

全国工程管理
专业学位研究生教育
核心课程规划教材

定量分析 模型与方法

卢向南 唐任仲 黄红选 车阿大 胡祥培 主编

清华大学出版社
北京

内 容 简 介

本书系统地介绍了工程管理问题中的定量分析模型和方法,涵盖以下内容:统计分析,包括数据及数据分析(分布分析、参数估计、假设检验、回归分析、方差分析等);预测分析,包括定性预测方法、平稳时间序列预测模型、线性趋势预测模型、季节性预测模型、具有趋势和季节成分的时间序列模型等;大数据分析,包括大数据基本概念、大数据基本决策思想、大数据基本分析方法等;线性规划,包括线性规划建模、线性规划应用及软件介绍、对偶及灵敏度分析等;线性规划的拓展,包括运输问题、最短路模型、网络优化、线性目标规划等;整数线性规划与非线性规划,包括整数规划的建模及计算机求解、指派问题、非线性规划模型与计算机求解等;决策分析,包括不确定型决策、风险决策、决策树、效用决策等;综合评价,包括专家打分法、层次分析法、模糊综合评价法等。

本书可作为工程管理硕士研究生以及其他专业学位硕士研究生的教材或参考书。

版权所有,侵权必究。举报:010-62782989, beiqinquan@tup.tsinghua.edu.cn。

图书在版编目(CIP)数据

定量分析:模型与方法/卢向南等主编. —北京:清华大学出版社,2023.3
全国工程管理专业学位研究生教育核心课程规划教材
ISBN 978-7-302-62513-1

I. ①定… II. ①卢… III. ①定量分析—研究生—教材 IV. ①O655

中国国家版本馆 CIP 数据核字(2023)第 021754 号

责任编辑:冯 昕 赵从棉

封面设计:傅瑞学

责任校对:欧 洋

责任印制:宋 林

出版发行:清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址:北京清华大学学研大厦 A 座 邮 编:100084

社 总 机:010-83470000 邮 购:010-62786544

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

质量反馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

印 装 者:三河市天利华印刷装订有限公司

经 销:全国新华书店

开 本:185mm×260mm 印 张:24.75

字 数:601 千字

版 次:2023 年 5 月第 1 版

印 次:2023 年 5 月第 1 次印刷

定 价:75.00 元

产品编号:088572-01

全国工程管理专业学位研究生教育核心课程规划教材 编委会

主 任：叶金福(西北工业大学)

副 主 任：郑力(清华大学)

委 员 (按姓氏拼音排序)：

车阿大(西北工业大学)

方志耕(南京航空航天大学)

郭波(国防科技大学)

胡祥培(大连理工大学)

蒋祖华(上海交通大学)

李启明(东南大学)

梁昌勇(合肥工业大学)

刘贵文(重庆大学)

卢向南(浙江大学)

鲁耀斌(华中科技大学)

苏秦(西安交通大学)

田琼(北京航空航天大学)

王孟钧(中南大学)

王青娥(中南大学)

王文顺(中国矿业大学)

王雪青(天津大学)

张伟(清华大学)

责任编辑：冯昕(清华大学出版社)

马宁(重庆大学出版社)

清华大学出版社

工程管理硕士(Master of Engineering Management, MEM)是经中国工程院多次建议,经严密论证,于2010年批准新设立的专业学位类别。MEM培养重大工程建设项目实施中的管理者,重要复杂新产品、设备、装备在开发、制造、生产、运维过程中的管理者,技术创新与改造、企业转型转轨以及与国际接轨中的管理者,以及产业、工程和科技的重要布局与发展战略的研究与管理者等工程管理人才。2011年3月18日,国务院学位委员会、教育部、人力资源和社会保障部在京联合召开了全国工程管理等29个专业学位研究生教育指导委员会成立会议。

2018年,按照国务院学位委员会办公室统一部署,全国工程管理专业学位研究生教育指导委员会(以下简称工程管理教指委)确定了《工程管理导论》《工程经济学》《系统工程》《定量分析:模型与方法》《质量与可靠性管理》《工程信息管理》6门MEM核心课,研究起草了核心课程大纲,并由国务院学位办统一发布。

2019年4月,工程管理教指委正式启动6门核心课程的教材编写工作,组建了核心课程系列教材编写委员会,由时任(第二届)副主任委员叶金福教授担任编委会主任,时任秘书长郑力教授担任副主任,十多位第二届工程管理教指委委员主动担任教材主编、副主编工作,并牵头组建了各门课程的编写小组。核心教材的出版工作得到了清华大学出版社和重庆大学出版社的大力支持。

根据工程管理教指委和教材编委会的统一规划设计,核心教材的编写充分考虑MEM培养要求,体现专业学位教育特点,根据发布的核心课程大纲选择知识点内容,精心设计编写方式,采用问题导向的思路,以工程管理实际问题引出各章节的知识点内容,并在各章节后提供了思考题目。

2020年新冠疫情期间,工程管理教指委克服困难,利用线下和线上工作方式,对教材草稿进行了初审、复审等工作,邀请全国多位工程管理重点培养院校有丰富教学经验的专家就教材知识框架、知识点和写作质量等内容给出详细意见和建议,秘书处逐一反馈至教材主编。教材编写小组在主编组织下开展了认真细致的修改工作。

在工程管理教指委和教材编委会统一指导下,经过众多专家们创造性的辛勤劳动,这套系列教材才得以出版。这套教材不仅适用于MEM人才培养,也适用于从事工程管理实际工作的广大专业人员深入学习工程管理核心知识。下一步,工程管理教指委(目前为第三届)将围绕本系列教材,开展6门对应核心课程的师资培训和交流工作,征集相应的精品配套资料(PPT、教学视频、延伸学习材料、精品案例等),全面提高课程质量,服务我国MEM高质量人才培养的教育目标。

因作者水平所限,时间仓促,加之工程管理发展迅速,故教材中不妥之处在所难免,欢迎广大读者批评指正,以便再版时修改、完善。

郑 力

2023 年 4 月于清华大学

清华大学出版社

工程管理硕士(master of engineering management, MEM)是为了适应我国现代工程事业发展对工程管理人才的迫切需求,完善工程管理人才培养体系,创新工程管理人才培养模式,提高我国工程管理的人才质量而设置的,培养掌握坚实、系统的工程管理理论,以及相关工程领域的基础理论和专门知识,有较强的计划、组织、协调和决策能力,能够独自胜任工程管理工作的高层次、应用型工程管理专门人才。

《定量分析:模型与方法》是 MEM 的核心课程教材。本书旨在介绍如何运用定量分析方法解决工程中的管理问题,即如何把工程管理中的问题抽象成数学模型,如何应用数学方法实现对工程管理中的问题进行定量分析决策。通过本书的学习,读者能够了解解决管理问题的基本的定量分析方法,掌握定量分析的基本原理与方法工具,根据实际问题建立定量分析模型,运用相应的软件求解模型,运用模型的结果进行问题的分析,并给出解决实际问题的方案。

本书系统地介绍了工程管理问题中的定量分析模型和方法,包括统计分析、预测分析、最优化模型、决策分析与综合评价等内容。本书的特色是,突出介绍了如何运用各种定量分析模型与方法解决工程管理中的实际问题,即如何针对实际问题建立数学模型,如何用软件求解模型,以及如何将这些模型运用在工程管理问题中的实际场景中。本书的编写紧扣管理实际问题,各章以案例为导入,针对案例问题给出定量分析模型,借助于软件介绍模型求解方法,最后,通过章后案例对相关的模型方法进行提炼和概括。

全书共分 10 章。编写分工为:第 1 章、第 2 章由陈芑熙、唐任仲编写;第 3 章由李永刚、胡祥培编写;第 4 章、第 10 章由唐任仲、陈芑熙编写;第 5 章、第 6 章由黄红选编写;第 7 章、第 8 章由车阿大编写;第 9 章由卢向南编写。全书由卢向南统稿。

由于编者水平有限,书中难免存在疏漏或错误之处,恳请广大读者批评指正。

编 者

2022 年 10 月

清华大学出版社

第 1 章 数据	1
1.1 数据的基本概念	1
1.1.1 数据的定义	1
1.1.2 工程管理数据	2
1.2 数据获取	2
1.2.1 观察数据及其获取方法	2
1.2.2 实验数据及其获取方法	4
1.3 数据整理	6
1.3.1 数据分组及其方法	6
1.3.2 分组数据图示	9
1.3.3 多变量数据图示	12
本章小结	15
习题与思考题	15
参考文献	16
第 2 章 数据分析	17
2.1 分布分析	17
2.1.1 数据分布的特征	17
2.1.2 数据分布的测度	18
2.1.3 与数据分析相关的常用概率分布	25
2.2 参数估计	27
2.2.1 点估计	27
2.2.2 区间估计	28
2.3 假设检验	30
2.3.1 假设检验的基本思想	30
2.3.2 假设检验的步骤	31
2.3.3 假设检验的内容	32
2.3.4 假设检验中的两类错误	35
2.4 回归分析	35
2.4.1 一元线性回归	36
2.4.2 多元线性回归	40

2.5 方差分析	41
2.5.1 单因素方差分析	41
2.5.2 双因素方差分析	44
本章小结	47
习题与思考题	48
参考文献	49
第3章 预测	50
3.1 定性预测方法	51
3.1.1 定性预测方法概述	51
3.1.2 主观概率法	51
3.1.3 德尔菲法	52
3.1.4 交叉影响分析法	53
3.1.5 经理评判意见法	55
3.1.6 定性预测的其他方法	56
3.2 时间序列预测法	56
3.2.1 时间序列预测法概述	56
3.2.2 简单时间序列预测法	57
3.3 移动平均法	58
3.4 指数平滑法	61
3.5 线性趋势预测模型	65
3.5.1 线性趋势预测模型概述	65
3.5.2 趋势性检验	66
3.5.3 平稳化方法	66
3.5.4 线性趋势预测模型方法	67
3.6 季节性预测模型	71
3.6.1 季节性预测模型概述	71
3.6.2 季节指数	71
3.7 具有趋势和季节成分的时间序列模型	73
3.8 平稳时间序列预测模型	77
3.8.1 平稳时间序列预测概述	77
3.8.2 时间序列的平稳性	78
3.8.3 时间序列的自相关分析	79
3.8.4 平稳时间序列预测模型及其应用	81
3.9 机器学习的预测方法	87
本章小结	88
习题与思考题	88
参考文献	91
第4章 大数据分析	92
4.1 大数据基本概念	92

4.1.1 大数据的含义	92
4.1.2 大数据的商业价值	94
4.1.3 大数据的关键技术	94
4.2 大数据基本决策思想	95
4.3 大数据基本分析方法	97
4.3.1 聚类分析	97
4.3.2 关联分析	104
本章小结	106
习题与思考题	107
参考文献	115
第 5 章 线性规划	116
5.1 线性规划建模	117
5.1.1 建模的含义	117
5.1.2 模型三要素	118
5.1.3 建模举例	120
5.2 线性规划的应用场景	125
5.2.1 生产管理	125
5.2.2 混合问题	131
5.2.3 人力资源规划	133
5.3 线性规划模型的常见形式	134
5.3.1 一般形式	135
5.3.2 标准形式	136
5.4 求解线性规划问题的图解法	138
5.4.1 可行域及作图	138
5.4.2 图解法	139
5.4.3 若干可能情形	142
5.5 求解线性规划问题的其他方法	145
5.5.1 Excel 的规划求解功能	145
5.5.2 LINDO 和 LINGO 优化求解举例	147
5.5.3 数值求解举例及单纯形法	148
5.6 灵敏度分析	154
5.6.1 灵敏度分析的实例	154
5.6.2 目标函数系数的灵敏度分析	160
5.6.3 约束右端项的灵敏度分析	162
5.6.4 灵敏度分析的 100% 规则	163
5.7 对偶问题的提出与求解	166
5.7.1 对偶问题的提出	166
5.7.2 对偶问题的求解方法	168
5.8 对偶问题的经济学解释——影子价格	171

5.8.1 对偶问题的经济学解释	172
5.8.2 影子价格	173
5.8.3 影子价格与对偶最优解	176
本章小结	177
习题与思考题	177
参考文献	182
第6章 线性规划的拓展	184
6.1 运输问题的基本模型	185
6.1.1 平衡运输问题	185
6.1.2 基本模型及特性	189
6.2 特殊形式的运输模型	191
6.2.1 供需不平衡情形	191
6.2.2 运输路线不完备的情形	194
6.2.3 转运问题	196
6.3 最短路问题	200
6.3.1 点弧关联矩阵	201
6.3.2 最短路模型	201
6.3.3 求解最短路的 Dijkstra 法	203
6.4 网络优化	205
6.4.1 最小费用流	205
6.4.2 最小生成树	208
6.4.3 最大流问题	214
6.4.4 最小割问题	221
6.5 线性目标规划	223
6.5.1 线性回归与目标规划	223
6.5.2 不等式约束的线性回归	226
6.5.3 目标规划的建模与求解方法	228
6.6 线性规划拓展的应用场景	231
6.6.1 运输模型与最小费用流	231
6.6.2 分布变换与最优传输模型	233
6.6.3 最短路模型与设备更新	236
本章小结	238
习题与思考题	239
参考文献	244
第7章 整数线性规划	245
7.1 整数线性规划问题的提出与建模	246
7.2 整数线性规划的图解法	248
7.3 0-1 整数线性规划	249
7.3.1 0-1 变量的作用	249

7.3.2	0-1 整数规划的应用	250
7.4	整数线性规划的计算机求解	254
7.4.1	分支定界法	254
7.4.2	割平面法	256
7.4.3	分支切割法	258
7.4.4	LINGO 求解整数线性规划示例	259
7.5	指派问题	260
7.5.1	指派问题的数学模型	260
7.5.2	匈牙利解法	262
7.5.3	指派问题的非标准形式	264
7.6	应用案例分析	265
7.6.1	案例简介	265
7.6.2	问题建模	268
7.6.3	优化结果及分析	271
	本章小结	272
	习题与思考题	272
	参考文献	276
第 8 章	非线性规划	278
8.1	问题的提出与建模	279
8.1.1	非线性规划问题引例	279
8.1.2	非线性规划问题的数学模型	280
8.2	最优解及其存在的条件	281
8.2.1	局部极值和全局极值	281
8.2.2	极值点存在的条件	281
8.3	无约束优化问题	282
8.3.1	梯度法	282
8.3.2	无约束优化问题的 MATLAB 求解	284
8.4	约束优化问题	285
8.4.1	最优性条件	286
8.4.2	制约函数法	288
8.4.3	约束优化问题的 MATLAB 求解	289
8.4.4	约束优化问题的 LINGO 求解	291
8.5	非线性规划的应用场景	292
8.5.1	生产管理	293
8.5.2	选址问题	296
8.5.3	人力资源规划	298
	本章小结	299
	习题与思考题	299
	参考文献	302

第 9 章 决策分析	303
9.1 决策的分类	304
9.2 决策的要素	306
9.3 决策的基本程序	306
9.4 确定型决策	307
9.5 不确定型决策	307
9.5.1 乐观主义准则	308
9.5.2 悲观主义准则	309
9.5.3 等可能性准则	309
9.5.4 最小机会损失准则	310
9.5.5 折中主义准则	311
9.6 风险决策	311
9.6.1 最大期望收益准则	312
9.6.2 最小期望机会损失准则	312
9.7 决策树	313
9.7.1 单项决策	313
9.7.2 序列决策	314
9.8 灵敏度分析	315
9.9 信息的价值	316
9.9.1 全信息的价值	316
9.9.2 不全信息的价值与贝叶斯决策	317
9.10 效用决策	320
9.10.1 效用及效用曲线	320
9.10.2 效用曲线的确定	320
9.10.3 效用期望值决策准则	322
9.10.4 效用期望值准则与期望值准则的区别	323
9.11 应用 Excel 和 TreePlan 求解决策树	325
9.12 行为决策理论简介	331
9.12.1 行为决策理论的发展阶段及研究方法	333
9.12.2 行为决策的几个主要理论	334
本章小结	336
习题与思考题	337
参考文献	343
第 10 章 综合评价	344
10.1 基本概念	344
10.1.1 综合评价的五个基本要素	345
10.1.2 综合评价的基本步骤	346
10.2 评价指标体系的建立	347
10.2.1 评价指标体系建立的一般原则和常用方法	347

10.2.2	评价指标的一致化处理	347
10.2.3	评价指标的无量纲化处理	348
10.3	指标权重的确定	349
10.4	综合评价方法	350
10.4.1	专家打分法	350
10.4.2	层次分析法	351
10.4.3	模糊综合评价法	356
	本章小结	360
	习题与思考题	360
	参考文献	369
附录 A	几种优化软件简介	370
A.1	Excel 软件: 规划求解	370
A.2	LINDO 和 LINGO 软件	370
A.3	MATLAB 软件: 优化工具箱	374
A.4	基于 CPLEX 软件求解	376

清华大学出版社

数 据

【教学内容、重点与难点】

教学内容：工程管理数据的定义与特点,观察数据与实验数据的获取方法,数据整理与可视化展示方法。

教学重点：工程管理数据的定义与特点,观察数据的问卷调查法获取,数据分组与图表化展示。

教学难点：工程管理数据、工程数据及管理数据之间的关系。



数据是智能制造的核心

智能制造是一种将新一代信息技术与先进制造技术结合,实现对产品全生命周期的实时管理和优化的新型制造系统,是制造企业实现转型升级的主要途径。

通过智能制造,采用产品全生命周期的实时管理、先进的传感技术和远程监控技术,可以缩短产品从研发到上市的时间,缩短从订单到交付的时间,减少生产中断时间,从而缩短产品的研制周期。通过智能制造,采用数字化、互联和虚拟工艺规划,可以实现大批量定制生产,提高生产的灵活性。通过智能制造,可以促进企业从“以产品为中心”向“以集成服务为中心”转变,帮助企业利用服务在产品全生命周期中创造新价值。

某特种钢铁企业实施矿渣粉磨系统智能制造项目,通过优化决策和动态执行达到提高生产效率、减少设备故障停机、满足环保要求的目标。优化决策的指令来自于对生产系统的海量信息的挖掘、分析和推理预测,需要大量的数据支持,那么,如何获取数据,如何展示数据整理和分析的结果,将是该项目面临的基础性问题。本章将重点介绍数据获取的主要方法、数据整理和可视化展示的主要工具,以便初步解答上述问题。

1.1 数据的基本概念

1.1.1 数据的定义

数据是用于描述客观事物的、可识别的符号。不仅有数值型数据,还有非数值型数据,如等级数据、属性数据等。数据的表现形式除了数字之外,还可以是文字、符号、图像、音频、

视频等。例如,“0、1、2、…”“合格、不合格”“生产计划、物料清单”等都是数据。

数据本身并不能完全表达出其所代表的内容,例如,170 是一个数据,它可以是某个专业的学生人数,可以是某个人的身高,也可以是某个人的体重,因此,对数据的理解必须与数据的语义结合起来。

1.1.2 工程管理数据

工程是改造世界的实践活动,是一种有组织、有计划、有目的的人工活动,通过科学和数学的某种应用,向社会提供有价值的产品或服务。

工程管理指通过决策、计划、组织、指挥、协调和控制以实现工程预期目标的过程。在这些过程中产生的、需要用到数据就是工程管理数据。

工程数据是涉及工程本身的数据,例如,大楼的高度、产品的功能、产品的性能,等等。管理数据是涉及管理过程的数据,例如,营运数据、市场营销数据,等等。工程管理数据是一种管理数据,是对工程活动进行管理所产生的和/或使用到的数据,例如,产品的质量、材料库存、材料利用率、生产计划,等等。

工程管理的对象是动态变化的,在工程全生命周期的各阶段,其任务和目标不尽相同,因此,工程管理数据是多方面、多维度的,会受工程发展方向、时空变化以及工程参与人员等因素影响而产生变化,具有复杂性、多样性、关联性、动态性、可度量性等特点。

1.2 数据获取

1.2.1 观察数据及其获取方法

观察数据是指在获得数据的过程中,不对被分析对象数据产生的条件施加任何控制所得到的数据。可以通过统计报表、普查、重点调查、典型调查、抽样调查、访谈、专家调查、问卷调查、基于传感技术的数据采集等方法获取观察数据。

1. 总体、个体和样本

被研究对象的全体称为“总体”,例如,一条生产线上生产出来的零件可以构成一个总体。从总体中抽取出来用于调查研究的那部分对象所构成的集合称为“样本”,例如,从一条生产线上生产出来的零件中抽取的若干个零件可以构成样本。组成总体的每一个研究对象称为“个体”,例如,一条生产线上生产出来的零件中的任意一个零件就是一个个体。

2. 抽样调查

抽样调查是从调查研究对象的总体中抽选一部分对象进行调查,并根据调查结果对调查研究对象的总体作出估计和推断的一种调查方法。在抽样调查中,样本数的确定是一个关键问题。当对总体中的所有个体进行调查时,即为普查。

抽样调查的目的是获取反映总体情况的信息,因此,抽样调查虽然是一种非全面调查,但也能起到类似全面调查的作用。抽样调查包含两层含义:一是抽样,从总体中抽取部分个体;二是调查,考查那些反映在个体上的所要研究的特征的数据。根据样本抽选方法的不同,抽样可分为概率抽样和非概率抽样。

概率抽样是以概率论和数理统计原理为依据,按照随机原则从总体中抽选样本。非概率抽样是按照非概率的原则,或者依据对个体和总体特征的判断,从总体中抽选样本。习惯上将概率抽样称为抽样调查,其特点是总体中所有个体被抽中的概率相同。常用的概率抽样方法有简单随机抽样、系统抽样、分层抽样、整群抽样等。

在从总体中抽选了研究所需的样本后,需要调查研究对象的特征,从样本中的每个个体上获得所需要的数据。常用的调查方法有访谈法和观察法。

抽样调查具有适应面广、准确性高、经济性好、时效性强等优点。

(1) 抽样调查可以根据调查目的设定调查项目和指标,适用于各个领域的调查。

(2) 抽样调查是用样本数据去推算总体,虽然会存在推算误差,但是可以应用概率论和数理统计原理计算和控制这种误差,从而保证调查的精度。

(3) 抽样调查抽选的样本是总体中的一小部分,调查的工作量小,调查效率高,调查费用低。

(4) 由于抽样调查工作量小,因此可以快速、及时地获得数据,从而保证了调查的时效性。

3. 问卷调查

问卷调查是指通过制定与研究目标有关的问卷,根据被调查者据此进行的回答收集资料的方法,它是人们在调查研究活动中用来收集资料的一种常用工具,其主要优点在于标准化和成本低。

问卷调查法通过设计、发放和回收调查问卷来获取调查数据,问卷设计的质量对问卷调查数据的获取以及问卷调查数据的质量有直接的影响。因此,进行调查问卷设计时既要安排合适的问卷内容,又要设计恰当的问卷格式。

调查问卷的设计一般分为以下几个步骤:

(1) 充分了解调查的目的,将需要调查的内容具体化和条理化,确定问卷设计所需的信息。

(2) 明确调查所要了解的内容和所要搜集的数据,确定需要通过问卷调查来获取的数据。

(3) 确定问卷中具体包括的问题和询问的内容。

(4) 确定问答题的结构,采用封闭式问题还是开放式问题。

(5) 确定问答题的措辞,要求问题措辞准确清楚,封闭式问题的答案要穷尽、互斥。

(6) 合理安排问题的顺序,通常将简单的、容易回答的问题以及被调查者感兴趣的问题放在前面,复杂的、较难的问题以及开放式的问题放在后面,并且按照问题的逻辑顺序来排列问题。

(7) 确定格式和排版,问卷上除了需要调查的问答题,还应该包括卷头语、卷尾语、问卷答题说明、调查对象的背景信息等。

(8) 问卷设计完成之后,需要选择少部分调查对象进行模拟调查试验,根据试验结果进行修正和定稿。

下面是某汽车企业为了获取用户对电动汽车关切的相关信息而设计的一份调查问卷。

×××汽车企业对用户关切的电动汽车信息的调查表

尊敬的客户：非常感谢您参与本次调查，此次调查是为了了解用户对电动汽车关切的相关信息，以便对今后电动汽车的研发、生产、销售提供参考和指导。感谢您的支持！

1. 您的年龄属于以下哪个年龄段？

☐ A. 28 周岁以下	☐ B. 28 周岁(含)~35 周岁以下
☐ C. 35 周岁(含)~45 周岁以下	☐ D. 45 周岁及以上
2. 您的学历是？

☐ A. 大专以下	☐ B. 大专
☐ C. 本科	☐ D. 研究生及以上
3. 您的工作所属行业？

☐ A. 公务员	☐ B. 互联网企业职员
☐ C. 私营企业主	☐ D. 其他
4. 您身边的同事或者亲朋好友是否购买过电动汽车？

☐ A. 有	☐ B. 没有
☐ C. 好像有	☐ D. 不太清楚
5. 您本人是否在未来有意愿购买电动汽车？

☐ A. 未来 1 年内有购买意愿
☐ B. 未来 2~3 年内有购买意愿
☐ C. 未来 3 年后可能会考虑购买
☐ D. 暂时没有购买电动汽车的想法和计划
6. 您最关注电动汽车的哪些指标？

☐ A. 性价比	☐ B. 续航里程
☐ C. 技术先进性	☐ D. 美观性
7. 您觉得电动汽车代替燃油汽车，是否是一个不可逆转的趋势？

☐ A. 是	☐ B. 不是	☐ C. 不一定
----------------------------	-----------------------------	------------------------------
8. 您觉得您会优先选择以下哪类企业生产的电动汽车？

☐ A. 传统国内车企	☐ B. 传统合资车企或者纯外资车企
☐ C. 互联网车企	
9. 您觉得什么原因可能会让您下定决心去买一辆电动汽车？

☐ A. 限牌/限行	☐ B. 环保
☐ C. 驾驶体验	☐ D. 其他
10. 您会优先考虑购买电动汽车还是租赁电动汽车？

☐ A. 购买	☐ B. 租赁
☐ C. 两者都可以接受	☐ D. 其他

1.2.2 实验数据及其获取方法

实验数据是指在实验中通过控制实验对象而获得的变量数据，是在实验中控制一个或

多个变量,在有控制的条件下得到的观测结果。

实验设计通过对数据产生条件的恰当限定、对数据产生过程的合理设计而获得数据。

用实验设计方法来获得数据的主要优点如下:

- (1) 可以获得在真实情况下用调查法无法获得的某些数据;
- (2) 可以在一定程度上直接观察到某些因素之间的相关关系;
- (3) 可以获得在抽样调查、问卷调查中难于获得诚实回答的数据。

在实际生产活动中,实验结果往往受到多方面因素的影响,因此需要考察各因素对实验结果的影响情况。在多因素、多水平实验中,如果对每个因素的每个水平的所有组合都进行实验,则实验的次数就会很多。

例如,对一个 3 因素 7 水平的研究对象进行实验研究,如果 3 个因素 7 个水平都互相组合进行全面实验,就要做 $7^3=343$ 次实验;如果因素数量增加到 6 个,那么对 6 因素 7 水平进行全面实验要做 $7^6=117\,649$ 次实验,这显然是不经济的。

如果能够在不影响实验效果的前提下,尽可能地减少实验次数,将极大地减轻实验工作强度,降低实验成本,正交试验设计是解决这个问题的一种有效方法。

正交试验设计是一种利用正交表,挑选试验条件,确定试验计划,在较少的试验次数下,找出最优或较优组合的实验设计方法,是研究与处理多因素多水平试验的一种科学方法。

正交表是进行正交试验设计的主要工具,是一整套规则的设计表格,用正交表来安排试验时,各因素的各种水平的搭配是均衡的,可以实现以最少的试验次数达到与大量全面试验等效的结果。

一般的正交表记为 $L_n(m^k)$, L 为正交表的代号; n 为表的行数,也就是要安排的试验数; k 为表的列数,表示因素的个数; m 为各因素的水平数。

常见的正交表如下:

2 水平的有 $L_4(2^3)$, $L_8(2^7)$, $L_{12}(2^{11})$, $L_{16}(2^{15})$ 等;

3 水平的有 $L_9(3^4)$, $L_{27}(3^{13})$ 等;

4 水平的有 $L_{15}(4^5)$ 等;

5 水平的有 $L_{25}(5^6)$ 等。

例如,某混凝土预制构件厂为提高预制构件的质量,需要通过实验确定最好的生产工艺。经初步分析,影响预制构件质量的主要因素为养护温度、水灰比和含砂率,每个因素都考虑有两个水平,如表 1-1 所示。如果每个因素的每个水平都互相组合进行全面实验,必须做实验 $2^3=8$ 次,如果采用正交试验设计方法,使用 $L_4(2^3)$ 正交表来设计实验方案,则只需要从 2^3 次中选出 4 次实验。以 C20 预制构件选取立方体试块为例,每次实验进行 100 个混凝土立方体试块配置,统计合格品的产出数量,实验方案和实验结果如表 1-2 所示。

表 1-1 预制构件质量的主要影响因素与水平

水 平	因 素		
	养护温度 $T/^{\circ}\text{C}$	水灰比 W	含砂率 S
1	18	0.54	0.48
2	20	0.60	0.33

表 1-2 预制构件质量正交试验设计

编 号	因 素			
	养护温度 $T/^{\circ}\text{C}$	水灰比 W	含砂率 S	合格数量
1	1	1	1	35
2	1	2	2	90
3	2	1	2	88
4	2	2	1	75

表 1-2 中记录的实验参数、实验参数设置以及实验的结果即为实验数据,都是实验中直接可以获得的数据。

1.3 数据整理

数据整理是对抽样、调查、观察、实验等研究活动中所搜集到的原始数据进行审核、分组、汇总等加工处理,使之条理化、系统化,以符合统计分析的需要,同时用图表形式将数据展示出来的工作过程。数据整理的目的是简化数据并使之更容易理解和分析。

工程管理数据既有数值型数据,又有等级数据、属性数据等非数值型数据。非数值型数据在进行数据整理之前需进行量化处理,转化为数值型数据。对等级数据和属性数据,一般可采用李克特量表(Likert scale)来量化。

数据整理一般包括两个方面:一是对数据的处理,主要是考虑如何进行统计分组;二是确定反映数据基本特征的相关指标。

频数和频率是描述样本数据基本特征的主要指标。频数表示的是数据值在样本数据中的出现次数。某个数据值的频数与样本数据总数的比值称为该数据值在样本数据中出现的频率。

整理后的数据可以利用 SPSS 等常用统计分析软件中的图形工具或 Excel 中提供的图表工具进行可视化展示。

1.3.1 数据分组及其方法

数据分组是根据统计研究的需要,按照预先设定的标准将原始数据划分成不同的组别,分组后的数据称为分组数据。数据分组的主要目的是统计原始数据在各组中出现的频数,编制频数分布表,观察数据的分布特征。

数值型数据依据数值的不同进行数据分组,非数值型数据依据数据属性的不同进行分组。

数据分组需要遵循穷尽性原则和互斥性原则。穷尽性原则要求调查的每个样本都能无一例外地划归到某一组中,不会产生“遗漏”现象。互斥性原则要求分组后各个组的范围应该互不相容、互为排斥,即每个样本在特定的分组标志下只能归属某一组,而不能同时或可能同时归属到几个组。

数据分组有单变量值分组和组距分组两种方法。

1. 单变量值分组

把一个变量值作为一个组的分组方法称为单变量值分组,对于离散变量,且变量值较少的情况,通常采用单变量值分组。

例 1-1 某企业生产线改造工程中,为了设计合适的生产节拍,需要对各工序进行作业测定,对其中某道工序进行作业测定时测得的 50 次重复作业的作业时间如表 1-3 所示。采用单变量值分组法对数据进行分组并计算各组中数据出现的频数,结果如表 1-4 所示。

表 1-3 某工序 50 次重复作业的作业时间 单位：s

108	110	112	137	122	131	118	134	114	123
125	123	127	120	122	118	119	127	124	125
109	112	135	121	129	116	118	123	128	138
122	131	119	124	106	133	134	113	115	117
126	127	121	119	130	122	123	117	128	123

表 1-4 某工序 50 次重复作业的作业时间单变量值分组结果

作业时间/s	频数/次	作业时间/s	频数/次	作业时间/s	频数/次
106	1	118	3	128	2
108	1	119	3	129	1
109	1	120	1	130	1
110	1	121	2	131	2
112	2	122	4	133	1
113	1	123	5	134	2
114	1	124	2	135	1
115	1	125	2	137	1
116	1	126	1	138	1
117	2	127	3		

2. 组距分组

将全部变量值依次划分为若干区间,并将每一个区间作为一个组的分组方法称为组距分组。对于连续变量或变量值较多的情况,通常采用组距分组。组距分组可采用等距分组,也可采用不等距分组。

组距分组具体的步骤如下:

(1) 确定组数。组数一般与数据本身的特点及数据的多少有关,组数的确定以能够反映数据的分布特征和规律为原则。在实际分组时,可以根据 H. A. Sturges 提出的经验公式确定组数 K :

$$K=1+\frac{\lg n}{\lg 2}$$

(1-1)

式中, n 为数据个数。

(2) 确定各组的组距。可以根据全部数据的最大值、最小值以及划分的组数来确定组距,组距=(最大值-最小值)÷组数。

(3) 根据分组编制频数分布表。

在组距分组中,一个组的最小值称为下限;一个组的最大值称为上限;上限与下限之

差称为组距；下限与上限之间的中点值称为组中值；上限和下限都齐全的组称为闭口组；上限或下限有一个没有的组称为开口组。

组限的确定有一个基本原则，即按这样的组限分组后，标志值在各组的变动能反映事物的质的变化。

连续变量分组时，由于相邻两组的上限和下限常是同一数值，每组的界限会重叠，为避免计算各组次数时出现混乱，组限表示的一般原则是“上组限不在内，或下组限不在内”。

离散变量分组时，由于相邻的上限和下限通常是以两个确定的不同整数值来表示，因此相邻两组的上下限可以不重合。

例 1-2 对表 1-3 记录的某工序 50 次重复作业的作业时间数据进行组距分组。

分别按上下组限重叠、上下组限间断和使用开口组进行等距分组，分组结果如表 1-5、表 1-6 和表 1-7 所示。

表 1-5 某工序 50 次重复作业的作业时间按上下组限重叠的等距分组结果

按作业时间分组/s	频数/次	频率/%
105~110	3	6
110~115	5	10
115~120	10	20
120~125	14	28
125~130	9	18
130~135	6	12
135~140	3	6
合计	50	100

表 1-6 某工序 50 次重复作业的作业时间按上下组限间断的等距分组结果

按作业时间分组/s	频数/次	频率/%
105~109	3	6
110~114	5	10
115~119	10	20
120~124	14	28
125~129	9	18
130~134	6	12
135~139	3	6
合计	50	100

表 1-7 某工序 50 次重复作业的作业时间使用开口组的等距分组结果

按作业时间分组/s	频数/次	频率/%
110 以下	3	6
110~114	5	10
115~119	10	20
120~124	14	28
125~129	9	18
130~134	6	12

续表

按作业时间分组/s	频数/次	频率/%
135 以上	3	6
合计	50	100

等距分组与不等距分组在反映频数分布上存在差异：等距分组各组频数的分布不受组距大小的影响,不等距分组各组频数的分布受组距大小的影响；等距分组的各组绝对频数反映了频数分布的特征和规律,不等距分组各组绝对频数的多少不能反映频数分布的实际状况,需要用频数密度来反映频数分布的实际状况。

3. 累计频数分布

在频数分布的基础上将各组频数逐一累计,称为频数分布累计,累计得到的频数称为累计频数。频数分布累计分为向上累计和向下累计两种方式。向上累计是从变量值最低组开始向变量值高的组累计,表明小于该组上限的频数(频率)一共有多少；向下累计是从变量值最高组开始向变量值低的组累计,表明大于该组下限的频数(频率)一共有多少。同一数值的向上累计和向下累计次数之和等于总体总次数。以变量值为横坐标、累计频数和频率为纵坐标,可以画出累计频数分布图。

例 1-3 对表 1-3 记录的某工序 50 次重复作业的作业时间数据进行组距分组并计算累计频数分布。

按上下组限重叠对表 1-3 记录的某工序 50 次重复作业的作业时间数据进行等距分组,分组数据的累计频数分布如表 1-8 所示,对应的累计频数分布图如图 1-1 所示。

表 1-8 某工序 50 次重复作业的作业时间的等距分组累计频数分布

按作业时间分组/s	频数/次	频率/%	向上累计频率/%	向下累计频率/%
105~110	3	6	6	100
110~115	5	10	16	94
115~120	10	20	36	84
120~125	14	28	64	64
125~130	9	18	82	36
130~135	6	12	94	18
135~140	3	6	100	6
合计	50	100	—	—

1.3.2 分组数据图示

1. 直方图

直方图是一种用矩形的宽度和高度来表示频数分布的图形,实际上是用矩形的面积来表示各组的频数分布。在直角坐标中,用横轴表示数据分组,纵轴表示频数或频率,各组与相应的频数构成的矩形即为直方图。在以频率表示的直方图中,直方图下的总面积等于 1。

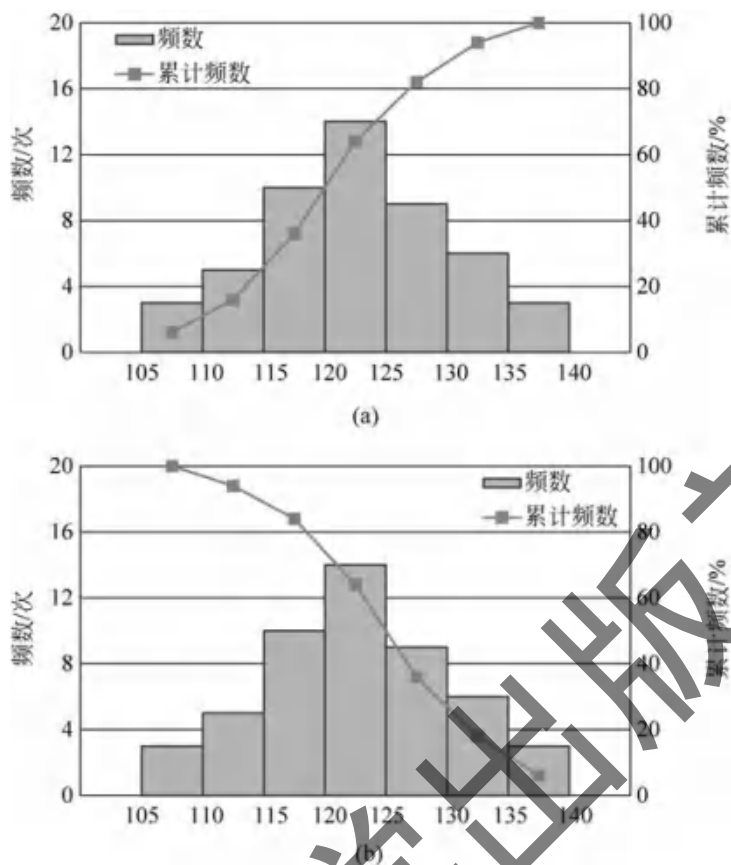


图 1-1 累计频数分布图

(a) 向上累计频数分布图 (b) 向下累计频数分布图

例 1-4 对表 1-3 记录的某工序 50 次重复作业的作业时间按上下组限重叠进行等距分组后的频数直方图如图 1-2 所示。

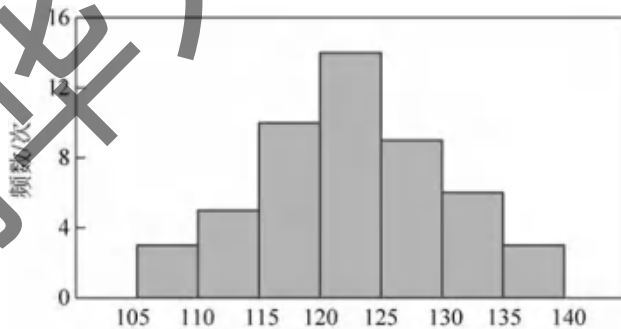


图 1-2 某工序 50 次重复作业的作业时间频数直方图

对本例中已经分组的数据,可以应用 Excel 的图表工具中的柱形图工具生成频数直方图。在 Excel 表中插入柱形图,选择已分组的数据为数据源,设置和编辑坐标轴等属性参数后即可生成该分组数据的频数直方图,主要过程如图 1-3~图 1-6 所示。

2. 折线图

折线图是在直方图的基础上,把直方图顶部的各组中点(各组中值对应的点)用直线连接起来,再去掉原来的直方图后的图形,也称频数多边形图。



图 1-3 选择柱形图模板



图 1-4 选择数据源

图 1-5 设置坐标轴选项

折线图的两个终点在横轴上,绘制折线图时,将第一个矩形的顶部中点与左侧竖边中点(即该组频数值的一半的位置)连接并延长到横轴,得到左侧终点,将最后一个矩形顶部中点与其右侧竖边中点连接并延长到横轴,得到右侧终点。折线图与横轴构成一个封闭的多边形,这个多边形的面积与原直方图的面积相等,二者所表示的频数分布是一致的。

例 1-5 用折线图展示表 1-3 记录的某工序 50 次重复作业的作业时间按上下组限重叠进行等距分组后的频数变化。

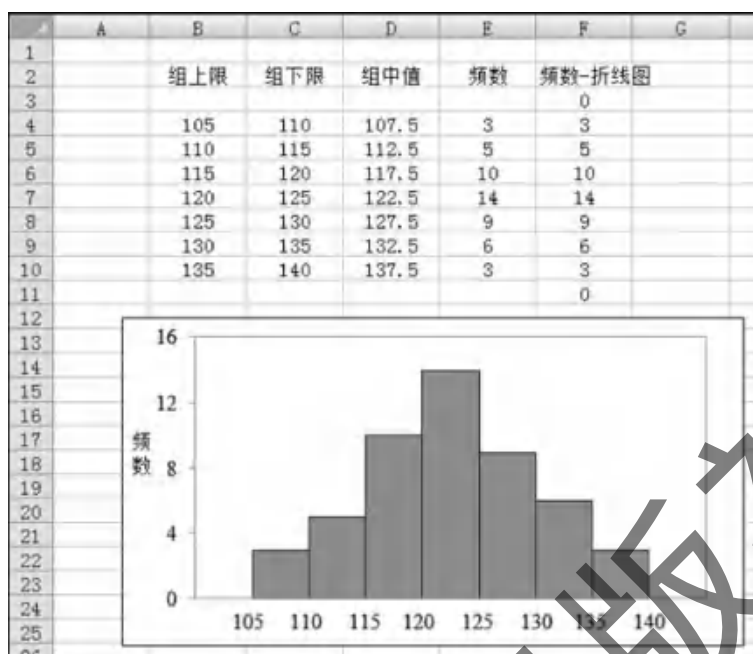


图 1-6 输出结果

图 1-7 是应用 Excel 的图表工具在频数直方图上绘制的频数折线图。在频数直方图的数据源中,通过添加数据系列,并将新添加的数据系列的图表类型更改为折线图类型后即可生成对应的频数折线图。

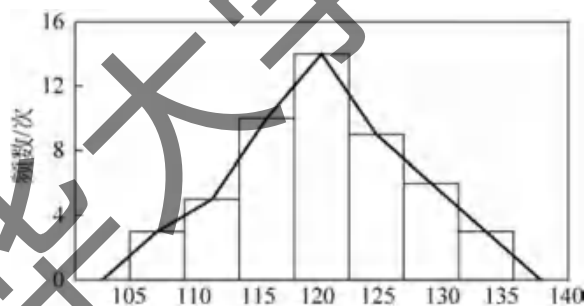


图 1-7 某工序 50 次重复作业的作业时间频数折线图

1.3.3 多变量数据图示

1. 雷达图

雷达图是以从同一点开始的轴上表示的三个或更多个变量的二维图形的形式显示多变量数据的图形方法。雷达图常用于显示或对比各变量的数值总和,也可用于研究多个样本之间的相似程度,当不同样本间各变量的取值具有相同的正负号时,总的绝对值与图形所围成的区域成正比。

例 1-6 某企业三相油浸式电力变压器的全生命周期成本(LCC)如表 1-9 所示,根据全生命周期成本的组成项绘制该产品全生命周期成本的雷达图。

图 1-8 所示是应用 Excel 图表工具中的雷达图工具生成的三相油浸式电力变压器的全

生命周期成本雷达图。

表 1-9 某企业三相油浸式电力变压器的全生命周期成本(LCC)

成 本 项	产品 LCC/元
研发成本	4172
制造成本	123 714
销售成本	2874
使用成本	200 348
退役成本	—2712
LCC 合计	328 396

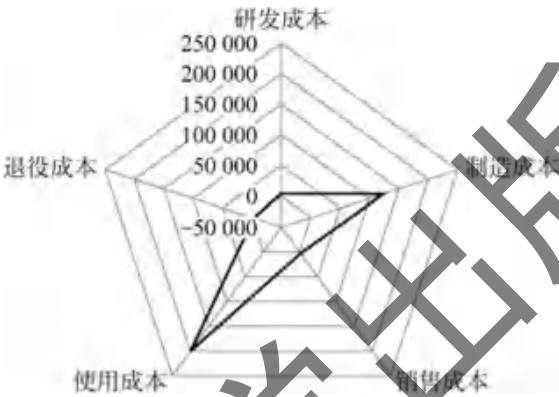


图 1-8 某企业三相油浸式电力变压器的全生命周期成本(LCC)的雷达图

2. 散点图

散点图是用二维坐标展示两个变量之间关系的一种图形,使用的数据是成对的数据,如 (x_i,y_i) ,以变量 x 为横轴,变量 y 为纵轴,在坐标系中描出各数据点。

例 1-7 某工序需要将零件安装到夹具上进行加工,装夹零件时取零件作业时间与零件质量有关,取零件作业时间与零件质量之间的关系的数据如表 1-10 所示。根据表 1-10 的数据,用散点图描述零件作业时间与零件质量变化的关系。

表 1-10 装夹零件时取零件作业时间与零件质量的关系

时间/s	4	10.5	7.5	6.5	6	8.5	8	7.5	5
质量/kg	2	8	6	4	3	9	7	5	1

图 1-9 所示为应用 Excel 图表工具中的散点图工具生成的取零件作业时间与零件质量变化关系散点图。

3. 气泡图

气泡图是用以展示三个变量之间关系的一种图形。它与散点图类似,在散点图的基础上,将点变成有大小的气泡。绘制时一个变量为横坐标,一个变量为纵坐标,而第三个变量则用气泡的大小来表示。

例 1-8 零件质量的变化将影响取零件作业时间,从而影响零件的装夹时间。表 1-11 所示为零件总的装夹时间与取零件作业时间和零件质量关系的抽样调查数据,图 1-10 所示

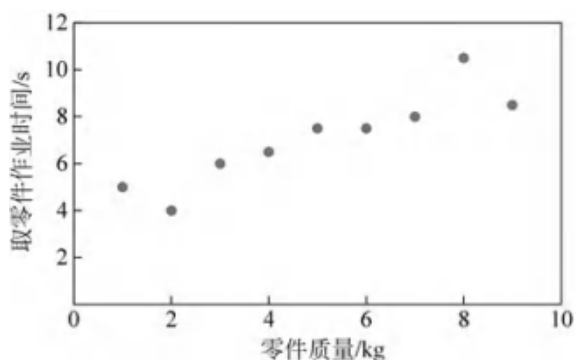


图 1-9 取零件作业时间与零件质量关系散点图

为应用 Excel 图表工具中的气泡图工具生成的零件总的装夹时间与取零件作业时间和零件质量关系的气泡图。气泡图上的数值表示该零件总的装夹时间,数值越大,对应的气泡就越大。

表 1-11 零件总的装夹时间与取零件作业时间和零件质量的关系

零件质量/kg	2	8	6	4	3	9	7	5	1
取零件作业时间/s	4	10.5	7.5	6.5	6	8.5	8	7.5	5
总的装夹时间/s	16	9.8	23.5	19	15	31.5	27	31.5	38.5

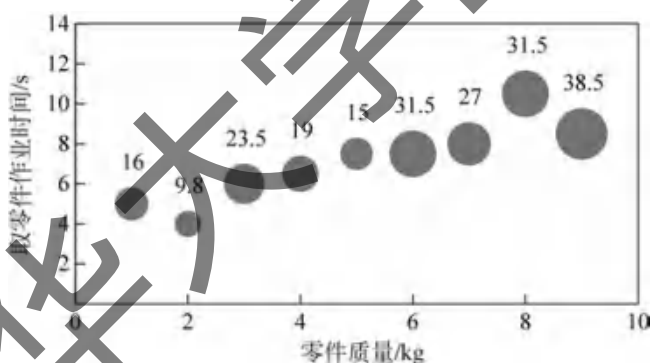


图 1-10 零件总装夹时间与取零件作业时间和零件质量关系气泡图

4. 时间序列图

时间序列是一组在不同时间采集到的按照时间顺序排列的数据序列,用于描述数据随时间变化的情况。时间序列图是以时间为横轴,变量为纵轴来展示时间序列数据的一种图形,也是折线图的一种形式。时间序列图可以直观地反映某一特征随时间的变化状态或程度。

例 1-9 表 1-12 所示为某操作者某年每个月生产的零件的合格率数据,为了分析该操作者的作业质量,可以用时间序列图来展示该操作者生产零件的合格率随时间变化的情况。

图 1-11 所示为应用 Excel 图表工具中的折线图工具生成的操作者月生产零件的合格率时间序列图。