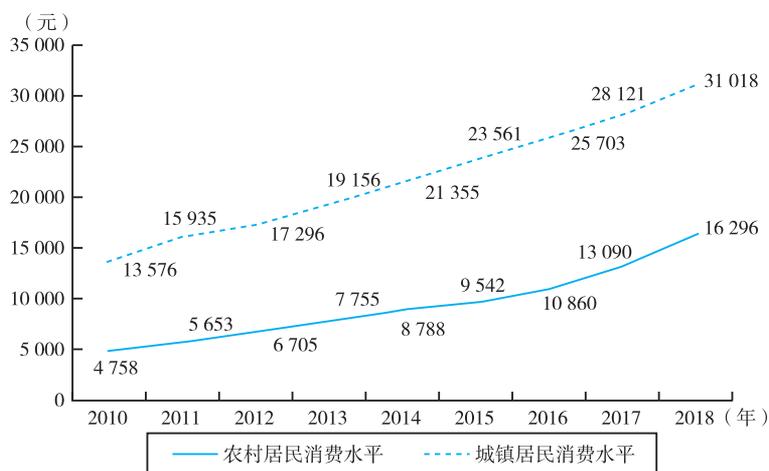


图 3-4 折线图



3.4 习题讲解

## 六、统计故事

### 南丁格尔玫瑰图

南丁格尔玫瑰图 (Nightingale Rose Diagram) 的由来与佛罗伦斯·南丁格尔有关。佛罗伦斯·南丁格尔是现代护理的奠基人,一位杰出的护士。她在 1854 年主动申请担任克里米亚战争的战地护士。由于当时医院的卫生条件极其恶劣,伤员的死亡率高达 42%。1855 年 3 月,卫生委员会来到医院并改善了医院的卫生环境,伤员死亡率骤然



下降至2.2%。南丁格尔当时敏锐地注意到了这一变化，并深信政府需要改善战地医院的条件以拯救更多年轻的生命。通过分析档案，她发现英军死亡的主要原因是战场外感染疾病及伤后缺乏适当护理，而并不是敌人的炮火。为了更清晰地呈现这一发现，南丁格尔使用了圆形图来展示数据。

图3-5来自南丁格尔建议英国政府加强公众医疗卫生建设和相关投入的文件，图中每个扇形代表战地医院每月的死亡人数，面积越大代表死亡人数越多。每个扇形中不同颜色部分代表因不同原因死亡的人数：蓝色区域代表死于受伤后得不到良好救治的人数；深灰色区域代表死于战场的人数；浅灰色区域代表死于其他原因的人数。

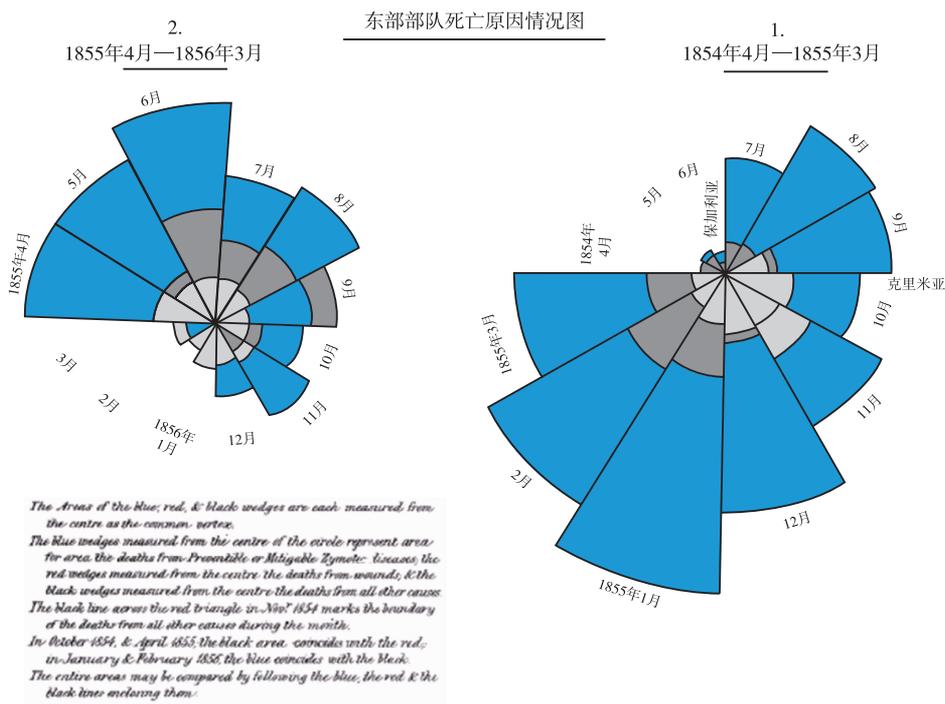


图3-5 南丁格尔玫瑰图（现存英国博物馆）

左右两个图展示的是同一个内容，且时间长度一样，均包含了12个月的数据，只是时间段不同。右侧的图展现的是1854年4月—1855年3月的数据，是卫生委员会到达医院之前的数据；左侧的图展现的是1855年4月—1856年3月的数据，是卫生委员会到达医院并改善医院卫生环境后的数据。

这样的展示让人一眼就看分明，左右两张图相比，右图有大片的蓝色区域，左图的蓝色区域显然要小得多，说明卫生委员会到达医院并改善医院卫生环境后，死亡人数明显下降。同时，蓝色区域的面积明显大于其他颜色区域的面积，这说明大多数的伤亡并非直接来自战争，而是由于恶劣的医疗环境造成的感染。

通过这幅图，南丁格尔让英国政府官员明白，改善医院的医疗状况能够显著降低英军的死亡率，同意加强公众医疗卫生建设和相关投入，最终挽救了无数人的生命。



这种新型的图形因为形状酷似绽放的玫瑰，因此被命名为“南丁格尔玫瑰图”。

南丁格尔玫瑰图，又称鸡冠花图、极差条图，是最古老的统计图形之一。类似于饼图，南丁格尔玫瑰图也是以圆心为起点，各扇形从圆心延伸出去，扇形的大小表示数据的大小。但不同于饼图用扇形角度表示数据的大小，南丁格尔玫瑰图用扇形的半径表示数据的大小，各扇形的半径不完全相同，但角度则保持一致。每个扇形区内部还用不同颜色划分了区域，不同颜色代表不同类别。

南丁格尔玫瑰图有如下的特点：

第一，南丁格尔玫瑰图采用极坐标系而不是直角坐标系。这意味着数据在图上按角度和距离定位，而不是按 x 轴和 y 轴的坐标定位。这种绘图方式使数据点在径向上展开，而不是在直角坐标系中垂直和水平展开。

第二，南丁格尔玫瑰图可以通过色彩的运用和扇形的分层来区分不同类别或组别的数据，在同一张图上同时展示多组数据的特征，有助于比较它们的差距和变化趋势，有助于提高图形的可读性和视觉效果，使数据呈现更加清晰明了。

第三，南丁格尔玫瑰图使用扇形的半径或面积表示数据大小，而不是使用扇形的角度。这使得在展示差异较小的数据时，玫瑰图的视觉对比强于饼图。因为扇形半径与扇形面积的关系是平方的关系，即半径的变化反映到扇形面积时，其结果是被“放大”的，而这种放大增强了数据间的差异，使其以更加明显的方式呈现出来。而玫瑰图的这种“放大”效应是合理的，其原理来自扇形面积公式，不是毫无来由地为“放大”而“放大”。

任何一种统计图都有其优劣势及适用范围。南丁格尔玫瑰图也不例外，它不适合于分类过少或分类数值过小的场景，也不适合表示数据占比。

南丁格尔玫瑰图是一种功能强大、表现力丰富、非常实用的数据可视化工具，能够直观地展示数据的分布和变化情况，尤其适合在需要对比分析各类数据时使用，为我们的研究和决策提供有力的支持。随着计算机技术的发展，南丁格尔玫瑰图已经从手工绘制发展到了计算机绘制。在数据可视化领域，南丁格尔玫瑰图的应用也越来越广泛，不仅用于医学、统计学等领域，还扩展到了商业、金融、市场调研等领域，通过合理运用其特点和技巧，可以有效地传达数据信息，促进理解和沟通。