

# 第1章

## 生产系统建模与仿真概述

制造业正在朝产品个性化、生产柔性化、过程自动化的方向转变,现代制造工厂的内部结构和运行逻辑也越来越复杂,如何科学地进行工厂设计、定量分析工厂性能(包括效率、质量、交期和成本等指标),并持续优化工厂运行,是智能工厂(smart factory)实施的重要内容。现如今,虚拟建模与仿真技术作为智能工厂重要的使能技术,已广泛应用于生产系统的性能评估分析过程中,与纯数学分析方法相比,它不仅能将各种复杂的生产与物流活动、调度与控制逻辑转换为直观的描述,还能够充分考虑系统内外的随机因素影响;不仅能够建立静态模型,还可以模拟其动态行为。通过多次仿真实验运行,得到大量数据,基于对仿真数据全面而深入的分析,有助于科学确定生产系统设计参数,准确预测系统性能,从而制定更优的生产策略。

本章介绍了生产系统的基本概念,概述了虚拟建模与仿真的基本目标、相关理论知识和软件产品,并总结了生产系统建模与仿真的基本过程。

### 1.1 生产系统概述

#### 1.1.1 系统

系统(system)是由若干部分相互联系、相互作用,形成的具有某些功能的整体。根据系统状态变化的时间连续性与否,可将系统分为连续系统(continuous system)和离散系统(discrete system)。其中,离散系统是指全部或关键组成部分的变量具有离散信号形式,状态在时间的离散点发生突变的系统。

描述系统的基本要素包括对象(object)、属性(property)、活动(activity)、输入输出(input/output, I/O)。“对象”又称为“实体”(entity),它确定了系统的构成和边界,可区分为临时对象与永久对象,在系统中只存在一段时间的对象叫作临时对象,比如顾客、工件、工人等,它们一般是流动的,永久驻留在系统中的对象则叫作永久对象,比如服务台、设备等,它们一般是静止的。“属性”描述了每个对象的基

本特征。“活动”定义了对对象之间的相互作用,从而确定了系统状态随时间发生变化的过程。“输入输出”描述了系统与外部环境的物质和信息交互。

现实世界有大量的系统,比如银行、医院、工厂、物流中心,稍加分析,就可以确定出这些系统的基本要素。以银行服务系统为例:对象包括顾客、服务台和银行工作人员;顾客的属性包括到达时间、是否为VIP、服务类型等,服务台的属性包括类型、服务时段等,银行工作人员的属性包括人员编号、所在服务台、处理任务时长等;活动包括顾客存取款活动、业务咨询活动、理财活动等;输入信息包括顾客的到达规律、顾客的类型分布等,输出信息包括资金变化、服务效率等。

### 1.1.2 离散事件系统

在生产活动和社会活动中,经常会遇到一类特殊的系统,这类系统中有许多事件时而出现、时而消失、时而动作、时而停止,而启动和停止都发生在一些离散的时刻,并带有一定的随机性。例如,港口中的船舶停靠码头、生产线上机床的启停、电话的接通和断开、计算机系统中某项作业的进行和退出等,这类系统统一称为离散事件系统(discrete event system, DES)。它属于离散系统大类,是由离散触发事件驱动状态演化的动态系统,其状态变量只在某个离散时间点上发生变化,有时也称为离散事件动态系统(discrete event dynamic system, DEDS)。

离散事件系统的基本特性如下:

(1) 时间离散性。时间本身虽然是连续的,但系统状态只在离散时刻点发生变化,未发生状态变化的时刻一般不需要关心,因此可将时间离散化成一连串间隔不等的时刻点。

(2) 不确定性。事件何时发生通常是不确定的,比如订单到达、设备故障、质量异常等,各种偶然因素导致了这些事件发生时间的随机性,一般通过随机变量来模拟不确定性。

(3) 动态性。事件的发生、状态变化、活动触发都是动态的。

(4) 层次性。体现为空间跨度、时间跨度、组织结构的层次性。

(5) 计算复杂性。由于组成对象数目大,事件状态多,常有“组合爆炸”的危险,给分析计算带来了很大困难。

常见的离散事件系统包括排队系统、库存系统、加工系统、装配系统、项目系统、物流系统等,这些单元性的离散事件系统在各种书籍中都有详细介绍,读者可自行查阅。

### 1.1.3 生产系统

生产系统(production system)是一种组合性的复杂系统,通常由若干个排队系统、库存系统、加工系统、装配系统、项目系统、物流系统等组合而成,通过它来完成产品的生产与物流作业,实现产品制造。生产系统有不同的层级范围,制造单



生产系统  
示例

元、生产线、车间、工厂、供应链就是从低到高的层级。生产系统的层级越高,其内部结构和活动逻辑也越复杂。

生产系统的基本构成要素包括:

(1) 生产与物流设施对象,包括设备、加工/装配/拆卸/检测站、机器人、缓冲区/仓库、传送线、小车系统、天车系统、工人调配系统、资源运输系统等;

(2) 工件对象,包括零件毛坯、在制零部件、配套零部件、产成品等;

(3) 资源对象,包括物流车辆、托盘、工具、工人等;

(4) 数据,包括布局数据、工艺数据、订单数据、物流数据、资源数据、工作日历数据等;

(5) 活动,包括加工、装配、质量检测等工艺执行活动,生产计划与调度活动,物料请求、调度、出入库、运输、上下料等物流活动,资源请求与调度活动,故障与维修活动,质量异常处置活动等;

(6) 输入输出信息,包括生产计划、生产过程记录、生产绩效等。

生产系统的类型多样,既有离散系统,也有连续系统或离散-连续混合系统。比如,汽车、电子、家电、航空、航天、船舶、制造装备、工程机械的生产系统均为离散系统,石油化工、钢铁的生产系统属于连续系统,而锂电池的生产系统属于离散-连续混合系统。

本书主要介绍离散型生产系统,它本质上就是一类离散事件系统。

## 1.2 系统建模与仿真的基本概念

### 1.2.1 模型与仿真

在工程领域,为了理解和分析的需要,将所研究的物理系统进行一定的简化,并用适当的表现形式或规则来描述其主要特征,所得到的模仿品称为模型(model)。模型的类型非常多,典型的有:

(1) 物理模型,比如一辆实验用的物理样车;

(2) 数学模型,比如故障诊断模型、生产计划调度模型、生产性能预测模型;

(3) 图形化模型,比如框架图、流程图;

(4) 仿真模型,比如产品设计有限元分析模型、生产系统仿真模型。

模型的建立过程称为建模(modeling)。系统建模的目的在于:

(1) 更直观地描述所研究的系统;

(2) 方便进行计算,从而定量分析系统的性能指标;

(3) 实现系统的模拟运行,进行仿真实验,分析不足,优化系统设计。

客观来讲,任何模型都无法完全模拟实际的系统,为了保证模型能够建立出来,并且建模代价(时间、费用等)可接受,在建模过程中,需要进行简化或抽象,具

体措施包括:

(1) 引入各种假设,比如在设备故障建模时,假设故障失效时间服从威布尔分布;

(2) 忽略次要因素,比如在生产建模时,不考虑工艺路线的临时调整;

(3) 分层次、分视图、分尺度建模,在进行高层级分析时,简化低层级的内部结构,在进行 A 视图分析时,简化 B 视图的影响,在大的时间尺度观察系统时,忽略小时间尺度范围内的细节变化。

仿真(simulation)是对现实世界的过程或系统进行时间允许的模拟,产生并观测系统的人造历史记录,从而推断出实际系统的运行特性,再通过计算机上运行的建模工具,采用适当的仿真语言或程序,将现实系统转变为仿真模型。关于不同控制策略或设计变量对系统的影响,或是系统受到某些扰动后可能产生的影响,最好是用系统本身进行实验,但许多时候这是不可行的,原因有多方面。例如,实验费用非常高;系统本身不稳定,实验可能破坏系统的平衡,造成危险;系统的时间常数很大,实验需要很长时间;待设计的系统尚不存在等。在这样的情况下,建立系统的仿真模型是必然的选择。

对于离散事件系统而言,数学建模和仿真建模是两种常见的建模方式,它们可以对系统的性能进行定量分析和优化,计算或解析方法的求解效率高,但其数学模型一般需要进行高度抽象与简化,仅能分析单元子系统,对于复杂的系统几乎无能为力。仿真方法更适用于复杂系统的模拟分析,受结构复杂性的影响较小,仿真过程不仅可以进行验证分析,还可以与各种优化算法相结合,实现仿真优化,获得最佳的参数设计、资源分配策略和调度策略。

### 1.2.2 离散事件系统仿真

在计算机环境中,采用基于离散事件系统的仿真模型模拟其运行过程,称为离散事件系统仿真(discrete event system simulation, DESS)。其过程如下:在离散时间点模拟事件发生,改变系统的状态,触发对象的活动,该活动持续一段随机时间后结束,结束后模拟产生新事件,再次改变系统状态,并触发新活动,如此反复,直到仿真结束。

离散事件系统仿真的关键在于事件的模拟、对象活动开始与结束的模拟,以及对对象状态变迁的模拟。下面介绍状态、事件、活动三个基本概念。

#### 1. 状态

状态(status)是系统中对象的重要属性。相较于对象的其他属性,状态具有两个基本特征:一是“动态变化”,即在不同的时刻点,状态的取值可以不同。以设备对象为例,设备的名称、尺寸、位置是静态属性,一般不能称为状态,而设备是否在工作、在制工件编号、加工参数、已完工工件数量等则是状态属性。二是“多层次性”,系统的多层次性也导致状态的多层次,在不同层次观测一个系统,其状态变量

是不同的。比如一条生产线,如果将其视为一个整体,则状态变量包括在制品总量、生产线工作或停线等;如果将其细分为各个工位,则每个工位的状态变量又包括工作或停机、在制工件编号、开始加工时间等。因此,一段时间内,如果从高层次来观测,系统的状态可能静止未变,但从低层次来观测,系统的状态已发生了很多变化。

## 2. 事件

事件(event)是指引起系统状态发生变化的行为或事情。生产系统中常见的事件包括工件创建、工件进入、任务安排、开始加工、完工、工件离开、开始服务、结束服务、发生故障、故障修复、计划停机、停机重启等。事件具有时间标签,并且事件的发生都是有条件的,最基本的条件是时间条件,即只有在某时刻及之后才能发生,还可以定义更为复杂的发生条件。此外,物理系统中的事件一般不能精确确定何时会发生,比如设备故障事件、工件完工事件等。

## 3. 活动

如 1.1.1 节所述,活动(activity)是描述系统的基本要素。广义活动确定了系统内部发生变化的过程,比如生产活动、物流活动等。离散事件系统中的狭义活动一般是指同一对象的两个相邻事件之间的过程,比如设备的加工作业、工人的物料搬运作业、机器人的上下料作业等。在这段过程中,该对象无新事件发生,状态保持不变。比如设备加工某个工件,从开工到完工这段时间的状态不变。注意,如果设备在加工期间发生故障事件,则需要将原活动进行分解,得到设备加工、故障、维修、设备重新加工等几个新活动。

在物理系统中,活动会持续一段时间,比如设备加工某工件共需 5min。但在仿真系统中,活动通常也就是一个事件处理程序,事件一旦发生,触发该程序执行,活动也就结束了。为了避免混淆,本书使用“操作”(operation)一词来代替狭义活动,它更具“原子性”,而广义活动可视为多个对象在一段时间内发生的多个操作的集合。

操作不会无缘无故发生,操作的发生由事件触发,操作过程中往往会产生新事件。事件、状态、操作的关系示例如下:某时刻点,“工件进入设备”是一个事件,触发了工件的加工操作,改变了设备的状态,从“等待”变为“工作中”。工件加工完成后,产生“工件完工”事件,触发完工操作。这时有两种执行结果,一是工件的后继对象允许进入(比如机械手就绪),则搬运工件,并更改设备状态为“等待”;二是工件后继对象不可进入(比如设备后置缓冲区已满),则更改设备状态为“阻塞”。

# 1.3 生产系统建模与仿真的目标

在生产系统建模领域,有许多经典的分析与优化问题,比如车间布局规划与重构、生产线平衡分析、车间计划调度、物流路径规划、物流调度、故障分析与维修决



策等。大量学者利用运筹学(operations research, OR)方法对这些问题进行了深入研究,取得了许多重要的理论成果。然而由于实际生产系统的复杂性,这些成果往往难以直接用于解决工程问题。通过建模与仿真手段对生产系统进行分析,由于更容易模拟实际生产过程,并且分析手段全面,越来越受到企业的重视。生产系统建模与仿真的根本目的在于:

(1) 在系统布局设计阶段,通过生产与物流活动的仿真,对系统运行性能进行定量分析,提前发现问题,为系统结构设计、资源分配、方案比选等提供数据决策支持,以保证系统设计的科学性、经济性、鲁棒性。

(2) 在系统运行与持续优化阶段,建立物理生产系统的数字孪生模型,通过基于数字空间的仿真实验与优化,识别生产瓶颈,优化运行参数,评估系统在不同调度策略下的性能,确定高效的作业计划和调度方案,辅助生产决策,提高物理系统的综合运行效率。

### 1.3.1 生产系统的性能指标

对于现有或预期建设的生产系统,为了评价其性能的优劣,进而找到改善的方向,必须量化分析系统的性能指标。常见的性能指标包括:

(1) 生产率(productivity)。它是一个相对指标,通常表示为产出和投入之比。根据投入要素的数量,可以分为单要素生产率、多要素生产率和总生产率。比如,投入要素为工人,则表示为劳动生产率。

(2) 生产能力(production capacity)。生产能力是指生产系统在一段时间内,在合理的技术条件下,生产某类产品或零部件的平均数量和最大数量,简称产能。比如一条加工生产线,如果工件投入充足,运行一段时间后,生产线单位时间内产出的工件数量会稳定到一个最高水平值,这就是生产线的最大能力(极限能力)。实际情况下,未必能满负荷投入工件,平均产量一般小于最大产量。进一步还可考虑设备故障、工件缺陷、工人离岗等因素,以定义有效生产能力。

(3) 在制品(work in process, WIP)数量。在制品数量是指已投产未完工的工件数量,包括处于工作、等待、运输等状态的工件数量之和。根据著名的 Little 公式,单位时间内投入系统的工件越多,并且工件通过系统的时间越长,在制品数量就越多。在保证生产能力的情况下,在制品数量当然越少越好,因此,工件的投产时机选择和生产调度策略很重要。

(4) 通过时间  $T$ 。通过时间是指工件进出系统(设备、产线或车间)的时间差,也称为系统逗留时间,它由加工时间、运输时间和等待时间构成。一般只统计一类工件的平均通过(逗留)时间。

(5) 工件加工/运输/等待时间占比。工件加工/运输/等待时间占比是指工件通过时间中,加工/运输/等待时间的比例。显然,如果加工时间占比偏低,则说明物流效率低或生产调度不合理。

(6) 等待队长。等待队长是指在等待加工或搬运任务的缓冲区中,最大和平均等待的工件个数。等待队列决定了缓冲区的容量设计大小,并且,队列越长,说明缓冲区之后的设备是瓶颈。

(7) 订单按期完成率。订单即生产作业计划,在订单下发时一般会给定期望完成时间。由于生产异常因素,或者生产负荷过重、调度不合理,部分订单可能无法按期完成,导致订单按期完成率达不到 100%。实际情况中,订单提前太长时间完成可能也不合理,这时可以设定期望完工时间区间,并计算订单准时完成率。

(8) 直通率(first pass yield rate)。直通率是指产品从第一道工序开始一次性合格到最后道工序的比例,它与每道工序的合格率相关。

(9) 设备利用率(utilization reatio)。设备利用率是指设备实际工作时间占总时间或有效时间的比例。

(10) 设备综合效率(overall equipment effectiveness, OEE)。用来表示实际的生产能力相对于理论生产能力的比率。OEE 的计算公式:  $OEE = \text{时间开动率} \times \text{性能开动率} \times \text{合格品率} = \text{合格产品的理论加工总时间} / \text{负荷时间}$ 。

(11) 资源利用率。资源利用率是指物流车辆、人、工装、工具的利用率。

(12) 能源效率。能源效率是指单位产品能耗或单位能耗产出。

(13) 生产线平衡率(line balance rate)。生产线平衡率 =  $\frac{\text{各工序作业时间和}}{\text{工序数} \times \text{瓶颈工序时间}}$ 。

(14) 物流运输距离/成本。物流运输距离/成本是指单位时间内的物流运输距离/成本或者产品的平均物流运输距离/成本。

(15) 平均故障间隔时间(mean time between failure, MTBF)。平均故障间隔时间指系统在两相邻故障间隔期内正常工作的平均时间,也称平均无故障工作时间。

(16) 平均修复时间(mean time to repair, MTTR)。平均修复时间指系统出现故障后到恢复正常工作时平均所需要的时间。

(17) 设备可用率(availability)。可用率 =  $\frac{MTBF}{MTBF + MTTR}$ 。

如果生产系统已经存在并运行了一段时间,理论上讲,通过历史数据统计可以计算出上述性能指标,但分析过程一般非常复杂;如果生产系统尚不存在,或者虽存在但亟待重构,新系统性能则往往难以测算。现实中一般靠经验估计或者简单计算来分析生产系统的性能,结果难以令人信服,此时,建立生产系统仿真模型,通过仿真运行来计算上述指标,并给出定量评价,就非常有必要了。

### 1.3.2 建模与仿真的意义

从某种程度上讲,生产系统仿真和计算机辅助工程(CAE)具有类似的含义。在产品的设计过程中,为了提前验证产品的性能,需要利用 CAE 软件建立有限元分析模型,对产品的功能、性能与安全可靠性进行计算,对产品的工作状态和预期行

为进行模拟仿真,及早发现设计缺陷,改进和优化设计方案,证实未来产品的可用性与可靠性。同样,在生产系统设计过程中也需要开展仿真分析,基于虚拟工厂模型,通过模拟、验证、优化手段,实现低成本、快速、科学的工厂设计。当产品、工艺或物流发生变化时,基于虚拟工厂先进行重构和优化,持续保障工厂效能。因此,生产系统仿真分析可认为是针对工厂所开展的 CAE 工作。

具体而言,生产系统建模与仿真的实际意义在于:

(1) 基于虚拟环境,可以较为真实地模拟生产与物流过程,从而快速、全面地获得生产运行数据。在物理生产系统中,能获取的数据往往是不全面的,尽管各种物联网技术、生产管控技术等正在逐步推广应用,但生产过程中仍然欠缺有价值的数  
据,这个时候,仿真技术就为数据的获取提供了另外一个手段。通过工厂建模,模拟设施、工艺、物流、人、控制与调度逻辑,导入或模拟生产计划,模拟各种随机因素,然后进行仿真运行,可以在较短的时间内得到大量运行过程数据,比如,在 1min 内完成 1 年生产订单的运行模拟,得到几千万条数据。

(2) 基于数据分析,定量评估分析工厂的性能,包括能力分析和效率分析。1.3.1 节定义了生产系统的 17 个性能指标,通过对它们的分析,可以迅速发现系统运行中存在的问题和有待改进之处,并及时进行调整与优化,减少后续生产执行环节对于物理系统的更改与返工次数,从而有效降低成本、缩短工期、提高效率。

(3) 通过指导性分析(prescriptive analysis),主动寻找优化方案并仿真验证。仿真分析和优化算法可以进行各种形式的融合,比如仿真分析与实验设计的结合、仿真分析与遗传算法的结合等。算法提供了基本的优化逻辑,而仿真则为算法中的一些关键步骤提供了“黑箱”式计算方式。比如,遗传算法中需要计算个体的适应度,然而对于复杂问题,很难给出适应度计算的解析公式,这时候就可以通过仿真来得到适应度值。

## 1.4 生产系统建模与仿真的相关理论与技术

### 1.4.1 概率论与统计分析

生产系统中存在许多随机事件,比如设备故障、工件加工完成、订单到达等,为了描述它们需要用到几种常见的概率分布,见表 1-1。

表 1-1 几种常见的概率分布及其统计特征

分布	参数	分布律或概率密度	数学期望	方差
均匀分布	$a < b$	$f(x) = \begin{cases} \frac{1}{b-a}, & a < b \\ 0, & \text{其他} \end{cases}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$

续表

分布	参数	分布律或概率密度	数学期望	方差
指数分布	$\theta > 0$	$f(x) = \begin{cases} \frac{1}{\theta} e^{-\frac{x}{\theta}}, & x > 0 \\ 0, & \text{其他} \end{cases}$	$\theta$	$\theta^2$
泊松分布	$\lambda$	$P\{x=k\} = \frac{\lambda^k e^{-\lambda}}{k!}$	$\lambda$	$\lambda$
正态分布	$\mu$ $\sigma > 0$	$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	$\mu$	$\sigma^2$
威布 尔分 布	$\eta > 0$ $\beta > 0$	$f(x) = \begin{cases} \frac{\beta}{\eta} \left(\frac{x}{\eta}\right)^{\beta-1} e^{-\left(\frac{x}{\eta}\right)^\beta}, & x > 0 \\ 0, & x \leq 0 \end{cases}$	$\eta \Gamma\left(\frac{1}{\beta} + 1\right)$	$\eta^2 \left\{ \Gamma\left(\frac{2}{\beta} + 1\right) - \left[ \Gamma\left(\frac{1}{\beta} + 1\right) \right]^2 \right\}$
三角分布	$a \leq c$ $\leq b$	$f(x a,b,c) = \begin{cases} \frac{2(x-a)}{(b-a)(c-a)}, & a \leq x \leq c \\ \frac{2(b-x)}{(b-a)(b-c)}, & c < x \leq b \end{cases}$	$\frac{a+b+c}{3}$	$\frac{a^2+b^2+c^2-ab-bc-ca}{18}$
伽马分布	$\alpha > 0$ $\beta > 0$	$f(x) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, & x > 0 \\ 0, & x \leq 0 \end{cases}$	$\frac{\alpha}{\beta}$	$\frac{\alpha}{\beta^2}$
$n$ 阶 埃尔 朗分 布	$n$ 为 正整 数 $\lambda > 0$	$f_n(x) = \begin{cases} \frac{\lambda^n}{(n-1)!} x^{n-1} e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0 \end{cases}$	$\frac{n}{\lambda}$	$\frac{n}{\lambda^2}$

如果随机变量的参数未知,就需要用到参数估计这一统计推断方法。参数估计(parameter estimation)是根据从总体中抽取的随机样本来估计总体分布中未知参数的过程。参数估计要处理两个问题:①求出未知参数的估计量;②在一定可信度下求出估计量的精度。可信度一般用概率表示,如95%;精度用估计量与被估参数之间的接近程度或误差来度量。对估计值的评价标准有3个:

(1) 无偏性。无偏性是指估计量抽样分布的数学期望等于总体参数的真值。无偏性的含义是:估计量是一随机变量,对于样本的每一次实现,由估计量算出的估计值有时可能偏高,有时可能偏低,但这些估计值平均起来等于总体参数的真值。在平均意义下,无偏性表示没有系统误差。

(2) 有效性。有效性是指估计量与总体参数的离散程度。如果两个估计量是无偏的,那么离散程度较小的估计量相对而言是较为有效的。离散程度是用方差度量的,因此在无偏估计量中,方差越小越有效。

(3) 一致性。一致性又称为相合性,是指随着样本容量的增大,估计量越来越

接近总体参数的真值。

点估计和区间估计是两个基本的参数估计方法。点估计(point estimation)是依据样本估计总体分布中所含的未知参数或未知参数的函数。通常它们是总体的某个特征值,如数学期望、方差和相关系数等。点估计问题就是要构造一个只依赖于样本的量,作为未知参数或未知参数的函数的估计值。构造点估计常用的方法有:矩估计、最大似然估计、最小二乘估计和贝叶斯估计。区间估计(interval estimation)是针对抽取的样本,根据一定的正确度与精确度要求,构造出适当的区间,作为总体分布的未知参数或参数的函数的真值所在范围的估计。求置信区间常用的3种方法是:①利用已知的抽样分布;②利用区间估计与假设检验的联系;③利用大样本理论。

以概率论为基础,对所观测数据进行的分析称为统计分析。常用的统计分析方法有聚类分析、因子分析、相关分析、对应分析、回归分析、方差分析等。

(1) 聚类分析(cluster analysis)。聚类分析是指将物理或抽象对象的集合分为由类似的对象组成的多个类的分析过程。通过聚类将数据划分为不同簇,同一个簇中的对象有很大的相似性,而不同簇间的对象有很大的相异性。聚类分析是一种探索性的分析,在分类过程中,不必事先给出一个分类标准,能够从样本数据出发,自动进行分类。聚类分析所使用的方法不同,常常会得到不同的结论。

(2) 因子分析(factor analysis)。因子分析是指研究从变量群中提取共性因子的统计技术。通过因子分析,可从大量数据中寻找内在的联系,减少决策的困难。因子分析主要有两大类方法:一是探索性因子分析(exploratory factor analysis);二是验证性因子分析(confirmatory factor analysis)。

(3) 相关分析(correlation analysis)。相关分析是研究现象之间是否存在某种依存关系,并对具体有依存关系的现象探讨其相关方向及相关程度。相关关系是一种非确定性的关系。例如,用 $X$ 和 $Y$ 分别表示一个人的身高和体重,则 $X$ 与 $Y$ 显然有关系,而又没有确切到可由其中的一个去精确地决定另一个的程度。

(4) 对应分析(correspondence analysis)。对应分析也称关联分析、R-Q型因子分析,通过分析由定性变量构成的交互汇总表来揭示变量间的联系,可以揭示同一变量的各个类别之间的差异,以及不同变量各个类别之间的对应关系。对应分析的基本思想是将一个联列表的行和列中各元素的比例结构以点的形式在较低维的空间中表示出来。

(5) 回归分析(regression analysis)。回归分析是确定两种或两种以上变数间相互依赖的定量关系的一种统计分析方法。回归分析运用十分广泛,其按照涉及的自变量的多少,可以分为一元回归分析和多元回归分析;按照自变量和因变量之间的关系类型,可以分为线性回归分析和非线性回归分析。

(6) 方差分析(analysis of variance, ANOVA)。方差分析是用于两个及两个以上样本均数差别的显著性检验。由于各种因素的影响,研究所得的数据呈现波