



天津商業大學

教材建設



财经类专业“十四五”规划新形态教材

RPA财务机器人应用

余冰冰 白默 倪江崑 / 主编

厦门网中网软件有限公司 / 组编



立信会计出版社

LIXIN ACCOUNTING PUBLISHING HOUSE



财经类专业“十四五”规划新形态教材

RPA财务机器人应用

余冰冰 白默 倪江崑 / 主编

李长山 杨行翀 叶银兰 / 副主编

厦门网中网软件有限公司 / 组编



立信会计出版社

LIXIN ACCOUNTING PUBLISHING HOUSE

图书在版编目(CIP)数据

RPA 财务机器人应用 / 余冰冰, 白默, 倪江崑主编
· 一上海: 立信会计出版社, 2024. 4
ISBN 978-7-5429-7505-8

I. ①R… II. ①余… ②白… ③倪… III. ①财务管理—专用机器人 IV. ①F275②TP242.3

中国国家版本馆 CIP 数据核字(2024)第 041131 号

策划编辑 王斯龙 王秀宇
责任编辑 王斯龙 王秀宇
美术编辑 吴博闻

RPA 财务机器人应用

RPA CAIWU JIQIREN YINGYONG

出版发行 立信会计出版社
地 址 上海市中山西路 2230 号 邮政编码 200235
电 话 (021)64411389 传 真 (021)64411325
网 址 www.lixinaph.com 电子邮箱 lixinaph2019@126.com
网上书店 <http://lixin.jd.com> <http://lxkjcs.tmall.com>
经 销 各地新华书店

印 刷 常熟市人民印刷有限公司
开 本 787 毫米×1092 毫米 1/16
印 张 18
字 数 393 千字
版 次 2024 年 4 月第 1 版
印 次 2024 年 4 月第 1 次
书 号 ISBN 978-7-5429-7505-8/F
定 价 49.80 元

如有印订差错,请与本社联系调换

前 言

新一轮技术革命,特别是数智技术的飞速发展,对会计行业产生了深远影响,推动了会计理论、职能、组织方式和工具的变革。我国财政部在2021年印发的《会计行业人才发展规划(2021—2025年)》和《会计信息化发展规划(2021—2025年)》中指出,RPA财务机器人等自动化工具的推广,使会计工作数字化转型得以深入,同时也使得行业对数智化复合型新型人才的需求愈发迫切。然而,现行高校财会类专业课程体系却难以满足这一市场需求,传统的财会专业人才培养方案也未能充分涵盖数智化技能和思维的培养,尤其在人工智能和RPA技术方面。尽管国内部分院校已尝试开设相关RPA课程,但多偏向于编程或技术工具的使用,知识难度较大,无法很好地融入财会类课程内容体系,不利于教师开展数智化的教学和学生学习,进而影响了财会类专业数智化人才的培养以及高校会计专业教育的数智化发展。基于此,本书应运而生。

本书共包含4篇,分别是理论篇、流程分析与设计篇、流程开发与部署篇、流程测试与运维篇,其中理论篇主要介绍了RPA财务机器人概论、全生命周期、项目关键文档及项目团队及职责;流程分析与设计篇主要阐述了财务机器人典型应用场景,从业务流程分析到自动化流程设计,向学生充分展现流程为何实施及如何实施RPA;流程开发与部署篇以华为WeAutomate为开发工具,详细阐述了RPA在数据处理、文本处理、UI自动化的三大板块的详细开发与部署;流程测试与运维篇主要是描述了流程测试与运维的相关内容。

本书特色如下。

1. 按照RPA项目实施全生命周期设置教学框架

本书围绕RPA项目实施全生命周期展开教学,分为理论篇、流程分析与设计篇、开发工具应用篇以及流程运维与优化篇,逐步深入讲解RPA项目全过程。

2. 精选五大经典RPA财务机器人应用案例,贴近企业实际

本书精选来自资金管理、采购到付款、订单到收款、费用报销、总账到报表的行业经典RPA财务机器人应用案例,以贴近实际的企业运营场景,以及RPA在数据处理、文本处理及UI自动化的三大经典应用。

3. 真实情景案例的开发实训

本书通过真实情景案例的开发实训,使读者能够从理论学习、案例应用到 RPA 工具的开发实践,全方位地掌握 RPA 技术。

4. 融入课程思政

本书在每章结尾,结合 RPA 财务机器人的行业现状、应用案例等融入了结合思政启示的思考题,致力于让读者掌握知识的同时,促进职业素养的提升。

本书提供华为 RPA 设计器教育版安装程序、教材内案例源程序等开发资源;也提供课程开设的教学大纲、教学课件、思政元素案例素材、教学案例讲解、教学方式设计及课程考核设计等教学资源;还将配套 RPA 项目设计交付文档样例。我们相信,通过学习本书的知识与软件平台的功能,读者将能够更好地理解和实践财务领域的机器人流程自动化,从而提升效率、降低成本,并实现业务的卓越发展。

本书适用但不限于作为普通高校本科会计学、财务管理、审计学、内部审计等财会类专业的 RPA 财务机器人相关课程的教材,也可作为企业业务人员和 IT 人员实施和管理 RPA 项目的学习用书。

本书由余冰冰、白默、倪江崑任主编,李长山、杨行翀、叶银兰任副主编,厦门网中网软件有限公司提供了技术支持。感谢昆仑银行 RPA 负责人郭宇博先生,从企业实践角度出发,对本书的编写提供了有力支持,也感谢立信会计出版社对本书出版提供的帮助。

限于作者水平,对于书中的疏忽及错漏之处,诚挚地希望广大读者给予批评指正。

编者

2024 年 4 月

目 录

理论篇

第一章 RPA 财务机器人概论	3
第一节 RPA 概述	4
第二节 RPA 产品	13
第三节 财务机器人	25
第四节 财务机器人的发展趋势	30
章节测试	31
第二章 财务机器人全生命周期	33
第一节 总体介绍	34
第二节 概念验证阶段	35
第三节 财务机器人实施阶段	39
第四节 RPA 卓越中心	45
章节测试	49
第三章 财务机器人项目关键文档	51
第一节 财务机器人项目交付	52
第二节 可行性分析文档	53
第三节 流程定义文档	55
第四节 流程详细设计文档	57
第五节 流程测试报告	60
章节测试	63
第四章 财务机器人项目团队及职责	65
第一节 RPA 项目团队组建	65
第二节 RPA 人才能力要求	67
第三节 财务人员数智认知能力提升	70

章节测试	77
流程分析与设计篇	
第五章 资金管理	81
第一节 RPA 在资金管理中的应用	82
第二节 现金归集流程分析	84
第三节 银行流水采集流程分析	90
第四节 银行流水采集自动化流程设计	93
章节测试	99
第六章 采购到付款	101
第一节 RPA 在采购到付款中的应用	101
第二节 发票验真流程分析	104
第三节 发票验真自动化流程设计	109
章节测试	115
第七章 订单到收款	117
第一节 RPA 在订单到收款中的应用	117
第二节 发票开具流程分析	120
第三节 发票开具自动化流程设计	125
章节测试	131
第八章 费用报销	133
第一节 RPA 在费用报销中的应用	133
第二节 报销付款业财系统核对流程分析	137
第三节 报销付款业财系统核对自动化流程设计	140
章节测试	145
第九章 总账到报表	147
第一节 RPA 在总账到报表中的应用	147
第二节 银企对账自动化流程分析	152
第三节 银企对账自动化流程设计	156
章节测试	161

流程开发与部署篇

第十章 财务机器人开发	165
第一节 华为 WeAutomate 的简介与结构	165
第二节 华为 WeAutomate 的功能与使用	170
第三节 财务机器人开发实践	173
章节测试	179
第十一章 数据处理自动化	181
第一节 数据处理自动化概述	181
第二节 数据处理自动化场景	183
第三节 RPA 在数据处理中的应用	190
章节测试	201
第十二章 文本处理自动化	203
第一节 文本处理自动化概述	203
第二节 文本处理自动化场景	205
第三节 RPA 在文本处理中的应用	214
章节测试	219
第十三章 UI 自动化	221
第一节 UI 自动化概述	221
第二节 UI 自动化场景	223
第三节 RPA 在 UI 中的应用	227
章节测试	233

流程测试与运维篇

第十四章 财务机器人测试与运维	239
第一节 财务机器人测试	239
第二节 财务机器人部署	244
第三节 财务机器人运维	248
章节测试	251

第十五章 RPA 在其他领域的核心应用场景.....	253
第一节 RPA 在制造领域的应用	254
第二节 RPA 在医疗领域的应用	256
第三节 RPA 在教育领域的应用	261
第四节 RPA 在物流领域的应用	263
第五节 RPA 在金融领域的应用	266
章节测试	271
附录 1 自动化流程需求评审表	273
附录 2 自动化流程需求评审标准	276

第十章

财务机器人开发

——以华为 WeAutomate 为例



本章目标

1. 了解财务机器人开发流程。
2. 掌握华为 WeAutomate 的结构与功能。
3. 掌握利用华为 WeAutomate 设计器开发财务机器人。

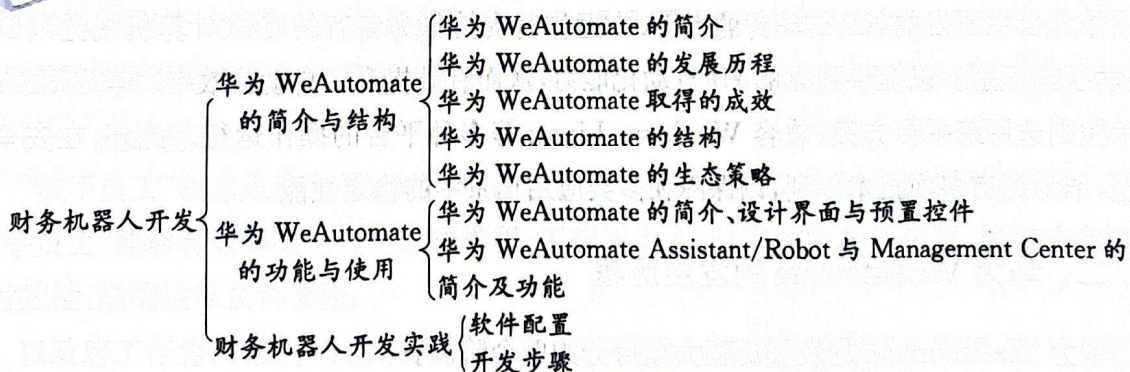


本章概览

本章主要介绍财务机器人的开发流程,华为 WeAutomate 的功能、基本使用方法,并以实际的财务机器人场景为例,介绍如何使用华为 WeAutomate 开发。



章节导航



第一节

华为 WeAutomate 的简介与结构

作为全球领先的通信设备供应商,华为在 RPA 领域也拥有深厚的技术积累和应用经验。华为 WeAutomate 是华为推出的一款智能 RPA 平台,集成了流程设计、自动化执行、智能决策等功能,能够为企业提供全方位的流程自动化解决方案。华为 WeAutomate 具有的

特色包括:①强大的流程设计能力,支持拖拽式流程设计,让流程设计更加简单易用。②高效的自动化执行能力,能够实现快速、准确的流程执行。③智能决策功能,能够根据业务需求进行智能决策,进一步优化流程。华为 WeAutomate 已在金融、制造、物流等多个行业得到了广泛的应用。华为 WeAutomate 作为华为旗下的 RPA 软件,可以实现“发现→设计→运行→管理→协作→运营&运维”的全生命周期管理。本章将深入介绍华为 WeAutomate 的应用结构与具体功能,以帮助您全面了解及运用华为 WeAutomate 实现企业流程自动化。

一、华为 WeAutomate 的简介

2015年,华为首次引入了RPA,将其应用于全球约20个区域和交付运维共享交付中心。这项技术主要用于性能告警、数据统计等方面,旨在提升客户服务质量和工作效率。通过数字机器人在无线话统优化领域的运用,涵盖作业、搬运和指令三类机器人,这项技术实现了从现网KPI监控、分类处理到问题排查作业全流程的自动化处理。这一自动化处理不仅降低了人工操作的频次和时间,还在网络问题实时监控和自动优化方面取得了显著成效,有效缩短了网络问题的响应时间,提高了客户服务满意度。然而,由于电信行业软件应用场景与通用商业软件有所不同,许多RPA供应商未能满足华为在运营商网络软件方面的各类要求。例如,运营商网络软件的对话框可能包含动态信息,如IMSI、MSISDN、小区信息、时间等。而当时的RPA供应商并不支持动态窗口信息的抓取,这导致很多场景无法通过RPA来提升服务质量。基于此,华为从2017年开始自研RPA,选择了相对于Java和C#更为简单易用的Python作为开发语言,以及使用了Springboot框架来开发流程设计、运行、管理的端到端全自动化平台。

作为自主研发的RPA,华为的RPA在运作时支持华为专有的CV(计算机视觉)、OCR(光学字符识别),以提供强大的UI自动化能力,从而有效提升了自动化效率。华为RPA的技术栈则选用跨平台方案,兼容Windows、Linux等多种平台的操作运行。此外,在安全能力上,华为拥有多项技术专利,以保障在多类应用场景中的稳定性能。

二、华为 WeAutomate 的发展历程

华为 WeAutomate 的发展历程大概可分为四个阶段。

(一) 初创期:2017年—2018年年初

华为 WeAutomate 起初从零自研 RPA 执行器,通过屏幕坐标的自动化和 XML 实现 RPA 流程处理,并构建管理中心和设计器雏形。此阶段,华为自研 RPA 已在全球 100 多家代表处支持客户服务场景中使用,通过各类电信软件和多场景的大量使用,逐渐提升 RPA 软件质量。

(二) 发展期:2018年中后期—2019年年初

在发展期,华为自研 RPA 设计器,实现通过拖拉拽的方式完成 RPA 流程的编排,并支持流程录制;同时投资低代码开发平台(ADC),提供模型编排、服务编排、流程编排、页面编

排、集成编排等端到端编排能力,并与 RPA 集成,加快了业务开发效率。在此阶段,华为内部多个职能部门使用了约 1 万个数字机器人,如财经、HR、制造、供应链等部门,实现了诸多公司内部流程末端的自动化处理,提升了处理效率。

(三) 探索期:2019 年中后期—2020 年

在经过华为内部不同职能部门的使用磨合,RPA 设计器、执行器和管理中心在稳定性和安全性上已具备商用能力,并开始面向政企,赋能各行各业。华为在与客户和伙伴的交流和实践中发现,除了流程自动化的挑战,客户还面临着信息化程度不高以及智能化水平不足的问题。单纯的 RPA 产品无论是从能力还是架构上都无法有效打通企业数字化转型的‘最后 1 公里’。于是华为开始将已运用在运营商领域的成熟的 AI 和大数据能力引入,增加大数据流处理和批处理编排能力,并将 AI 能力集成到 RPA 设计器中,如 OCR、NLP 等 AI 算法模型,技术上达到了开箱即用。

(四) 成熟期:2021 年至今

华为 WeAutomate 产品升级为超级自动化平台,聚焦政务、财务领域,通过“自动化+”的方式,将 RPA 与低代码、AI、大数据开发平台整合,能够构筑更强的智能自动化能力和场景解决方案。同时面向 Office 办公人员,华为 WeAutomate 提供了以业务中心视角开发的设计器 StudioE,进一步降低使用门槛,通过更简单、更直观的业务编排方式,使更多的用户能快速上线应用。

三、华为 WeAutomate 取得的成效

发展至今,华为 WeAutomate 在实际工作中取得了显著的应用成果。在过去的 5 年里,华为 RPA 服务于大型国有企业的数字化改革,提升了政府政务服务水平,推动了高水平智慧医疗服务建设等方面。在不同行业、不同领域,华为 WeAutomate 都发挥着关键作用。

(一) 效率层面

“数字员工”概念是华为 WeAutomate 经过了一系列流程设计和部署的最终形态。通过“数字员工”能够有效解决工作中数据堆积、工作压力大、时间成本高等问题,提高流程运转的时效性、准确性以及科学性。

以政府工作为例,数字化劳动力投入后,能够实现一体化的项目表格填报,通过自动匹配相同字段,显著减少了人工录入信息的工作量,取而代之的是系统自动填写数据。此外,“数字员工”还能协助工作人员完成表格填报、整理生成主表并进行自动发送。原本需要 5 天完成的工作现在仅需 5 小时,整体表格梳理工作更加高效,数据准确率几乎达到 100%。

以医疗机构为例,药剂仓库的发票收集和整理需要耗费大量的人力和时间。通过引进“数字员工”,医疗机构可实现无人看守的仓库发票验真和对比工作,将原来 200 张发票入库需要 6~7 小时的时间缩短至仅需“数字员工”处理 20 分钟。整项工作处理时间缩短远超过 80%,大幅提高了业务效率及数据准确性。医疗机构运用华为“数字员工”所取得的成效示意如图 10-1 所示。

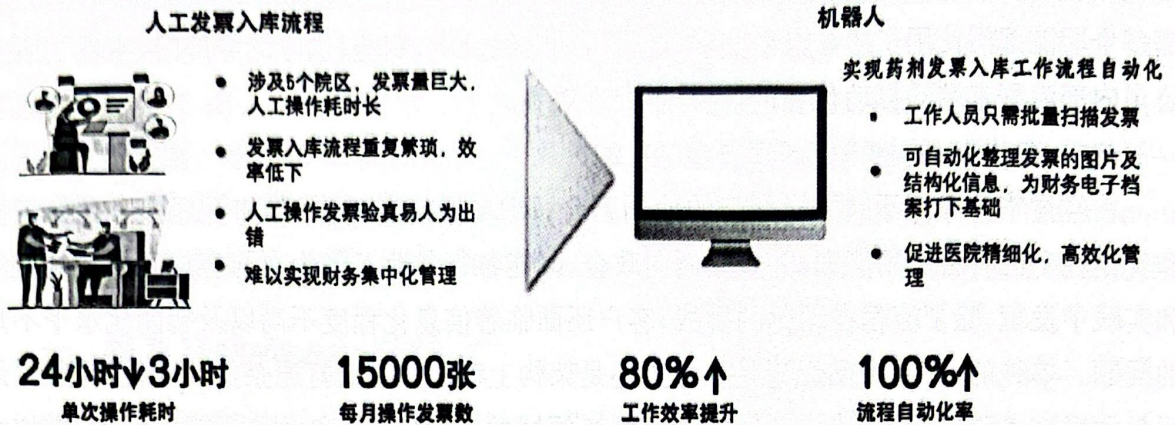


图 10-1 医疗机构运用华为“数字员工”所取得的成效

(二) 收益层面

在衡量 RPA 项目投资与收益时,我们经常会使用 ROI(投资回报率)这一关键指标进行评判。据 2020 年度 Forrester 对四家采用华为 WeAutomate 的企业进行调查研究显示,华为 WeAutomate 的 ROI 为 116%。在考虑风险因素的情况下,华为 WeAutomate 能够在产品设计检查效率、制造流程效率、物流、单据信息核对等方面带来收益,累计创收超 1 800 万元。研究中还发现了存在的非量化收益,包括华为 WeAutomate 能够提升员工的体验感以及员工留存率、减少自身产品的质量问题的提升产品交付速度,以及在流程创新的途中不断为组织带来变革。

四、华为 WeAutomate 的结构

华为 WeAutomate 的结构包括 Studio(机器人流程设计器)、Management Center(机器人管理中心)和 Assistant/Robot(机器人执行器),它们之间的关系可以类比为电影编剧、导演和演员的关系。Studio 根据需求设计和实现 RPA 自动化流程(类似于编剧设定场景和对白),Assistant/Robot 负责执行 Studio 设计好的自动化流程(类似于演员完成根据剧本完成表演),Management Center 负责调度和编排各个自动化流程(类似于导演现场调度演员的表演)。

RPA 主要专注于处理那些涉及标准数字输入的流程,这些流程具有高度手动性、重复性、基于规则的特点,并且系统异常率相对较低。对于传统的工作流自动化工具,软件开发人员会使用内部应用程序编程接口(API)或专用语言生成一系列操作,作为自动化任务或者后端系统的接口。相反,RPA 系统通过观察用户在应用程序的图形用户界面(GUI)中执行该任务,然后通过直接在 GUI 中重复这些任务来执行自动化操作。一方面,这可以降低在没有 API 的产品中使用自动化的障碍。另一方面,业务的快速发展和迭代使基于 API 的开发方式很难跟上业务发展的需求。与 API 集成的开发方式相比,使用 RPA 可显著提高开发速度,其发展进步可谓是数量级的飞跃。当然,RPA 在实施中也存在一些风险。例如,

RPA 为现有软件提供了一个更为复杂的维护环境,在某些场景下,部分应用使用图形用户界面来部署 RPA 流程并不是最高效和稳定的。总体来说,RPA 被认为是数字化转型最有效的方法之一,RPA 解决方案可看作虚拟机器人劳动力,其操作管理由业务线(仅由 IT 支持)进行,就像人工劳动力一样。

华为 WeAutomate 端侧软件 Studio、Assistant/Robot 一般独立单机部署和使用,能够实现轻量级的用户操作自动化及机器人调度运行。其中华为 WeAutomate 设计器提供图形化机器人脚本编排能力,支持录制和回放,封装上通过 Electron 将 nodeJS 实现的 WEB 应用打包成桌面应用;华为 WeAutomate RPA Assistant/Robot 软件架构保持一致,一套软件支持两种运行模式,助手程序提供运行页面,同时和设计器复用同一套执行程序。

华为 WeAutomate 的主要用户是客户或合作伙伴的流程开发人员、业务管理员与系统管理员;其中流程开发人员通过设计器对目标应用进行录制、编排与回放等,并通过执行器/连接器、管理中心在其他 PC 机、服务器上进行部署和调度。华为 WeAutomate 三件套与周边系统/人员的关系如图 10-2 所示。

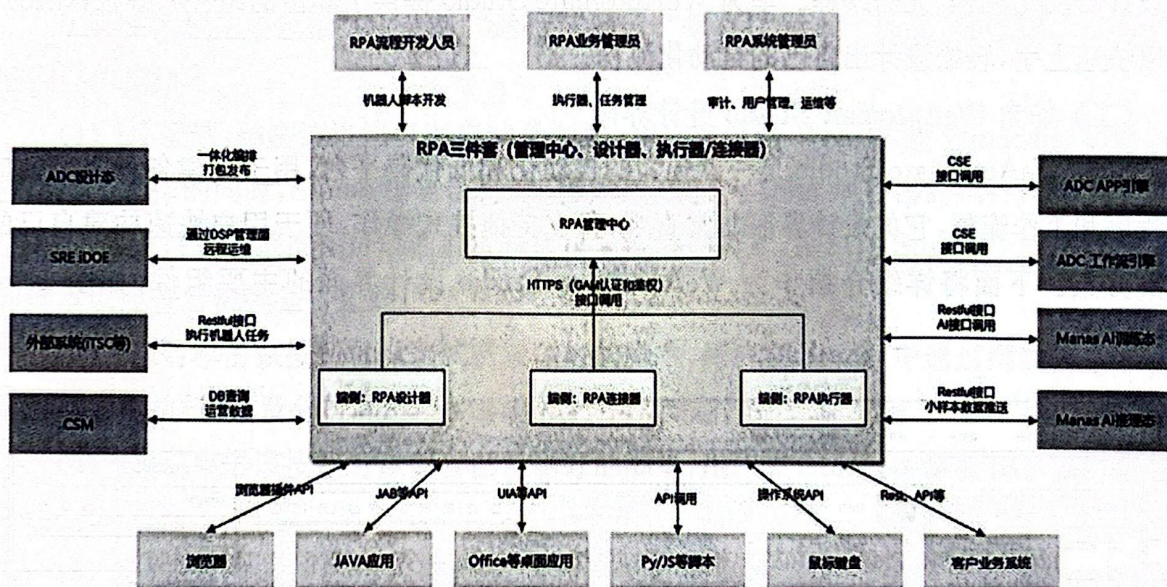


图 10-2 RPA 三件套与周边系统/人员的关系

五、华为 WeAutomate 的生态策略

要助力各大政企实现数字化转型,不仅需要过硬的技术和优秀的落地能力,更需要携手产业上下游的伙伴,共同打造优质服务帮助客户实现业务价值,聚焦提供具备竞争力的产品及套件,坚持“被集成”的生态策略,通过赋能生态,由合作伙伴面向海量客户提供端到端业务咨询、开发及运营服务。目前华为已建立完整的伙伴支撑体系,涵盖伙伴认证、伙伴赋能、联合拓展、伙伴激励全生命周期,未来会持续加大对合作伙伴的投入力度,将最新、最具优势的资源和支

收益。华为 WeAutomate 的生态伙伴已呈现百花齐放的状态,包括具有行业影响力的德勤咨询及普华永道,具有聚焦实施能力的软通动力、中软国际、神州数码、德成技术等,新发展的伙伴有翰智、西辰软件、元年科技等。

第二节 华为 WeAutomate 的功能与使用

华为 WeAutomate 包含 Studio、Assistant/Robot 及 Management Center,下面将详细介绍其功能与使用方法。在应用华为 WeAutomate 时,Studio 的重要程度最高,因此将详细介绍。

一、华为 WeAutomate 的简介、设计界面与预置控件

(一) 华为 WeAutomate Studio 简介

华为 WeAutomate Studio 是一款可视化的智能自动化设计工具,用户可以通过它来轻松设计自动化流程,无需编程。华为 WeAutomate Studio 提供了丰富的组件和模板,让用户可以快速上手,轻松设计出自己的自动化流程。

(二) 华为 WeAutomate Studio 设计界面

华为 WeAutomate Studio 是一款强大的自动化和低代码平台,用于创建各种类型的应用程序和 workflows,它的设计界面非常直观,且支持拖拽式操作,便于用户快速构建自己的解决方案。下面将详细介绍华为 WeAutomate Studio 设计界面的主要组件,如图 10-3 所示。

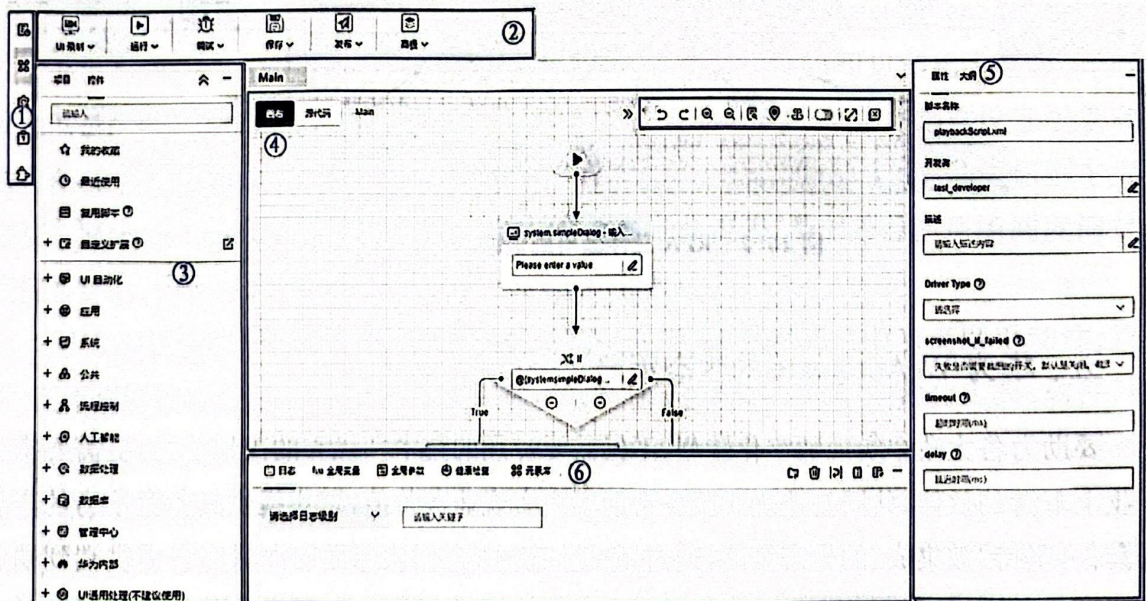


图 10-3 WeAutomate Studio 设计界面

1. 区域①:菜单栏区域

菜单栏区域主要包括“开始”“设计”“设置”“帮助”和“扩展管理”五个一级菜单。这些菜单的主要功能如下:

(1) 开始: Studio 的欢迎页面, 主要用于打开已有项目、导入已录制好的脚本、创建项目、查看最近编辑过的项目以及直接通过模板新建项目。

(2) 设计: Studio 的流程设计页面, 主要用于开发 RPA 业务流程, 并进行调试验证。

(3) 设置: Studio 的设置页面, 主要用于设置 Studio 的常规、许可证、Web 驱动、管理中心、快捷键、编排模式等配置。用户可根据实际场景需要, 设置 Studio 的相关信息。

(4) 帮助: Studio 的帮助中心页面, 主要用于打开帮助文档、进入论坛、查看 I Learning 课程等。用户在使用 Studio 的过程中, 可查看对应的帮助信息以及论坛信息, 获取所需知识和帮助。

(5) 扩展管理: Studio 的扩展包管理页面, 主要用于查看已安装的扩展包、安装所需的扩展包等。

2. 区域②:工具栏区域

工具栏区域主要包括“UI 录制”“运行”“调试”“保存”“发布”“高级”等按钮。这些按钮的主要功能如下:

(1) UI 录制: 单击该按钮可以启动录制器, 用户在录制 Web/本地 Windows/本地 Java 应用程序操作的同时, Studio 会自动将对应的操作转换为 RPA 执行器的执行脚本。

(2) 运行: 单击该按钮可以回放或执行已录制好的脚本, 用户在 Studio 中回放已录制好的脚本以便查看流程是否正确。

(3) 调试: 单击该按钮可以调试已录制好的脚本, 用户在 Studio 中通过调试脚本以验证脚本是否可执行, 以及分析和解决脚本在执行中的问题。

(4) 保存: 单击该按钮可以保存已录制好的脚本, 用户在 Studio 中保存脚本以防录制信息丢失。

(5) 发布: 单击该按钮可以将已录制和调试好的脚本发布到机器人管理中心或本地磁盘。

(6) 高级: 单击该按钮可以在项目内搜索子模块或检查项目和脚本, 用户可通过搜索功能快速定位到对应的源代码位置对应的控件, 或通过检查项目和脚本了解。

3. 区域③:项目和控件展示区域

项目和控件展示区域主要用于展示项目和可用控件的相关信息, 用户在 Studio 中开发时可以将所需的控件拖拽至画布区域。区域: 画布和源代码编辑区域, 主要用于开发 RPA 业务流程。

4. 区域④:画布工具区域

画布工具区域主要包括“撤销”“恢复”“放大”“缩小”“概览”“定位到开始节点”“自动布局”“启用自动布局”“收起侧边栏”“清空”等按钮。用户在画布中开发 RPA 流程时, 可根据

需要选择对应的工具进行操作。这些按钮的主要功能如下：

(1) 撤销：回到上一步操作的状态。

(2) 恢复：当停留在之前的操作时，单击恢复可以回到之前操作的下一步，直至所做的操作是最新的。

(3) 放大、缩小：对画布中的 workflow 进行放大和缩小。

(4) 概览：查看画布的整体缩略图。

(5) 定位到开始节点：将开始节点定位到画布中间位置。

(6) 自动布局：让 workflow 显示在画布的正中间并自动整理流程线路。

(7) 启用自动布局：自动让 workflow 显示在画布的正中间和整理流程线路。

(8) 收起侧边栏：收起画布两边的侧边栏。

(9) 清空：清空当前画布上的所有控件。

5. 区域⑤：属性区域

在画布中单击某一控件时，用户可选择“属性”页签查看和编辑控件相关属性，抑或选择“大纲”页签查看该控件的功能描述、参数以及使用示例说明。

6. 区域⑥：调试区域

调试区域用户在运行或调试脚本时，可在该区域查看脚本的运行日志，设置调试脚本所需的全局变量和全局参数，对脚本进行健康检查，抑或查看脚本相关的元素。

(三) 华为 WeAutomate Studio 预置控件

华为 WeAutomate Studio 中会预置用户常用的控件列表，用户在开发 RPA 流程中可根据需求灵活取用，控件的类别及功能列表如表 10-1 所示，用户可以根据自己的需求，选择合适的组件来设计自动化流程。

表 10-1 控件类别及功能列表

控件类别	控件功能
UI 自动化	针对 Web 应用、桌面应用和基于图像应用可使用的控件，如打开网页、鼠标单击、图片比对等
应用	针对 Excel 表格、Word 文档、邮件可使用的控件，如 Excel 表格读取、打开 Word、删除邮件等
系统	针对剪贴板、时间处理、操作系统等可使用的控件，如保存到剪贴板、截图等
公共	针对文本处理、调用等可使用的控件，如正则搜索、调用 C# 等
流程控制	针对流程控制、异常处理等可使用的控件，如退出循环、抛出异常等
人工智能	针对 Manas 引擎、EI-OCR 等可使用的控件，如图像识别、智能分类识别等
数据处理	针对二维数据处理、二维码、数学运算等可使用的控件，如获取表格数据、生成二维码、数学运算等
数据库	针对关系型数据库、非关系型数据库等可使用的控件，如连接数据库、向 MongoDB 插入数据等

(续表)

控件类别	控件功能
管理中心	针对管理中心中队列操作、人机协同等可使用的控件,如写入数据、创建人机协同任务等
华为内部	针对消息可使用的控件,如发送短信、语音电话等
UI通用处理	针对网页等可使用的控件,如追加文本等

二、华为 WeAutomate Assistant/Robot 与 Management Center 的简介及功能

华为 WeAutomate Assistant/Robot 作为一个助手工具,随时待命执行编排好的流程。华为 WeAutomate Assistant/Robot 可以执行本地计算机的自动化流程包,也可以接收华为 WeAutomate Management Center 的命令执行相应的自动化流程包。

华为 WeAutomate Management Center 作为一个集成作业调度管理中心,统一管理机器人流程资产、连接器、执行器等资源,调度机器人执行任务、监视流程执行状态和统一敏感信息安全加密等,同时能够为开发提供低代码 APP 开发能力,方便设计人机交互场景。连接器,即用于远程连接执行器的机器。无人值守类型执行器执行机器人任务前,连接器会远程连接至执行器完成屏幕解锁操作,连接器需安装于华为 WeAutomate Management Center。

RPA 管理中心具有云化管理中心(OC)和轻量化管理中心(OP)两种形态,OC(On Cloud)是由华为统一搭建,用户申请权限后即可使用。OP(On Premises)则需用户自行搭建配置并维护。

华为 WeAutomate Management Center 的主要功能包含集中调度、管理和监控所有 Assistant/Robot 的平台;存储可重用组件、资产,以及进行任务管理和配置执行器;提供低代码 App 开发平台,方便设计人机交互场景。

第三节

财务机器人开发实践

本节将通过华为 WeAutomate 实现场景,利用 Web 录制的技术,录制一个能够登录华为技术支持网站下载华为资料电子文档管理软件 ICSLite 到本地 D:\file 目录下并解压的流程,并能成功回放,具体实现步骤如下。

一、软件配置

在华为 HiLens 控制台下载设计器和执行器后,需要对产品进行激活,激活完成后才能够正常运行华为 WeAutomate 工具。

(一) 申请许可

(1) 打开 Studio 设计器后,点击“设置”,再点击设置主页中的“许可”页签,复制本机 ESN,许可证设置模块如图 10-4 所示。

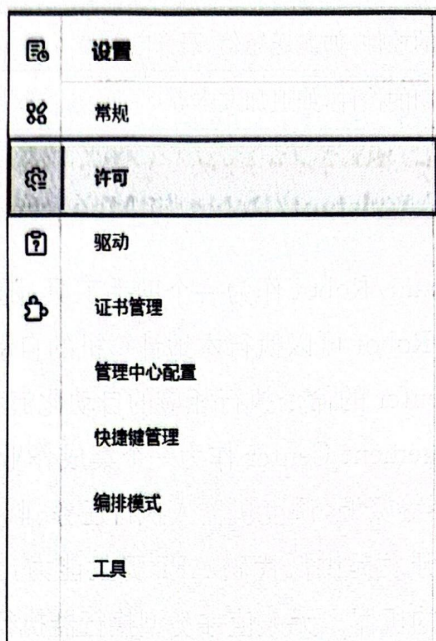


图 10-4 许可证设置模块

(2) 返回华为 RPA 工具下载页面,点击“试用激活”,在弹窗页面粘贴已复制的本机 ESN,点击“获取 License”进行下载,文件格式为 bin 格式。

(3) 返回 Studio 设计器申请页面,点击导入许可文件按钮并选择刚刚下载的 License 文件,页面显示为已导入许可证,则表示 License 已导入成功,具体如图 10-5 所示。

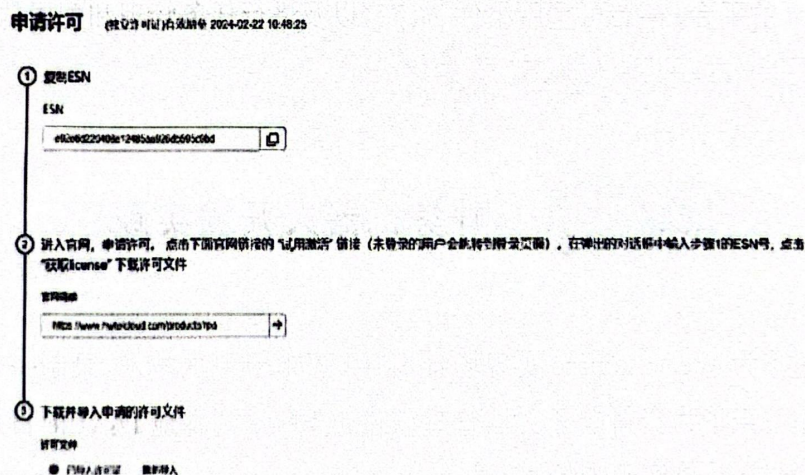


图 10-5 软件配置

(二) Web 驱动配置

考虑到后续的实际运用过程当中需要在浏览器中执行流程,因此需要提前安装相应的浏览器插件,具体方式如下:

(1) 在华为 WeAutomate Studio 中单击“设置”下方的“驱动”模块,可以根据自己的需要选择对应的浏览器插件进行下载。

(2) 下载完成之后,单击“导入驱动”导入刚刚下载的驱动版本即可,具体如图 10-6 所示。

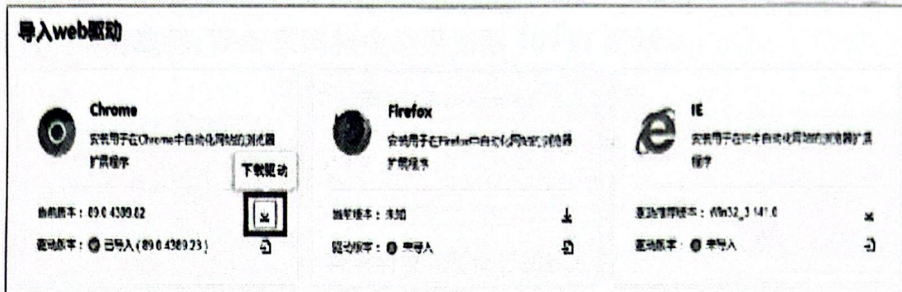


图 10-6 Web 驱动配置

二、开发步骤

完成软件配置之后,即可在项目模块构建自己的流程。本部分的 Web 录制功能是最基础、最简单的流程制作技术。简单来讲,Web 录制是通过录制用户的界面操作,并将录制环节中的各种操作节点转化为相应的流程模块,从而实现 RPA 流程的自动编写。运用 Web 录制的具体操作方式如下。

(一) 创建项目

(1) 通过软件左上角竖向菜单栏的“开始”按钮打开主页,点击“新建项目”开始创建项目。

(2) 输入项目名称,中英文均可,建议使用有意义的项目名称,以方便后续的项目跟踪和使用。

(3) 输入保存路径,此处的路径为工程目录的上级目录。(注意不要手动创建工程目录并配置在保存路径中,这只会导致软件在手动创建的工程目录下再创建一个子目录作为工程目录)

(4) 点击创建可以创建一个工程监理工程,具体开发步骤如图 10-7 所示。

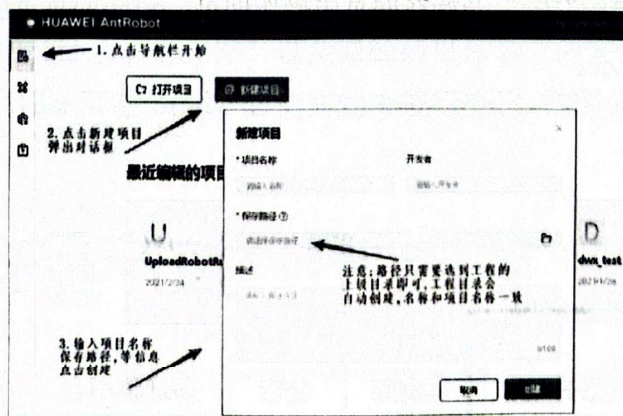


图 10-7 开发步骤

(二) 录制自动化流程

(1) 选择控件“创建文件/目录”，本地创建 D:\file 目录，在控件路径中输入“D:\file”，如图 10-8 所示。注意：接下来下载的路径需要同样更换为 D:\file，否则下载的内容将不在此文件目录之下。

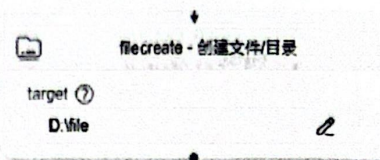


图 10-8 创建文件控件

(2) 首先，选择左上角“UI 录制”，再选择“录制并生成功能块(Ctrl+R)”，如图 10-9 所示，从而进入录制界面。

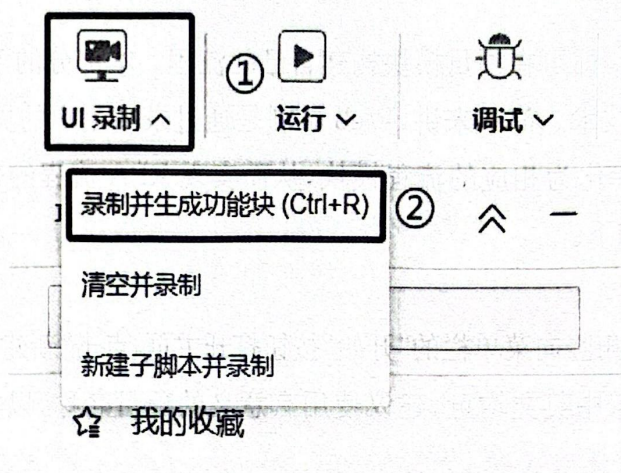


图 10-9 UI 录制模块控件

其次，进入“录制器”后，点击“启动浏览器”按键，输入“http://support.huawei.com/”，按下回车键后，进入华为企业产品技术支持网站。

再次，接着点击“继续”按键，开始在网页中操作即可。跟随网页页面点击“软件下载”按钮，具体如图 10-10 所示。

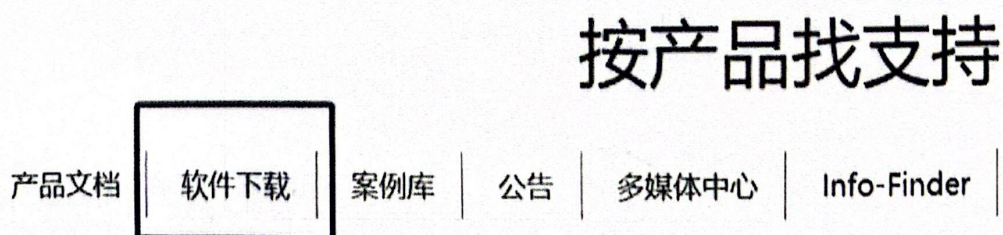


图 10-10 点击“软件下载”以进入下载页面

然后,在产品搜索栏中,搜索“ICSLite”。需要注意的是,此时需要在跳出的输入框中输入而非在原搜索栏输入。输入完成后,点击右侧搜索,即可找到相应的软件应用。在新界面中点击红色字体的“ICSLite”进入下载界面。

最后,在新界面中点击“立即下载”即可。完成以上步骤后,回到“录制器”界面,点击“保存”即可生成相应的流程,Web 录制相应结果如图 10-11 所示。

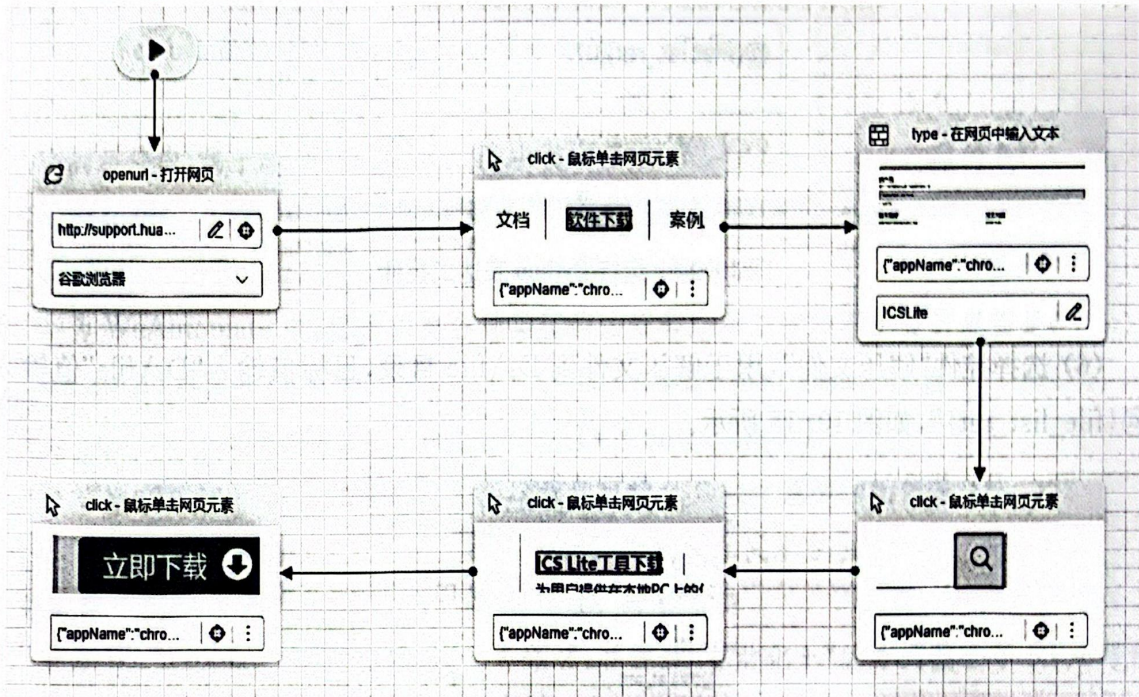


图 10-11 Web 录制相应结果

(3) 选择控件“关闭当前窗口”,关闭当前打开的浏览器窗口,如图 10-12 所示。

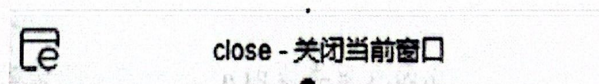


图 10-12 关闭当前窗口控件

(4) 选择控件“列出目录下的文件”,以全路径列出目录下的文件,返回为一个 list,如图 10-13 所示。注意:list 可在 Studio 右侧属性中进行查看,本部分中 list 的具体名称为“fileList_ret”。

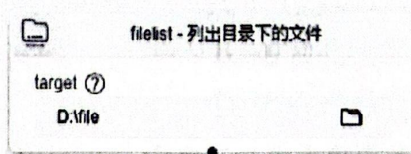


图 10-13 以全路径列出目录下的文件控件

(5) 选择控件“运行 Python 表达式”,Python 中的 eval 表达式,将上一步列出的 list 根据索引取出具体的值,如图 10-14 所示。其中@{fileList_ret}[0]为上一步下载文件的路径,eval_ret 表示为存储变量。

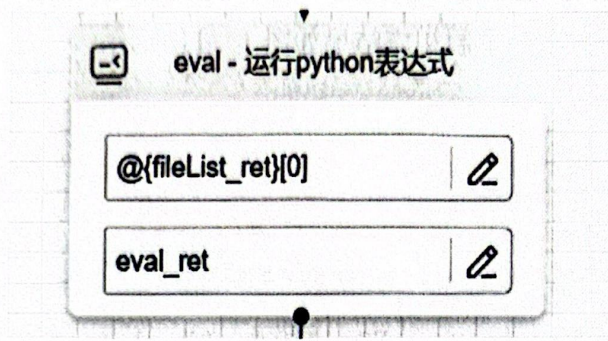


图 10-14 运行 Python 表达式控件

(6) 选择控件“解压文件”,用于解压文件到 D:\file 目录,目标值输入“D:\file”值输入“@{file_list_ret}”,如图 10-15 所示。

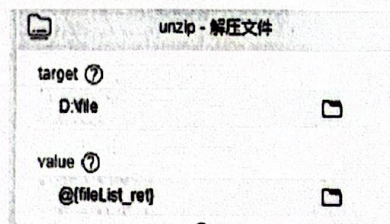


图 10-15 解压文件控件

(7) 完成以上所有步骤之后,将各个控件用“→”连接,即可运行整个流程,从而得到解压后的 ICSLite 文件。

章节测试

一、单选题

1. 华为 WeAutomate 的发展历程大概可分为()阶段。
A. 两个 B. 三个 C. 四个 D. 五个
2. 下列组件中,属于华为 WeAutomate 中负责调度和编排各个自动化流程的是()。
A. Studio B. Management Center
C. Assistant/Robot D. Connector
3. 华为 WeAutomate 中,用于创建各种类型的应用程序和工作流程的设计界面是()。
A. Studio 界面 B. Management Center 界面
C. Assistant/Robot 界面 D. Connector 界面
4. 华为 WeAutomate 中,允许用户无需编程即可设计自动化流程的功能是()。
A. 拖拽式流程设计 B. 低代码开发平台
C. 模型编排 D. 流程录制
5. 华为 WeAutomate 中,聚焦政务、财务领域,并通过“自动化+”的方式整合了 RPA 与低代码、AI、大数据开发平台的版本是()。
A. 初创期版本 B. 发展期版本
C. 探索期版本 D. 成熟期版本

二、多选题

1. 下列选项中,属于华为 WeAutomate 的主要用户的有()。
A. 流程开发人员 B. 业务管理员
C. 系统管理员 D. IT 支持人员
2. 下列选项中,属于华为 WeAutomate 的结构部分的有()。
A. Studio B. Management Center
C. Assistant/Robot D. Connector
3. 下列选项中,属于华为 WeAutomate 的功能的有()。
A. 流程设计 B. 自动化执行
C. 智能决策 D. 低代码开发
4. 下列选项中,属于华为 WeAutomate 在实际工作中取得的成效的有()。
A. 效率提升 B. 收益增加
C. 员工满意度提高 D. 产品交付速度提升

5. 下列选项中,属于华为 WeAutomate 的生态策略的有()。
- A. 伙伴认证 B. 伙伴赋能 C. 联合拓展 D. 伙伴激励

三、判断题

1. 华为 WeAutomate 的设计器提供图形化机器人脚本编排能力,支持录制和回放。 ()
2. 华为 WeAutomate 的 Robot 执行器负责执行 Studio 设计好的自动化流程。 ()
3. 华为 WeAutomate 的 Management Center 仅用于调度各个自动化流程,不涉及流程的编排。 ()
4. 华为 WeAutomate 在成熟期阶段才开始聚焦政务、财务领域。 ()
5. 华为 WeAutomate 的生态策略强调“被集成”,通过赋能生态,由合作伙伴面向海量客户提供服务。 ()

四、思考题

某国有银行为了提高财务管理的效率和准确性,决定开发一款财务机器人。在开发过程中,银行注重将社会主义核心价值观融入项目中,培养员工的责任感和创新能力。

项目团队积极采用最先进的 RPA 技术,不断优化财务机器人的功能,体现了创新驱动发展战略的实施。开发过程中,IT 部门、财务部门和业务部门紧密合作,共同解决开发中遇到的问题,展现了集体主义的力量。银行对财务机器人的性能和安全性进行了严格的测试和审查,确保产品的质量,体现了对客户和社会的高度负责。财务机器人的投入使用,极大提高了银行的财务管理效率,为客户提供了更高质量的服务,体现了服务社会的宗旨。请读者试着思考以下几个问题:

1. 请列举几种常见的 RPA 开发工具,并简要介绍它们的特点和优势。
2. 在进行 RPA 开发时,如何选择合适的开发工具? 请给出你的建议。
3. 请解释 RPA 开发中的“流程设计”和“自动化规则”是什么? 它们在 RPA 开发中起什么作用?
4. 在 RPA 开发过程中,如何确保开发出的机器人能够稳定、高效地运行? 请列举一些方法和技巧。
5. 在实际应用中,RPA 可能会遇到哪些挑战和问题? 请给出你的看法。
6. 请谈谈你在 RPA 开发过程中遇到的一个问题,以及你是如何解决这个问题的。
7. 如何评估一个 RPA 项目的成功与否? 请给出你的观点。



财经类专业“十四五”规划新形态教材

书名	第一主编
财务管理实务	王炳华
智慧化税费申报与管理	贾瑞敏
智能化成本核算与管理	王爱玲
企业内部控制	张长胜
智能财务共享服务	高晓华
业财一体信息系统应用	张建峰
智能财税	闫 菲
Python数据分析与可视化	张洪忠
RPA财务机器人应用与开发——基于UiPath	王 浩
● RPA财务机器人应用	余冰冰
Python企业财务应用	李靠队

ISBN 978-7-5429-7505-8



9 787542 975058 >

定价：49.80 元

ACCA
ASSOCIATION OF
COST ACCOUNTANTS

中国注册会计师
资格考试丛书

2020年版
中国注册会计师
CMA 认证中文教材

战略财务管理

● PART II

中国注册会计师

CMA

战略财务管理

PART II · STRATEGIC FINANCIAL MANAGEMENT

ISBN 978-7-119-10442-9

中国商业出版社


 中华会计网校
www.chinaacc.com
正保远程教育旗下品牌网站
美国纽交所上市公司(代码:DL)

梦想成真
系列辅导丛书

2020年版美国注册管理会计师(CMA)认证考试教材

战略财务管理

白默 中华会计网校 编

 中国商业出版社

图书在版编目(CIP)数据

战略财务管理 / 白默, 中华会计网校编. —北京:
中国商业出版社, 2019. 12 (2020. 9 重印)

ISBN 978-7-5208-1024-1

I. ①战… II. ①白…②中… III. ①财务管理-高
等学校-教材 IV. ①F275

中国版本图书馆 CIP 数据核字 (2019) 第 266978 号

责任编辑: 朱丽丽

中国商业出版社出版发行
010-63180647 www.c-cbook.com
(100053 北京广安门内报国寺1号)

新华书店经销
北京市群英印刷有限公司印刷

*

787 毫米×1092 毫米 16 开 24 印张 661 千字
2020 年 1 月第 1 版 2020 年 9 月第 2 次印刷

定价: 98.00 元

* * * *

(如有印装质量问题可更换)

前 言

2011年一次偶然的机会有幸受邀在北京讲授英文CMA课程,开始与CMA考试结缘。彼时,我刚刚在南开大学会计系拿到博士学位,自恃通读过财务和管理学的英文文献,然而现实是CMA的备课和讲授过程让我倍感压力。因为它的课程体系不仅有理论深度,还非常重视贴近实务。题目中有很多情景或案例的分析,考查多个知识点的衔接和具体应用。课堂上,我也曾有过被学员追问知识点如何落地的尴尬。所谓知耻后勇,九年期间,我有幸与国家电网、北京联想股份有限公司、北汽集团等公司合作交流,并利用自己受聘天津市科技特派员的机会,深入一些制造业、服务业公司,认真了解、学习管理会计应用的现状和趋势,并开始参与一些企业具体的管理会计项目的实施。经过不断充实积累,我已经在北京、上海、南京、天津、珠海、大连等十二个城市讲授过CMA课程,教授学员超过3万人,取得了良好的教学效果,我也逐渐可以与实务管理经验丰富的学员深入交流管理会计的落地应用。我非常珍视和感激这段教学相长的经历。

2018年年末,正保教育集团邀请我撰写CMA教材。此时,我也正有把讲授CMA课程的收获、经验系统总结出来的想法。而且在与正保教育长期合作的过程中,我也感受到公司员工专业、敬业、一丝不苟的工作态度,所以我们一拍即合开始工作。期间经历了2020年开始实施的新考试大纲,对相应知识点做了仔细修改,终于成稿。

这套CMA教材的特点,或者说“差异化”之处,主要体现在以下三点:

第一,作为一本应试考试的教材,笔者的第一要务当然是帮助考生顺利通过考试。依据2020年开始实施的新考试大纲,我对所有考点进行了全面细致地梳理、对重点难点进行深入分析和总结、对经典例题加以解析、精选了章节习题帮助考生熟悉考试思路 and 模式,相信考生读后会有所受益。

第二,我将不同章节、包括P1(财务规划、绩效与分析)和P2(战略财务管理)两本教材中相关联的知识点加以提示和标注,目的不仅在于帮助读者在复习时注意知识点的前后衔接,还希望读者能够以战略和全局的角度来掌握管理会计的知识体系,真正将其学以致用。

第三,我将自己在实务领域学习和掌握的管理会计知识与考纲中的知识点相融合,希望帮助读者认识到:管理会计的知识点不是冰冷枯燥的计算公式和模型,而是能够帮助企业实现价值增值的工具。

2019年,随着中国经济增速步入“新常态”,原来一些企业的野蛮扩张模式必然遭遇更大的挑战,而向管理要效益、重视决策的科学性、使用恰当的分析工具和方法进行价值判断和预测,会被越来越多的企业所重视。因此,管理会计会有更广阔的发展空间。

在讲授CMA课程近十年之际,我把授课过程的所学、所思、所想总结出来奉于读者。“妆罢堂前问夫婿,画眉深浅入时无”,读者是最好的评委,请“孜孜焉唯进修是急”的各位来评判!

白默

目 录

第一章 财务报表分析	1
第一节 基本财务报表分析	1
第二节 财务比率和盈利能力分析	7
第三节 特殊问题	28
课后习题	33
第二章 公司财务	53
第一节 风险和报酬	55
第二节 长期财务管理	67
第三节 筹集资本	89
第四节 营运资本管理	100
第五节 公司重组	121
第六节 国际金融	127
课后习题	137
第三章 决策分析	174
第一节 本量利分析法	175
第二节 边际分析	184
第三节 定价	199
课后习题	220
第四章 风险管理	250
第一节 风险管理	251
第二节 企业风险管理	260
课后习题	269

第五章 投资决策	272
第一节 资本预算过程	273
第二节 贴现现金流分析	286
第三节 投资回收期和折现回收期	297
第四节 资本投资中的风险分析	300
课后习题	305
第六章 职业道德	332
第一节 商业道德	333
第二节 管理会计和财务管理专业人士的职业道德规范	335
第三节 组织对道德的考虑	341
课后习题	350
附录 1 考试大纲	360
附录 2 系数表	371
参考文献	378



中华会计网校 CMA 学习交流群

- 实现规模经济
- 获取资源,如通过整合营销团队、设施和分销渠道获取市场资源。
- 获取客户
- 实现比内部增长更快的增长机会
- 提供多元化的产品和服务
- 剥离非核心资产
- 利用标的公司结转以后年度弥补的经营亏损来抵税
- 获取必要资金
- 稳定股价
- 迫于监管层的压力

(二)公司重组的定量分析

评估并购标的价值的最重要的方法是现金流量折现法(简称 DFC, Discounted Cash Flow Method)。

该方法通过计算自由现金流量,将其按照必要报酬率进行折现,计算出标的物资产的价值。

自由现金流量是指扣除税收、必要的资本性支出和营运资本增加后,能够支付给所有的求偿者(包括债权人和股东)的现金流量。

基本公式为:

企业自由现金流量 = 息税前利润(EBIT)×(1-所得税税率) + 折旧和摊销 - 资本性支出 - 营运资本净增加额

例:A公司2018年销售收入\$7,000,000,实现息税前利润\$1,500,000。公司的营运资本占销售收入的比重为5%,2019年销售收入和息税前利润增长率均为10%。资本性支出与折旧分别为\$700,000和\$550,000。所得税税率为40%。则计算A公司2019年的自由现金流量。

2019年的销售收入 = $\$7,000,000 \times (1+10\%) = \$7,700,000$

2019年新增的营运资本按照销售收入百分比法计算,因为营运资本占销售收入的5%,而2019年新增的销售收入为\$700,000,所以2019年新增的营运资本 = $\$700,000 \times 5\% = \$35,000$ 。

2019年自由现金流量 = $\$1,500,000 \times (1+10\%) \times (1-40\%) + \$550,000 - \$700,000 - \$35,000 = \$805,000$

如果Z公司在2020年年初考虑收购A公司,预计A公司每年\$805,000的自由现金流量是稳定的,可以持续15年。Z公司的必要报酬率为14%,在年金现值系数表中,15期折现率14%的年金现值系数为6.142。所以Z公司并购A公司现金收益的现值 = $\$805,000 \times 6.142 = \$4,944,310$ 。

因此,Z公司报出的标的公司A公司的收购价格不应超过\$4,944,310。

第六节 国际金融

一、汇率的基本概念和影响汇率变化的因素

(一)与汇率相关的基本概念

1. 汇率

把一个国家的货币用另一个国家的货币价值来表示的指标就是汇率。具体来讲,汇率表示两种货币之间兑换的比率,或者说是一个国家的货币对于另一个国家货币的价值。在实际经济

活动中,美元是全球几乎所有货币交易中通用的货币。比如,人民币和英镑都是按照美元报价的价格来进行交易的。由于汇率一直处于变动之中,所以企业的管理者也会时刻关注汇率给企业带来的一系列影响。

2. 直接标价法和间接标价法

通常情况下,要想确定两种不同货币之间的比价,先要确定使用哪个国家的货币作为标准。由于确定的标准不一样,于是就产生了两种不同的外汇汇率标价方法。

直接标价法是指一单位的外币可以兑换多少本国货币的标价方式,相当于计算购买一定单位外币所应付的本币。在国际外汇市场上,包括中国在内的世界上绝大多数国家目前都采用直接标价法。例如,对于日本央行而言,日元兑美元汇率为 $\$1:\text{¥}100$,即1美元可以兑换100日元,该方法即为直接标价法。

间接标价法是指一单位的本国货币可以兑换多少外币的标价方式,相当于计算购买一定单位本币所应付的外币。例如,对日本央行而言,日元兑美元的汇率为 $\text{¥}1:\$0.01$,即1日元可以兑换0.01美元,这种标价方式即为间接标价法。

直接标价与间接标价互为倒数关系。一般来说,一个国家在一定时期内,其外币的折算方法只能采用其中一种。世界上大部分国家都采用直接标价法,只有少数发达国家采用间接标价法,如英国、美国。

3. 贬值与升值

(1) 货币贬值

货币贬值(Depreciation)是指一国货币币值的降低。若A国货币贬值,则单位A国货币将购买更少的B国货币及B国的商品。

例如,当前日本的尼康(Nikon)数码相机用日元的标价为 $\text{¥}100,000$,而美国的iPhone手机用美元的标价为 $\$1,000$ 。对日本而言,原汇率为 $\text{¥}1:\$0.01$ 。若日元贬值,贬值后的汇率为 $\text{¥}1:\$0.008$,即一日元仅能兑换0.008美元。不考虑关税等其他因素,则原来日本的尼康相机按照 $\text{¥}1:\$0.01$ 的汇率,在美国市场的标价为 $\$1,000$ 。而按照贬值后的汇率 $\text{¥}1:\$0.008$, $\text{¥}100,000$ 的尼康相机在美国市场的售价变为 $\$800$ 。由于日元贬值导致日本商品售价的下降,意味着美国会增加对日本商品的需求量,即日本向美国出口商品的增加。

而对美国而言,在 $\text{¥}1:\$0.01$ 的汇率水平下, $\$1,000$ 的iPhone手机,在日本的售价为 $\text{¥}100,000$,但是当日元贬值,汇率变为 $\text{¥}1:\$0.008$ 后, $\$1,000$ 的iPhone手机,在日本的售价变为 $\text{¥}125,000$ 。日元贬值导致美国商品价格上涨,需求下降,即日本从美国进口商品的减少。

所以**本币贬值,意味着对本国出口商品的补贴,和对进口商品的征税。**

(2) 货币升值

货币升值(Appreciation)是指一国货币币值的上升。若A国货币升值,则单位A国货币将购买更多的B国货币及B国的商品。

本币升值,意味着对本国出口商品的征税,和对进口商品的补贴。推导过程为上例相反的方向,不再赘述。

4. 即期汇率与远期汇率

(1) 基本概念

有一句话“你永远不知道,明天和意外哪个会先到来”,汇率也是一样,我们永远不会知道明天汇率的涨与跌,而这种“不知道”即为一种不确定性,这种不确定性意味着汇率风险的存在。

即期汇率是指某货币目前在现货市场上进行交易的价格。即交易双方达成外汇买卖协议

后,在两个工作日以内办理交割的汇率。远期汇率指的是买卖双方事先约定的,据以在未来的特定日期进行外汇交割的汇率。

(2) 升水与贴水

远期升水(Forward Premium)是指远期汇率高于即期汇率的百分比。**远期贴水**(Forward Discount)是指远期汇率低于即期汇率的百分比。升水和贴水一般均以年为报价基础。

$$\text{升水或贴水的幅度} = \left(\frac{F_i}{S_i} - 1 \right) \times \frac{360}{\text{远期的天数}} \quad (2.6.1)$$

式中, F_i = 间接标价法的远期汇率; S_i = 间接标价法的即期汇率。

如果题目中涉及的已知条件是直接汇率,则通过求倒数即可算出间接汇率。

如公式(2.6.1)所示,升水(贴水)幅度的计算均默认以年为单位。

【经典例题-19】 A国货币的当前汇率为0.41,90天后汇率为0.40,则A国货币90天后将()。

- A. 升水 9.76% B. 贴水 9.76% C. 升水 2.44% D. 贴水 2.44%

【答案】 B

【解析】 本题有两个考点:一是远期的期限为90天,不足一年,而升水和贴水的计算默认为年效应,所以要进行调整;二是题目并未说明是直接标价法还是间接标价法。

首先验证直接标价法,即当前汇率为1外币=0.41 A国货币。

90天后汇率为1外币=0.40 A国货币。

$$\text{升水幅度} = \left(\frac{F_i}{S_i} - 1 \right) \times \frac{360}{\text{远期的天数}} = \left(\frac{1}{\frac{0.40}{0.41}} - 1 \right) \times \frac{360}{90} = 10\%, \text{显然没有符合该条件的选项。}$$

再验证间接标价法,即当前汇率为1个A国货币=0.41 外币。

90天后汇率为1个A国货币=0.40 外币

$$\text{贴水幅度} = \left(\frac{F_i}{S_i} - 1 \right) \times \frac{360}{\text{远期的天数}} = \left(\frac{0.40}{0.41} - 1 \right) \times \frac{360}{90} = 9.76\%$$

所以答案为B。

(3) 抛补利率套利

抛补利率套利(Covered Interest Arbitrage)是指,对两国利息差进行套利的同时还能规避汇率风险的过程。

抛补利率套利包括两部分内容:一是利率套利,指对两国之间的利率差异进行套利;二是抛补,是指规避汇率风险。

例如,你希望用\$800,000在外汇市场进行投资,期限一年。美国的市场利率为2%,英国的市场利率为4%。英镑的即期汇率为\$1.60,如果你希望在英国获得相对较高的利息收益,由于两个国家的利率是确定的,只是360天后英镑兑换美元的市场汇率不确定。假定英镑360天后的汇率不变,仍为\$1.60。

表 2-6-1 利率套利过程

	美元	汇率	英镑
投资金额	\$ 800,000	\$ 1.60	£ 500,000

(续表)

	美元	汇率	英镑
市场利率			4%
利息收入			£ 20,000
收到的本息和	\$ 832,000	\$ 1.60	£ 520,000
实际回报额	\$ 32,000		
实际回报率	$\$ 32,000 / \$ 800,000 = 4\%$		

由表 2-6-1 可知,英镑的即期汇率与到期后的市场汇率相等。但是当远期汇率发生变化时,如果只考虑利率套利,不进行抛补,收益就可能面临风险。

在上例中,其他条件不变,即美国的利率为 2%,英国的利率为 4%。但是英镑的即期汇率改为 \$ 1.62,如果不考虑汇率风险,不进行抛补,若英镑 360 天后的汇率为 \$ 1.5888。则:

表 2-6-2 利率变化但未进行抛补的投资过程

	美元	汇率	英镑
投资金额	\$ 800,000	\$ 1.62	£ 493,827
市场利率			4%
利息收入			£ 19,753
收到的本息和	\$ 815,976	\$ 1.5888	£ 513,580
实际回报额	\$ 15,976		
实际回报率	$\$ 15,976 / \$ 800,000 = 2\%$		

从表 2-6-2 可见,如果只考虑利率套利,不进行抛补,则汇率的波动造成收益的下降。此时,在英国投资并不能获得更高的收益。

为了避免这种情况的发生,投资者可以出售英镑的远期合约,以此来确保在未来某一时刻将英镑兑换成美元的汇率。在上例中,仍假设其他条件不变,美国的利率为 2%,英国的利率为 4%,英镑的即期汇率为 \$ 1.62。但是英镑 360 天后远期合约的汇率为 £ 1 = \$ 1.60。投资者可以出售英镑的远期合约(持有空头头寸)。此时,一个完整的抛补利率套利的过程如表 2-6-3 所示。

表 2-6-3 抛补利率套利的投资过程

	美元	汇率	英镑
投资金额	\$ 800,000	\$ 1.62	£ 493,827
市场利率			4%
利息收入			£ 19,753
收到的本息和	\$ 821,728	\$ 1.60	£ 513,580
实际回报额	\$ 21,728		
实际回报率	$\$ 21,728 / \$ 800,000 = 2.72\%$		

可见,一个抛补利率套利的抛补过程,是将投资者的英镑投资可以兑换的美元金额予以固化,从而降低了汇率风险,保证了投资者取得了两国利率差额的收益。

投资者从抛补利率套利中所获得的回报可以通过以下的过程确定：

$$R = (1 + I_f) \times \frac{F_d}{S_d} - 1 = (1 + I_f) \times \frac{S_i}{F_i} - 1 \quad (2.6.2)$$

式中, R = 投资回报率; I_f = 外国的存款利率; I_h = 本国的存款利率; F_d = 直接标价法的远期汇率; F_i = 间接标价法的远期汇率; S_d = 直接标价法的即期汇率; S_i = 间接标价法的即期汇率。

【经典例题-20】 一家美国企业在加拿大投资 \$ 100,000, 一年期的投资回报率为 10%。在该投资开始时, 加元的即期汇率为 \$ 0.65; 在投资到期时, 加元的汇率为 \$ 0.70。那么该投资的实际收益率为()。

- A. 2.14% B. 7.69% C. 10.00% D. 18.46%

【答案】 D

【解析】 $R = (1 + I_f) \times \frac{F_d}{S_d} - 1 = (1 + 10\%) \times \frac{\$ 0.70}{\$ 0.65} - 1 = 18.46\%$

(二) 影响汇率波动的因素

1. 相对的通货膨胀率

通货膨胀使一国的货币在国内购买力下降, 同时也会导致该国货币相对贬值。因为汇率是两国币值的对比, 假定 A 国通货膨胀率高、发行货币过多, 则以 A 国货币计价的商品价格将上涨, 涨价意味着 B 国进口 A 国商品时, 就要支付更多 B 国货币才能购得等量的 A 国商品。这导致 B 国对 A 国商品的需求减少, 对 A 国货币的需求也随之减少。如果 B 国未发生通胀, 商品价格也未变化, 则 B 国商品相对于通货膨胀中的 A 国商品更便宜, A 国对 B 国商品的需求会增加, 即为采购 B 国商品而出售 A 国货币、买入 B 国货币、进而对 B 国货币的需求增加。所以, A 国货币供给增加(出售数量增加)、需求下降, 将导致 A 国货币贬值。

因此, 当一国通胀率增幅小于另一国时, 前者的币值相对升值, 后者的币值相对贬值。通胀率与币值存在负相关关系。

2. 相对的收入水平

如果 A 国收入水平较 B 国增加, 则 A 国货币相对贬值, B 国货币相对升值。以美国和英国为例。如果美国民众收入水平增加, 而英国民众收入水平不变。则美国人收入水平提升后, 原来收入较低时, 有意愿但没能力进行消费的英国商品, 如 Minicopper 的轿车, 随着美国人收入的增加, 他们成为有意愿且有能力购买的真实需求者。美国会增加对英国商品的进口, 在外汇市场中表现为出售美元购入英镑。而由于英国人的收入水平不变, 则对美国商品的需求不会发生变化, 即愿意出售英镑购入美元, 继而采购美国商品的数额不会发生改变。所以, 美元的供给增加而需求不变, 导致价格下跌, 美元相对于英镑发生贬值。当然, 英镑相对于美元发生升值。

因此, 当一国收入水平相对于另一国上升时, 则该国货币发生贬值, 即收入水平与币值存在负相关关系。

3. 相对利率水平

当 A 国的利率相对于 B 国利率提升时, A 国货币趋于升值, B 国货币趋于贬值。利率水平的提升意味着资金的收益率更高。例如, 原来 A、B 两国的利率相等, 均为 1.5%, 如果 A 国央行加息, 将利率提高到 1.75%, 而 B 国的存款利率仍为 1.5%, 则 B 国的资金所有者会出售 B 国货币兑换 A 国货币, 并将兑换后的 A 国货币存至 A 国金融机构, 以期获得更高的收益。这一过程导致 B 国货币的供给增加, 而 B 国货币的需求量没有变化, B 国货币价格下跌, 相对于 A 国货币贬

值。当然,A国货币也相对于B国货币发生升值。

因此,当一国利率增幅小于另一国时,前者的币值相对贬值,后者的币值相对升值,即相对利率与币值存在正相关关系。

需要注意的是,这里的利率指的是实际利率水平,而非名义利率。实际利率和名义利率的关系,可以用费雪公式来表示:

$$I_n = (1+I_r) \times (1+I_{INF}) - 1 \quad (2.6.3)$$

式中, I_n = 名义利率; I_r = 实际利率; I_{INF} = 通货膨胀率。

因为公式(2.6.3)中, I_{INF} 两者相乘的数额很小,可以忽略不计,所以可得:

$$I_n = I_r + I_{INF} \quad (2.6.4)$$

在CMA考试中,我们通常直接使用公式(2.6.4)。

例如,俄罗斯和委内瑞拉的实际利率均为2%,俄罗斯卢布的名义利率为18%,委内瑞拉货币的名义利率为15%,则卢布的币值相对委内瑞拉货币将发生升值还是贬值?

根据公式(2.6.4),已知名义利率和实际利率,可计算出卢布的通货膨胀率为16%,委内瑞拉货币的通货膨胀率为13%,前者高于后者,因此卢布将贬值。

再举一例,俄罗斯和委内瑞拉的通货膨胀率均为15%,俄罗斯卢布的名义利率为20%,委内瑞拉货币的名义利率为18%,则卢布相对委内瑞拉货币将发生升值还是贬值?

根据公式(2.6.4),已知名义利率和通货膨胀率,可计算出卢布的实际利率为5%,委内瑞拉货币的实际利率为3%,前者高于后者,所以此例中,卢布将升值。

4. 影响汇率变化的其他因素

CMA考试中,最重要的引发汇率变化的是上述三个因素,但是还有一些因素也会引发汇率的波动。

(1) 政府管制(对进口限制)

如果政府采取更严格的贸易保护主义政策,对进口商品实施更严格的限制,则本币兑换外币的需求减少。外币供给不变,需求减少,将导致外币相对于本币贬值。所以,币值与贸易保护主义的程度呈正相关关系。

(2) 市场预期

如果市场普遍预期一国货币会升值,则对该国货币的需求增加,将导致该国货币升值。反之,如果普遍预期一国货币会贬值,则该国货币会遭到大量的抛售,导致币值贬值。

(3) 资本流动方向

如果A国吸引投资的能力增强,有大量资本流入A国,则资本流入的过程需要其他国家出售本国货币来购买A国货币,A国货币的需求增加,币值相对升值。反之,若A国大量资本流出,资本流出的过程,A国出售本国货币兑换其他国家货币,则A国货币将发生贬值。

(4) 货币政策

货币政策有宽松(Expanding)和紧缩(Contracting)两种。宽松的货币政策,即增加流通中的货币投放数量,货币供给量增加,货币价格(利率)将下跌,其经济后果是实际利率下降,所以将导致汇率的降低。

反之,如果采用紧缩的货币政策,如美联储在渡过2008年金融危机后,开始逐步放弃量化宽松政策,降低流通中美元的数量,随之而来的是新一轮加息周期。实际利率的上升,将导致美元汇率的上升。

表2-6-4为各种影响汇率变化的因素的汇总。

表 2-6-4 汇率变动因素汇总表

影响因素		本币	外币	备注
本国通胀率相对于外国	↑	↓	↑	币值与相对通胀率成反比
	↓	↑	↓	
本国收入水平相对于外国	↑	↓	↑	币值与相对收入水平成反比
	↓	↑	↓	
本国实际利率相对于外国	↑	↑	↓	币值与相对实际利率成正比
	↓	↓	↑	
政府管制(对进口限制)	↑	↑	↓	币值与贸易保护主义的程度成正比
	↓	↓	↑	
市场预期	↑	↑	↓	币值与市场预期成正比
	↓	↓	↑	
资本流动	流入本国	↑	↓	流入增加需求 流出资本逃离
	流出本国	↓	↑	
货币政策	宽松	↓	↑	与实际利率相同,即宽松政策, 利率下降
	紧缩	↑	↓	

应该注意的是,考试中涉及影响汇率变化的原因,一般都是敏感性分析。即如表中所示,假定其他条件不变,单一因素(如收入水平)发生变化,将引发该国汇率发生怎样的变动。当然在现实中,一国汇率的变化都是由多个因素共同作用的结果。

【经典例题-21】若美国人青睐于瑞士腕表,而英国相对于美国通胀率高,则()。

- A.美元相对于瑞郎和英镑升值 B.美元相对于瑞郎升值,相对于英镑贬值
C.美元相对于瑞郎贬值,相对于英镑升值 D.美元相对于瑞郎和英镑贬值

【答案】 C

【解析】为进口更多瑞士腕表,需出售美元购入瑞士法郎,导致瑞士法郎的需求增加,币值上涨,则美元相对于瑞士法郎贬值。美国通胀低于英国,则美元相对英镑升值。

二、汇率风险管理

我们在上一部分刚刚介绍了抛补利率套利的过程,它实际上是对汇率风险的应对措施。除了抛补利率套利,我们还将介绍一系列应对汇率风险的其他措施。

(一) 早付迟收和早收迟付法

早付迟收和早收迟付法(Leads & Lags)是通过调整支付时间来防范外汇风险的一种方法。如果预计外币汇率上升,进口商和出口商分别采用早付和迟收的方法;如果预计外币汇率下跌,则进口商和出口商应分别采用迟付和早收法。当国际交易当事人对货币汇率变化趋势的预测有较大的把握时,可以通过提前或延迟外汇资金的收付时间来防范外汇风险。

具体而言,对赊购商品的进口商来说,如果预期支付外币的汇率将要上升,该进口商可以提前支付货款,因为如果延迟支付,随着外币汇率的不断上升,需要用更多的本币才能兑换等额的外币,满足支付的需求。所以此时,越早支付越有利;如果预期支付外币的汇率将要下跌,则该进口商会尽可能延迟支付货款,等到汇率下跌后再支付,从而降低其本币购汇成本。

对于赊销商品的出口商的收款策略则正好相反,即预期应收的外币的汇率将要上升,他可以延迟收取货款,因为外币汇率不断上升,收到外币货款后可以兑换为更多的本币。所以此时,延迟收款对出口商有利;如果他预期应收外币的汇率将要下跌,他会尽可能加速收款,以防外币汇率下跌兑换本币的金额变少。

在题目中我们还需要注意以下两个问题。

第一,我们关注的是外汇风险的管理。如果在题干中,出现一家美国公司向德国赊销商品,形成对德国公司的€250,000的应收账款,则要根据对欧元汇率升值(贬值)的预判,做出延迟(加速)收款的决策。但是如果题干中,销售方美国公司形成了对德国公司\$250,000的应收账款,则美国公司收取的是美元,对公司而言不存在外汇风险管理的问题。

第二,如果题干中,涉及一家美国的A公司向德国的B公司赊销商品,形成对德国B公司的€250,000的应收账款,到期日为90天;A公司从法国C公司赊购一批原材料,形成对C公司€250,000的应付账款,到期日为90天。则A公司不存在任何外汇风险。因为应收和应付款项到期日、币种和金额完全相同,两者实现了外汇风险的对冲。即如果A公司在一笔外汇交易发生后,有意地再进行一笔与该笔交易在到期日、币种和金额上完全相同,但资金流向正好相反的交易,从而将汇率变动的风险抵消,这也是一种外汇风险管理方式。如果本例中,A公司的应收账款为€250,000,但是应付账款为€220,000,则净外币敞口(Net Foreign Currency Exposure)为应收账款€30,000。A公司只需要对€30,000应收账款的净外币敞口进行外汇风险的管理。

(二) 合约性套期保值

套期保值的方法我们在本章第二节已经介绍过。使用套期保值法管理汇率风险,是指在金融市场上进行操作,使外汇净流量为零,把外汇风险转化出去。具体方法包括:在远期合约市场、外汇期货市场和外汇期权市场上进行套期保值。

1. 赊销方的套期保值

例如,某日本公司(本币为日元)向美国出口货物,美国公司6个月后支付100万美元的货款。

赊销方日本公司可以采取以下几种防范外汇交易风险的方法:

(1) 远期合约市场套期保值

卖出美元远期合约。在远期合约中,卖出远期合约100万美元,即持有空头头寸。对于6个月后收取100万美元外币的日本公司而言,主要担心6个月后美元贬值,导致收到的100万美元能够兑换的日元数额变少。而通过远期合约出售美元,锁定了美元的出售价格,可以规避美元贬值给公司带来的外汇风险。

(2) 期货市场套期保值

卖出美元期货。道理同上,通过卖出100万美元的期货,锁定了美元的出售价格,可以规避美元贬值给公司带来的外汇风险。

(3) 期权市场套期保值

买入美元看跌期权。看跌期权赋予持有人以行权价格(固定的)出售特定数量标的物的权利。买入行权日为6个月后、100万美元的看跌期权,日本公司拥有了以固定价格卖出100万美元的、兑换为固定金额日元的权利,从而规避美元贬值给公司带来的外汇风险。

2. 赊购方的套期保值

例如,某德国公司(本币为欧元)从美国进口货物,德国公司6个月后支付100万美元的货款。

赊购方德国公司可以采取以下几种防范外汇交易风险的方法:

(1) 远期合约套期保值

买入美元的远期合约。在远期合约中,购入远期 100 万美元,即持有空头头寸。对于 6 个月后支付 100 万美元外币的德国公司而言,主要担心 6 个月后美元升值,导致公司需要用更多的欧元才能兑换 100 万美元用于支付。而通过远期合约购入 100 万美元,锁定了美元的购入价格,可以规避美元升值给公司带来的外汇风险。

(2) 期货市场套期保值

买入美元期货。道理同上,通过买入 100 万美元的期货,锁定了美元的购入价格,可以规避美元升值给公司带来的外汇风险。

(3) 期权市场套期保值

买入美元看涨期权。看涨期权赋予持有人以行权价格(固定的)购入特定数量标的物的权利。买入行权日为 6 个月后、100 万美元的看涨期权,德国公司拥有了以固定价格购入 100 万美元的权利,即锁定了购汇成本,从而规避美元升值给公司带来的外汇风险。

三、对外投资

(一) 对外投资

对外投资是指投资者将其拥有的资本投入海外,从而进行直接或间接经营的投资形式。对外投资不仅是在经济发达的国家之间进行,而且在经济发达国家和发展中国家也广泛地进行着。经济发达国家的公司,常利用对外跨国投资形式,在发展中国家设立分支机构或子公司,从事跨国经营活动,可以通过多元化投资降低风险,规避关税壁垒。我们在本章第一节中已经介绍,通过构成投资组合降低风险的条件是:本土公司与海外分部的业绩不能呈现完全正相关关系,即业绩的相关系数不为 1。只要不是完全正相关,则通过对外投资即可降低公司风险。

(二) 国际贸易的融资手段

根据《巴塞尔协议》,国际贸易融资是围绕国际贸易结算的各个环节所发生的资金和信用的融通活动。具体来讲,国际贸易融资是银行对进口商或出口商提供的与进出口结算业务相关的资金融通或信用便利。

1. 国际保理

国际保理业务在欧美发达国家是一种成熟的金融结算工具。保理是保付代理的简称,是卖方与保理商之间的一种契约关系,主要是为赊销方式而设计的一种综合性金融服务。它指的是接受出口商的委托,向出口商提供进口商的信用额度调查、信用证风险担保、应收账款管理和贸易融资的综合性服务项目。其是集融资、结算、财务管理和信用证担保于一身的融资结算方式。

2. 信用证

信用证是银行有条件保证付款的证书,是国际贸易活动中常见的结算方式。按照这种结算方式的一般规定,买方先将货款交存银行,由银行开立信用证,通知异地卖方开户银行转告卖方,卖方按合同和信用证规定的条款发货,买方银行代买方付款。

信用证方式有三个特点:一是信用证不依附于买卖合同,银行在审单时强调的是信用证与基础贸易相分离的书面形式上的认证;二是信用证是凭单付款,不以货物为准,只要单据相符,开证行就应无条件付款;三是信用证是一种银行信用,它是银行的一种担保文件。

3. 银行承兑汇票

银行承兑汇票是商业汇票的一种,企业作为承兑人的汇票称为商业承兑汇票,银行作为承兑人的汇票称为银行承兑汇票。承兑即承诺兑付,即承担这张商业汇票项下的付款义务。

4. 福费廷

福费廷是在延期付款的大型设备贸易中,出口商把经进口商承兑的,或经第三方担保的,期限在半年至五六年的远期汇票,无追索权地出售给出口商所在地的银行或大金融公司,提前取得现款的一种资金融通形式,所以属于国际融资方式的一种类型。一般而言,银行或大金融公司买断票据后,放弃对出口商所贴现款项的追索权,承担票据拒付风险。

5. 对等贸易

对等贸易是以货物或劳务作为偿付手段的一种贸易方式。交易双方都有进有出,并求得各自的收支基本平衡。对等贸易有多种表现形式,基本的三种形式分别是易货贸易、互购和补偿贸易。其中,最简单最古老的一种方式易货贸易,就是双方都用自有的货物来交换对方的同价值的货物。对等贸易是伴随着国际间贸易的不断发展而逐步递增的。

WILEY

DATA AND ANALYTICS IN ACCOUNTING

AN INTEGRATED APPROACH

会计领域中的 数据与分析

[美] 安·杜拉宁 (Ann Dzurainin)


[美] 吉多·吉尔茨 (Guido Geerts)

[美] 玛格丽塔·伦克 (Margarita Lenk)

著

刘志远 主译

中国财经出版传媒集团

 经济科学出版社
Economic Science Press

会计领域中的数据与分析

[美] 吉多·吉尔茨 Guido Geerts

[美] 安·杜拉宁 Ann Dzurainin 著

[美] 玛格丽塔·伦克 Margarita Lenk

刘志远 主译

中国财经出版传媒集团



经济科学出版社
Economic Science Press

·北京·

图字号：01 - 2024 - 0288

Data and Analytics in Accounting: An Integrated Approach by Guido Geerts, Ann Dzurainin, Margarita Lenk, ISBN: 9781119722991

Copyright © 2023 by John Wiley & Sons, Inc.

All Rights Reserved. This translation published under license. Authorized translation from the English language edition, Published by John Wiley & Sons, Inc.. No part of this book may be reproduced in any form without the written permission of the original copyrights holder. Copies of this book sold without a Wiley sticker on the cover are unauthorized and illegal.

图书在版编目 (CIP) 数据

责任编辑：李 林
责任校对：隗立娜
责任印制：范 艳

会计领域中的数据与分析

[美] 吉多·吉尔茨 Guido Geerts

[美] 安·杜拉宁 Ann Dzurainin 著

[美] 玛格丽塔·伦克 Margarita Lenk

刘志远 主译

经济科学出版社出版、发行 新华书店经销

社址：北京市海淀区阜成路甲 28 号 邮编：100142

总编部电话：010 - 88191217 发行部电话：010 - 88191522

网址：www. esp. com. cn

电子邮箱：esp@ esp. com. cn

天猫网店：经济科学出版社旗舰店

网址：http://jjkxcbs. tmall. com

北京中科印刷有限公司印装

889 × 1194 16 开 30 印张 750000 字

2024 年 3 月第 1 版 2024 年 3 月第 1 次印刷

ISBN 978 - 7 - 5218 - 5634 - 7 定价：99.00 元

(图书出现印装问题，本社负责调换。电话：010 - 88191545)

(版权所有 侵权必究 打击盗版 举报热线：010 - 88191661)

QQ: 2242791300 营销中心电话：010 - 88191537

电子邮箱：dbts@ esp. com. cn)

推荐语一

会计师职业以“确认—计量—记录—报告”四段式循环模式处理数据为基本业务，在当前大数据—机器学习—人工智能时代，面对着怎样的挑战？

现实社会中，任何经济或非经济组织的运行，都面对着各种各样的“事件（events）”。会计系统的缺陷在于，仅仅针对经济事件中的“会计事件”加以处理，所以基于会计系统的信息报告，总是与现实存在脱节，总是不能满足信息报告使用者的管理需求和决策需求。当今商业环境，管理者可以获得数量空前的数据。譬如企业，日常运营过程中会产生大量与客户偏好、供应商行为和营销运营相关的数据。在信息和数据技术迅速发展和普及的当下，企事业单位已经不需要数量庞大的基础会计队伍，存量会计人员转型和新进入会计人员适应新时代，必须注重数据分析，通过数据分析技术，为管理者和决策者提供除常规报表之外更有价值的信息产品。

《会计领域中的数据与分析》为数据与分析教材，该教材专门为高校会计专业编写，注重培养学生从利益相关者需求出发，综合利用内部、外部以及财务与非财务数据，有助于培养学生在当今快速变化的会计行业中取得成功所必需的数据分析技能，通过有效规划、执行和沟通数据分析，提高综合分析能力，助力管理者和决策者解决复杂的现实问题。

北京大学光华管理学院教授 王立彦

2023年10月

推荐语二

第四次浪潮扑面而来，从会计电算化到企业资源计划（ERP），再到财务共享服务，技术演进推动财务的变革，财务连接更加广泛、流程更加智能。作为财务转型与技术创新的引领者，中兴新云深度参与了中国大型企业集团的财务转型进程，帮助其实现从财务共享服务，到财务一体化，再到财务数字化转型的全过程，在这个历程中，我们愈加确信：未来十年，将是财务数字化的十年。利用数据技术赋能会计工作，让数据走向决策、财务走向财经，这是财务变革的必然趋势。

财务数字化时代，企业财务部门需要确立新的组织定位，进化成为企业的财经数据中心，广泛收集财务、业务、产业等数据，经过数据加工和算法模型，为战略、经营和管理赋能；财务人员需要适应更加丰富的财务职能工作，从财务会计走向管理会计，健全成本、绩效、预算、合规、风控管理能力，同时拓展经营支持职能，成为企业数据的分析者和经营状况的描绘者，利用数字技术采集、处理和展示数据背后隐藏的价值。数字化变革趋势对财会从业者的知识及能力结构提出了新的要求。不少高等院校已在进行财会专业的革新，会计学科与数据科学深度融合，会计数据分析、数字技术应用等内容已经逐步融入财会人才培养体系中。企业财务人员也不应故步自封，面对新的发展要求、新的职能定位、新的岗位需求时，财务人员应该抓住变革契机，主动转变观念，学习新的技能，成为“懂会计+懂管理+懂技术+懂数据+懂战略”的综合型人才。

《会计领域中的数据与分析》在以刘志远教授为首的教授团队的努力下顺利出版，是财务和会计领域的重要参考教材。本教材对数据、分析、技术的深刻认识；对计划—分析—报告这一数据分析过程的系统解读；对发现机会、理解变化、解决问题、流程和绩效评估等分析动机的详细阐释；以及对描述型、诊断型、预测型、规范型分析类型的全面介绍，都是重要的分析理念和数据思维。本教材还列举了20种数据提取、转换和加载模式，算法的7种信息建模模式，以及6种常见的会计数据结构，具有很强的实践性，也是财务人员应当具备的数据技能。

本教材也是会计领域数据分析的应用工具书，具备较强的可读性和启发性，不仅能够帮助读者培养数据意识和思维，而且向读者展示了如何在财务实务中运用数据分析能力解决问题的过程和方法，培养读者利用数据与技术工具解决财会管理中实务问题的实践能力，对于高校会计专业学生和企业财务工作从业者而言，都具有较高的学习和参考价值。

深圳市中兴新云服务有限公司总裁 陈 虎

2023年10月

译者前言

随着数智时代的到来和快速发展，数据、分析和技术正在深刻地改变着会计行业的所有领域。智能化等数智技术在优化会计任务处理方面具有巨大优势，诸如创建会计分录、编制财务报表、税务申报等重复性的会计工作都可以通过这些技术高效率地完成，使大量从事基础会计工作的会计师面临着被技术替代的严峻挑战。而数智技术的不断迭代与升级，使商业行为所产生的数据被完整地捕获并记录下来成为可能。持续增长的数据可得性，不断增加的用来分析数据的技术和工具，又为会计师利用数据和分析提供财务洞见，支持业务决策提供了巨大的机遇。这意味着，未来的会计师必须要跳出传统会计思维与会计工作的局限，成为思维机敏、有创造力、通晓数据和技术的专业精通人士，具备运用数据和技术推动业务变革，利用数据和分析创造价值的高超能力。

如何培养适应数智时代需要的未来会计师，是会计教育界必须要回答的时代命题。近年来，我国高等院校的会计专业纷纷开始行动，尝试将数智技术、数据分析等内容融合到会计专业人才培养中，并取得一定的成效。但无可否认的是，这种探索和变革依然面临着很多的现实问题，需要从多个方面持续努力加以解决。缺乏合适的可供在教学中使用的教材可能就是其中迫切、重要的问题之一。

《会计领域中的数据与分析》一书由威利（WILEY）出版社于2022年11月在美国出版，是一部专门针对高校会计专业而编写的分析与分析教材。旨在培养学生在当今快速变化的会计行业中取得成功所必需的数据分析技能，使其获得通过有效规划、执行和沟通数据分析解决不同会计领域现实问题的职业能力。该教材在美国出版后受到美国高校会计专业的广泛关注和采用。本教材具有以下五个特点。

一是建立了一个数据分析的计划、分析、报告（PAR）三阶段框架。该框架可以帮助学生了解其在未来的会计职业生涯中将会遇到的实际数据分析项目从开始到结束的完整过程，并为之做好充分的准备。

二是自始至终强调数据分析过程中批判性思维技能的重要性。建立了一个利益相关者、目的、替代方案、风险、知识、自我复盘（SPARKS）的批判性思维框架，并将其整合进数据分析过程中的每一步。学生在学习数据分析技能的同时，也可以学习到如何将批判性思维应用到该技能中，进而精进自己的批判性思维技能。

三是将数据分析有机地嵌入和应用到不同会计领域的现实问题中，强调通过现实世界的真实数据集进行数据探索与解释。本教材教学案例及课后习题紧扣财务会计、审计、管理会计、税务会计、会计信息化等会计领域实际工作中可能面临的真实问题，训练学生练习如何将数据分析应用于这些会计话题。每章还配备了与案例配套的数据集用于实操训练或课后作

业，以帮助学生建立现实情景下运用数据分析技能解决现实问题的能力。

四是通过自助式商业智能软件（SSBI）的应用与演示，强化学生从确立分析目标、数据获取、数据整理、数据导出、数据分析、数据可视化、数据分析报告到沟通与演讲全流程的综合能力。本教材主要以 Excel 为教学演示工具，同时也简要介绍了 Tableau、Power BI、SQL 和 Alteryx 等其他数据工具的使用，允许教师在创建课程的学习结构时选择一个或多个软件进行讲解。

五是以对数据的敏感度及使用数据解决问题的逻辑素养这一数据分析工作的核心基础为经、以各会计领域的现实问题为纬构建了完整而丰富的教学内容体系，为教学选择提供了灵活性。学校可以根据自己的培养方案和师资状况，将本书作为会计专业数据分析相关的单独一门课程的教材使用，也可以按照会计领域将相关内容分别整合应用到不同的会计专业课程中，实现会计专业课程与数据分析的有机融合。

“他山之石，可以攻玉”。本教材在美国出版后，版权引进方——中国图书进出口（集团）有限公司对国内十几所高校进行了调研，获得了相关高校会计专业负责人和课程教师的积极响应与参与，最终于 2023 年 3 月确定引进该教材并翻译出版，以为我国高校会计专业的会计数据分析相关课程的教学提供参考教材。

为了高效率、高质量地翻译本教材，由笔者担任主译，联合了其他九所高校会计专业的相关教师共同承担了该教材的翻译工作。翻译工作的具体分工如下。

第 1 章为南开大学商学院刘志远教授；第 2 章为江西财经大学会计学院吴志斌老师；第 3 章为石河子大学经济与管理学院鲁春洋老师；第 4 章为华中科技大学管理学院郭炜教授；第 5 章为天津商业大学会计学院白默教授；第 6 章为重庆工商大学会计学院钟炼老师；第 7 章为北京工业大学经济与管理学院陈硕老师；第 8 章为上海立信会计金融学院会计学院林振兴副教授；第 9 章为上海商学院财务金融学院孙莉副教授；第 10 章为上海对外经贸大学会计学院张天舒教授。原中国图书进出口（集团）有限公司，现香港会计师公会的苗雯女士在本教材翻译工作的具体组织、第 1 章初稿的翻译、全书翻译和初稿整理等过程中做了大量工作。

翻译初稿完成后，由刘志远对全部译稿进行了审校和修订，并最终定稿。

本教材的引进翻译工作也得到了北京大学王立彦教授、中兴新云陈虎总裁的持续关注与支持，并欣然作序对本教材做出独到解读。

翻译过程中，对于演示软件直接输出的图表、与数据集有关联的图表等，我们基本保留了原图表的相关内容，没有翻译成中文。这样做一是为了方便教学过程中演示软件的使用；二是保持相关图表变量与数据集变量之间的一致性，以便使用例题配套的数据集。同时，我们一方面删除了少量与中国制度背景不相符的内容；另一方面适当补充了中国制度背景的相关内容，目的是使本教材的内容更加贴近中国的实际。

由于译者水平所限，本教材的翻译定会存在一些疏漏乃至错误之处。我们诚挚希望广大教师和同学在使用过程中指出这些不足，以便再版时进行修订和完善。让我们共同为培养适应数智时代需要的未来会计师而努力！

刘志远

2023 年 9 月 21 日于南开园



目 录

Contents

第 1 章 会计中的数据与分析	1
本章概览	1
本章学习任务	2
1.1 数据与技术如何改变会计行业?	2
1.2 数据分析过程包括哪些阶段?	10
1.3 什么是数据分析思维?	17
1.4 如何应用数据分析思维?	22
章节回顾与练习	28
第 2 章 数据分析的基本技能	32
本章概览	32
本章学习任务	33
2.1 为何了解数据存储有助于解答问题?	33
2.2 电子表格函数如何分析大量数据?	40
2.3 如何为分析构建数据集?	45
2.4 哪些描述性度量值有助于理解数据?	58
2.5 可视化在数据分析中的应用	75
章节回顾与练习	83

第3章 数据分析的动机与目标	88
本章概览	88
本章学习任务	89
3.1 动机如何影响基于目标的数据分析问题?	90
3.2 什么是描述性目标?	95
3.3 什么是诊断性目标?	100
3.4 什么是预测性目标?	104
3.5 什么是规范性目标?	112
3.6 会计领域实务中的数据分析动机与目标	118
章节回顾与练习	123
第4章 规划数据与分析策略	129
本章概览	129
本章学习任务	130
4.1 会计人员如何设计数据分析项目?	130
4.2 选择用于分析的数据时应考虑的因素	141
4.3 选择分析类型时应考虑的因素	147
4.4 数据策略和分析策略在不同会计实务领域的区别	155
章节回顾与练习	165
第5章 分析：数据准备	171
本章概览	171
本章学习任务	172
5.1 什么是数据概要分析?	173
5.2 提取、转换和加载（ETL）数据的含义	182
5.3 数据提取模式	187
5.4 转换列的模式	192
5.5 转换表格的模式	203
5.6 转换模型的模式	209
5.7 哪些模式适用于数据加载?	216
章节回顾与练习	222

第6章 分析：信息建模	232
本章概览	232
本章学习任务	233
6.1 什么是信息建模?	233
6.2 信息建模算法的实现模式	240
6.3 哪些模式有助于开发和实现会计信息模型?	252
章节回顾与练习	277
第7章 分析：数据探索	286
本章概览	286
本章学习任务	287
7.1 什么是数据探索?	287
7.2 数据探索中的数据关系可视化	297
7.3 通过整合数据关系进行数据探索	315
章节回顾与练习	327
第8章 解释数据分析结果	334
本章概览	334
本章学习任务	335
8.1 如何从数据分析中获得结论?	336
8.2 批判性思维和数据分析解释之间的关系	341
8.3 如何知道分析是合理的?	346
8.4 如何确定描述性和诊断性分析的效度和信度?	351
8.5 如何评估预测分析和规范性分析的效度和信度?	362
章节回顾与练习	371
第9章 沟通数据分析结果	375
本章概览	375
本章学习任务	376
9.1 如何讲述一个数据故事?	376
9.2 创建有效数据可视化的步骤	385
9.3 有效可视化的特点是什么?	393

9.4 什么会使数据可视化产生误导?	408
9.5 在现场演讲中如何使用数据?	416
章节回顾与练习	420
第 10 章 会计领域中数据与分析的最新进展	425
<hr/>	
本章概览	425
本章学习任务	426
10.1 哪些数据趋势正在影响会计实务?	426
10.2 会计人员采用了哪些最新的分析工具?	438
10.3 数据和分析的发展对会计职业领域有何影响?	454
章节回顾与练习	462

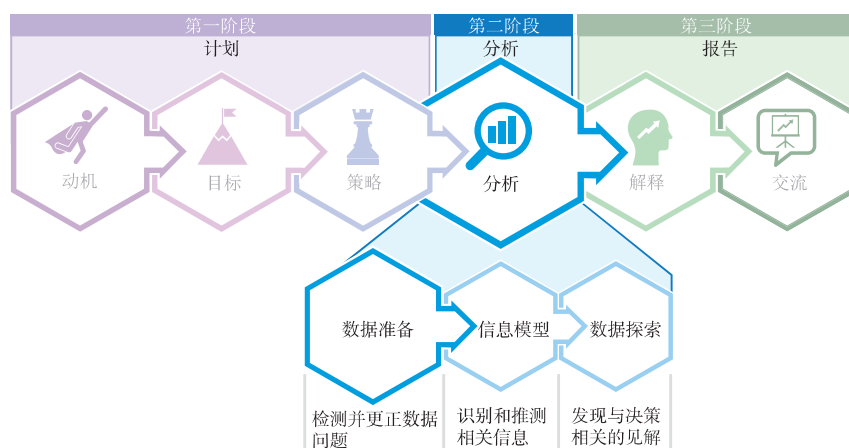
分析：数据准备

本章概览

截至目前，我们已经掌握了对数据分析项目进行计划的方法，包括阐明项目动机、确定项目目标和设计可以成功完成项目的策略。完成计划阶段后，就进入了数据分析阶段。分析阶段有三个任务：数据准备、构建信息模型和数据探索。本章将重点讲述数据分析阶段的第一个任务——数据准备。数据准备是数据分析项目过程中最耗时的任务。仅此一项任务，就可能占据项目总时长的75%！这是因为数据准备涉及多种活动。在进行数据准备时，要注意两个关键因素：

1. 数据的质量将影响基于数据的洞见和决策的质量。换言之，不良的数据会导致低质量的决策。
2. 数据的结构将决定可以在多大程度上对其进行有效的分析。

本章首先对这两个关键的数据准备过程进行详细讨论；其次将介绍20种用于数据分析的数据提取、转换和加载模式。



专业洞察：如何应对数据准备的挑战？

Bill 是一家大型会计师事务所的审计数据分析业务总监。

“在获得会计学位之后，我逐渐转向了数据分析业务。我着迷会计与技术 在审计业务中如何实现融合，其中总有新的机遇和挑战需要去努力。

在会计师事务所处理数据时，一个关键的挑战是我们客户的规模和形式的各异，他们的 ERP 系统包括了从传统的专有系统到日益现代化的、基于云计算且拥有复杂报告和分析能力的整合系统。这意味着数据的清洁度、质量、颗粒度、数量和结构方面可能有较大差异。这提升了 ETL（提取—转换—加载）解决方案的重要性，以便帮助我们的专业人员有效地从客户端系统中提取数据，将数据转换为通用格式，并最终将数据加载到我们的分析平台中。

为成功获取数据，公司的数据专业人员必须能够理解和评估客户的数据，确保其相关性和可靠性。这包括了解怎样与我们的客户进行有效对话，以便准确地传达我们的数据要求，对话人员通常包括客户的 IT 部门和我们内部数据提取方面的专家。在本章中将要学习的技能，与你想要在今注册 会计师行业中成为一个成功的数据分析师的职业生涯，都是高度相关的”。



本章学习任务

学习目标	主题	应用
LO 5.1 解释数据概要分析的过程	<ul style="list-style-type: none"> 📍 调查数据质量 📍 调查数据结构 📍 决定并告知 	确定数据质量问题
LO 5.2 描述提取—转换—加载（ETL）过程	<ul style="list-style-type: none"> 📍 数据提取 📍 数据转换 📍 数据加载 	合并表格进行分析
LO 5.3 将模式应用于提取数据	📍 两种数据提取模式	使用模式提取数据
LO 5.4 将模式应用于转换列	📍 八种列转换模式	使用列转换模式
LO 5.5 将模式应用于转换表	📍 四种表转换模式	使用模式对表进行转换
LO 5.6 将模式应用于转换模型	📍 三种模型转换模式	绘制星型架构
LO 5.7 将模式应用于数据加载问题	📍 三种数据加载模式	评估表之间的关系

5.1 什么是数据概要分析？

学习目标 ①

解释数据概要分析的过程

企业因“脏数据”、不准确或不能完全描述企业经济活动的的数据，每年损失数十亿美元。使用“脏数据”可能会导致各种问题，从不准确的定价到将账单发送给错误的客户（或完全不收取这些账单），再到无法检测出欺诈行为。数据准备有助于避免这些问题或更多其他问题。

数据准备是在对数据进行加工和分析之前进行概要分析、清洗、重组和集成的过程。它有助于确保数据的高质量，进而改进决策。数据概要分析是调查数据质量和结构的过程。它包括三个部分：

1. 调查数据质量：搜寻数据中的异常部分。即数据是“脏数据”吗？
2. 调查数据结构：找到组织数据和改进分析的最佳方法。
3. 决策与通知：决定是否可以解决已发现的问题，解决问题所付出的成本代价是多少，并考虑如果不解决问题可能产生的后果。

最后一个阶段是确定必须更改的内容，并引导提取、转换、加载（ETL）过程。如图 5.1 所示，数据准备是数据概要分析（本段内容讨论的主题）过程与 ETL（接下来介绍的）过程之间的持续协作。目前需要记住，数据概要分析会检测数据问题，ETL 过程会纠正它们。

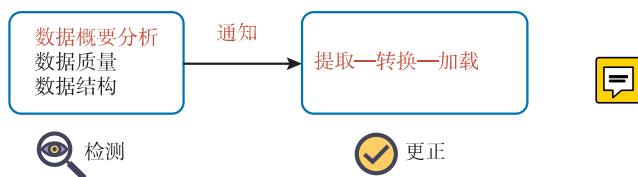


图 5.1 数据准备过程

调查数据质量

当我们讨论将要使用的数据“质量”时，意思是数据对于决策的适用性。通过评估数据质量识别出数据集里存在的缺陷值，可以揭示是否存在必须加以清洗的数据。以下是可以执行此操作的不同方法：

- 规则驱动法是一种自上而下的方法。在数据之间定义逻辑关系或规则，并对其进行测试，确定数据是否符合该关系或规则。可以指定的规则数量几乎不受限制。示例包括：
 - 职责分工。
 - 库存数量不能为负数。

- 汽车不能租给 16 岁以下的未成年人。
- 销售只能通过有效订单执行。
- 探索和推理法是自下而上的方法。它的目标是通过从多种不同角度检查数据来发现异常。排序、频率分布和离群值分析是一些强大的适用于探索目的的技术示例。第二种自下而上的方法——推理，是一种依靠计算机算法来识别异常的方法。

这些不同的方法均可识别数据异常，当数据不符合正确性、有效性、一致性和完整性的预期时，就会显示异常。

正确性

数据描述关于实体的事实，例如客户名称、产品价格或交易日期。当赋予某一实体特征的值有误时，数据就是不正确的。例如，客户可能居住在新泽西州（NJ），但记录成纽约（NY），或者产品的价格为 252 美元，但列出的价格为 225 美元。

数据不正确很常见，以下是一些真实发生的数据差错的案例。对于其中的企业而言，数据错误带来的后果可能会很严重：

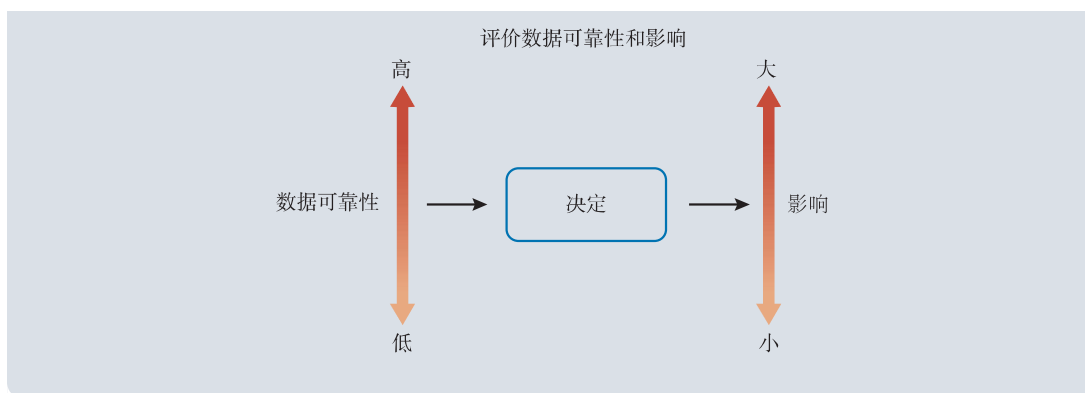
- 美国某航空公司以不到 7 美元出售了一张从美国大陆出发去往夏威夷的往返机票。另一个类似的例子是，美国某航空公司的一个客户只用了 7 个常飞里程就兑换了一张从美国到中国香港地区的头等舱往返机票。
- 客户以每晚 10 美元的价格预订了加利福尼亚州帕萨迪纳市的一个豪华酒店的房间。
- 某著名商超以不到 35 美元的价格出售了一台跑步机。

有以下几种方法可以检测此类错误。当定价不正确时，例如，当由于错误导致大幅度折扣出售价格昂贵的机票时，需求突然激增可能表明存在问题。概要分析技术，例如，探索离群值也很有效。无论如何，我们需要始终保持批判性地思考正在分析的数据并注意发生的异常情况。

应用批判性思维 5.1：评估数据质量

使用可靠的数据可以消除决策中的偏见和猜测，使用低质量数据会增加风险。因此，评估与不完美数据相关的风险非常重要。需要考虑以下两个特征（风险）：

- 可靠性：从企业（ERP）系统中提取的经过数百次内部控制的数据，比从社交媒体中提取的带情绪的数据更可靠。
- 影响：同样地，推出新产品线的影响远大于投资新名片的决定。影响越大就应该越愿意对提高数据的可靠性进行投资。



有效性

数据概要分析的另一个组成部分是设计有效规则。这些规则定义了什么是可接受的，什么是不可接受的，并且它们可以采取多种样式和形式。例如，订单应在发货之前完成，或者电子邮件地址应包含符号@。

有效性和正确性有时会被混淆。想象一下，一家企业将它们的物品运送到中西部各州：伊利诺伊州、印第安纳州、艾奥瓦州、堪萨斯州、密歇根州、明尼苏达州、密苏里州、内布拉斯加州、北达科他州、俄亥俄州、南达科他州和威斯康星州。以下是一些方案：

- 货物被运送到俄亥俄州（OH），并将俄亥俄州作为系统中状态字段的值输入。该值既有效又正确。
 - 有人不小心将明尼苏达州（MN）记录为运往密歇根州（MI）货件的目的地。输入的值有效（明尼苏达州是可接受的输入状态），但不正确。
 - 企业收到宾夕法尼亚州（PA）客户的订单，并且货物已发货。PA 作为状态输入系统。虽然系统视该值为无效，但它是正确的（PA 是正确的状态，但公司不能在那里销售商品）。
 - 收到来自宾夕法尼亚州（PA）客户的相同订单，货物将被运送到那里。NA（不适用）作为状态输入系统。在这种情况下，输入的值既无效又不正确。
- 不正确的值意味着赋予了错误的值，无效的值意味着赋予了不被接受的值。

一致性

除了正确性和有效性之外，数据还应保持一致性。当相同的特征以多种方式表示时，就会发生数据不一致。同时使用 MI 和 Michigan 来表示同一州会导致数据不一致。使用 mgr、mng 和 manager 来描述相同的工作职位也会导致数据不一致。这些不一致给分析过程带来困难。例如，如果密歇根州（MI）和密歇根州（Michigan）有单独的合计收入，这种情况下很难确定整个州的总收入。

我们如何识别这样的不一致性？以下是两个分析技术：

- 创建包含所有非重复值的列表，然后排序和查看。mgr、mng 和 manager 的不一致值可能会立即引起注意。

- 生成频率表，或生成计算值出现次数的表。频率较低的值可能表示数据不一致。

完整性

正确、有效和一致的数据只有在完整的情况下才是准确的，因为不完整的数据可能会导致歪曲的洞见。数据可能因为两种情况不完整。

一种情况是缺失实例，指概念发生但未记录的情况，例如，发生销售货物的行为，但未记录销售交易。缺失实例（如缺失的销售交易）可以通过间隔分析来识别。如果每个销售发票都有一个序号，则缺少某一序号可能表示发生了但未记录的销售。

另一种情况是交易记录时出现缺失值，但我们并没有所有特征值的信息。记录结果会有空单元格。这种情况往往发生在记录了客户信息，但并没有记录客户的电子邮件地址。术语 null 通常表示缺少或未知的值。因此评估缺失的信息如何影响决策尤为重要。

调查数据结构

除了数据的质量外，还要调查评估数据结构是否使数据分析简单高效。

明确的描述

在数据探索和解释过程中，表中的列名成为变量。我们应用数学运算，例如加总求和，并将它们用作可视化（如饼图）的一部分。但命名不当的列名，会使开发信息模型和进行分析变得复杂化。

列名称应正确、直观且清晰，因为它们分析数据库的一部分，而分析数据库可能有许多不同的用户。分析数据库是用于分析目的的集成数据集。在开发分析数据库时，明确的描述应作为优先选项。以下是正确和清晰的列标题的一些示例：

- 年龄列应包含数字（如 32）而不是日期（如 6/2/1990）。
- 客户名称加 ID 比只写 ID 更具描述性。
- “客户生日”比“C 生日”更容易理解。

应用批判性思维 5.2：创建一个共享的，可理解的词汇表

创建共享词汇表需要确定谁将使用分析数据库（利益相关者）：

- 如果为其他人准备数据，请确保为列名和表名创建一个易于理解和使用的词汇表。
- 例如，数据科学家通常使用诸如 DCustomer 和 FSales 之类的表名，其中，D 和 F 指代某些类型的表。但是这些术语可能会让不是数据科学家的其他利益相关者难以理解！

使分析简化的表格结构

分析包括对数据进行聚合或分组，然后以多种方式从不同的角度检查它们，称为“切

片”。以下是在多个地区销售的企业对销售交易数据进行切片的示例：

- 合计：计算总销售额。
- 切片：按区域细分总数，更详细地检查区域销售额。

有些数据结构非常适合这些合计和切片过程，但有些数据结构不适合，因此学习结构化数据的最佳实践方法至关重要。两种常见的做法是单值列和平面表。

单值列出于分析目的，每个单元格应包含一个描述一个特征的值。也就是说，每一列都应该是单值的。同一单元格中的两个或多个值加大分析难度。下面是违反单值规则的两种情况。

复合列组合了两个或多个特征的值。表 5.1 (A) 中的姓名列结合了姓氏、名字和认证资格。如果您需要生成包含所有拥有 CPA（注册会计师）面板 (A) 或面板 (B) 的员工的列表，您希望使用哪个数据集？虽然两个数据集的数据相同，但面板 (B) 使回答这个问题变得更加容易。

表 5.1 复合列与单值列

(A) 复合列	(B) 单值列	
姓名及认证资格	姓名	认证资格
科尔拉·基沙, 助理护士	科尔拉·基沙	助理护士
洛佩兹·亚历杭德罗, 注册会计师	洛佩兹·亚历杭德罗	注册会计师
基·金, 助理护士	基·金	助理护士
大卫·朱莉, 注册会计师	大卫·朱莉	注册会计师
马龙·摩西, 助理护士	马龙·摩西	助理护士
布斯莱普·比尔, 注册会计师	布斯莱普·比尔	注册会计师
波托姆斯·凯姆, 助理护士	波托姆斯·凯姆	助理护士
德斯庞廷·马克, 注册会计师	德斯庞廷·马克	注册会计师
卡明斯基·伊万卡, 助理护士	卡明斯基·伊万卡	助理护士

在多值列中，单元格包含具有相同特征的多个值。表 5.2 (A) 中的“认证资格”列列出了员工的所有认证资格。同样地，如果需要生成包含所有拥有 CPA 资格的面板 (A) 或面板 (B) 的员工的列表，你希望使用哪个数据集？虽然两个数据集的数据相同，但使用面板 (B) 中的单值列更容易回答此问题。

表 5.2 多值列与单值列

(A) 多值列		(B) 单值列	
姓名	认证资格	姓名	认证资格
科尔拉·基沙	国际信息系统审计师	科尔拉·基沙	国际信息系统审计师
洛佩兹·亚历杭德罗	注册会计师、国际信息系统审计师	洛佩兹·亚历杭德罗	国际信息系统审计师
基·金	注册管理会计师	洛佩兹·亚历杭德罗	注册会计师

续表

(A) 多值列		(B) 单值列	
姓名	认证资格	姓名	认证资格
大卫·朱莉	注册会计师、注册管理会计师、国际信息系统审计师	基·金	注册管理会计师
马龙·摩西	注册会计师	大卫·朱莉	国际信息系统审计师
布斯莱普·比尔	注册会计师、注册管理会计师	大卫·朱莉	注册管理会计师
波托姆斯·凯姆	注册管理会计师	大卫·朱莉	注册会计师
德斯庞廷·马克	注册会计师	马龙·摩西	注册会计师
卡明斯基·伊万卡	注册管理会计师、国际信息系统审计师	布斯莱普·比尔	注册管理会计师
		布斯莱普·比尔	注册会计师
		波托姆斯·凯姆	注册管理会计师
		德斯庞廷·马克	注册会计师
		卡明斯基·伊万卡	国际信息系统审计师
		卡明斯基·伊万卡	注册管理会计师

平面表结构化数据的另一种最佳做法是使用平面表，其中，列标题不包含对分析目的有用的数据值。对于数据分析，平面表结构优于交叉制表结构。

表 5.3 (A) 显示了一个交叉制表，表示汽车车型和销售人员这两个变量之间的关系。

表 5.3 交叉制表结构与平面制表结构

	(A) 交叉制表				(B) 平面制表		
	吉姆	珍	埃洛迪	卡洛斯	车型	销售人员	已售单位
福克斯	0	1	5	4	福克斯	卡洛斯	4
野马	1	2	3	3	野马	卡洛斯	3
捷安特	4	1	0	2	捷安特	卡洛斯	2
探险者	2	2	0	1	探险者	卡洛斯	1
					福克斯	埃洛迪	5
					野马	埃洛迪	3
					捷安特	埃洛迪	0
					探险者	埃洛迪	0
					福克斯	珍	1
					野马	珍	2
					捷安特	珍	1
					探险者	珍	2
					福克斯	吉姆	0
					野马	吉姆	1
					捷安特	吉姆	4
					探险者	吉姆	2

- 第一列是不同车型的值：福克斯、野马、捷安特和探险者。
- 顶端的列标题显示销售人员的值。
- 横截面单元格显示销售人员已售出特定型号的单位数。销售人员埃洛迪已售出五台福克斯车型。

交叉制表具有分析限制。面板（A）列标题中的名称是对数据分析有用的数据值的示例。列标题不能用于筛选或分组依据操作。

表 5.3（B）显示了使用扁平结构的相同数据。吉姆、珍、埃洛迪和卡洛斯为“销售人员”列的值，可用于筛选和分组操作。思考：你希望面板（A）中的表格还是面板（B）中的表格来回答诸如“Jane 和 Jim 卖了多少野马”这样的问题？表 5.3（B）中的平面结构使分析更容易，因为可以按销售人员进行筛选，然后对“已售单位”列中的值求和。

使分析简化的数据模型

数据模型定义不同表之间的相互关系。分析数据模型应便于业务用户（如会计人员）理解，并能够被计算机高效地进行处理。分析数据模型的推荐结构是星型架构，它具有两个特征（见图 5.2）。如图 5.2 的顶部所示，星型架构由两种类型的表组成：事实数据表和维度表，以及它们之间的关系。

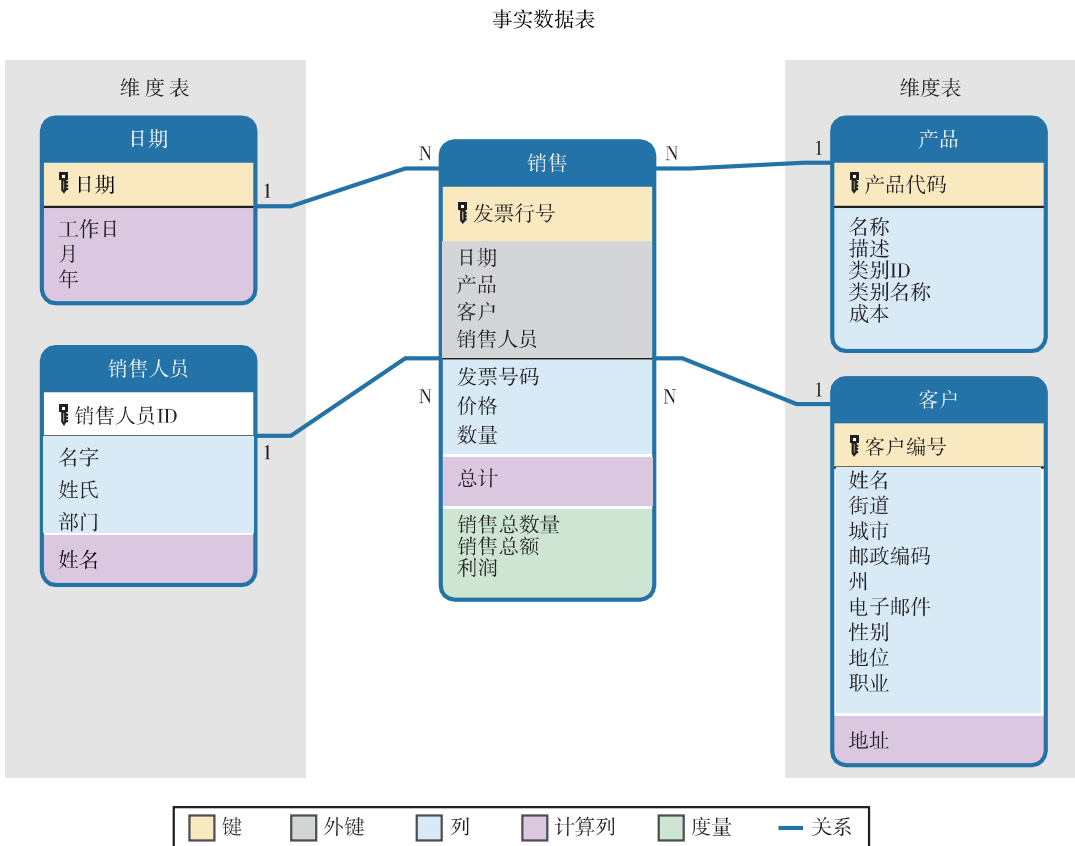


图 5.2 星型架构

事实数据表在会计环境中，事实对应于业务交易，如订单、销售、采购和付款。图 5.2 中心的事实数据表包含销售数据。事实数据表中的数据以高粒度级别捕获，可以避免对分析施加额外的约束（细节级别称为粒度）。在图 5.2 中，粒度是一个发票行，它是在特定日期发运给特定客户并涉及特定销售人员的特定产品。

虽然事实数据表中的列数通常很小，但它们通常有很多行。公司通常在同一天有数千笔销售额，如果这还不够多的话——想想亚马逊这样的线上零售巨头。事实数据表中的列也大多是定量的，更具体地说，它是累加的。这些列中的值可以作为度量值轻松地加总或组合在一起。然后，可以使用以下类型的表（维度表）中的列对这些度量值进行切片。

维度表维度为分析提供背景，并赋予事实意义。100 美元的销售额并不能提供太多信息。企业想知道的是销售交易何时发生，谁需要支付 100 美元，以及购买了什么商品。图 5.2 中的星型架构具有销售发生日期、销售产品、客户和销售人员的维度表。维度表中的列是可以通过多种方式组合对度量值进行切片的变量。图 5.3 展示了八个问题，这些问题可以通过将销售指标按图 5.2 所示的一个或多个维度切片来回答。



图 5.3 按维度划分的切片指标

虽然维度表中的行数通常很小，但它们通常包含许多列。图 5.2 中的“客户”维度表有 10 列，但没有度量值。实际上，客户维度表通常具有 100 多列。

关系是数据模型的最后一个元素，它连接表格并由图 5.2 中的行表示。星型架构中的所有关系都具有一对多（1-N）基数模式。基数是一种约束，用于定义实体的实例可以参与关系的次数。它可以采用两个可能的值，即 N 或 1：

- N：一个实例可以多次参与关系，没有任何限制。
- 1：一个实例在关系中只能参与一次。

维度和事实数据表之间关系的 1-N 基数模式可以按以下方式解释：

- 1：对于每个事实，每个维度表中只有一个对应的值。对于销售表（发票行）的每

个实例，都有一个日期、产品、客户和销售人员。

- N：每个维度都可能有很多事实。同一天可以有多个销售，同一产品类型可以销售很多次。客户和销售人员都可以参与许多销售交易。

决定并告知

在调查了数据的质量和结构之后，可以进行下一步：

- 停止前进：如果数据不适合，例如，数据质量差，那么使用这些数据进行决策风险太大。同样地，如果数据结构很差，则重组和处理数据可能过于复杂，并且这在经济上是不可行的。

- 重新设计源系统：做出此决定通常是因为需要更正源系统生成的错误。修复后，应对数据源重新进行概要分析。

- 确认数据是否适合分析并继续前进：处理数据需要具有风险意识。也就是说，我们可以在了解潜在可靠性问题及其对决策的影响的情况下继续处理数据。

如果决定向前推进，则下一步就要收集所有必须得到解决的问题的相关信息。



应用练习 5.1

确定数据质量问题

Stufan 是特拉华州冈博罗的一家小商店，售卖毛绒玩具。创始人兼首席执行官 Shanice Parker 希望应用数据分析来更好地了解小店业务。她要求你的公司帮助准备分析使用的数据。

以下是 Stufan 产品数据的示例。请确定数据的三个质量问题。

代码	描述	现有数量
BASHL	害羞的大号，矮人	13
BASHS	害羞的小号，矮人	67
BERL	贵族，柏辽兹	23
CAPH	丁丁，哈多克船长	14
DOCL	大矮人，医生	23
DOCS	小矮人，医生	45
DOPL	笨拙的，大矮人	43
DOPS	笨拙的，小矮人	-3
DUCH	贵族，公爵夫人	44

答案

1. 描述不是单值列。它结合了“描述”和“类别”。
2. 存在一个一致性问题：“贵族”和“贵族”。
3. 存在有效性问题：QOH（现有数量）不能为负数。

5.2 提取、转换和加载（ETL）数据的含义

学习目标 2

描述数据提取—转换—加载（ETL）过程

当数据概要分析过程中检测到问题时，可以通过提取—转换—加载过程（ETL）来纠正这些问题。如图 5.4 所示，ETL 流程通过一个或多个原始数据来构建分析数据库。

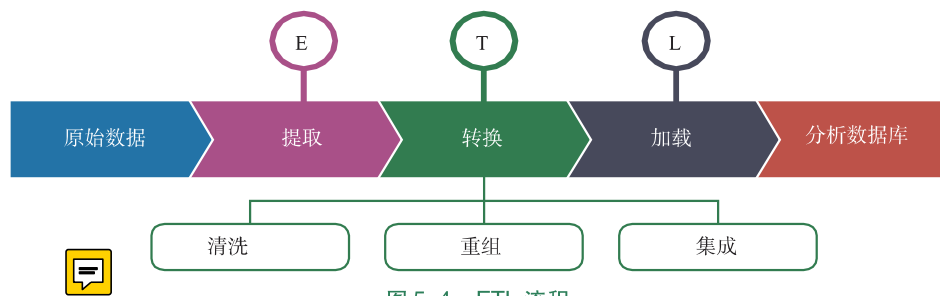


图 5.4 ETL 流程

数据提取

数据提取包括数据传输，是 ETL 流程的第一步：

- 将源数据移动到一个可以转换它们的平台，这个平台通常是一个数据仓库，一个可以存储和分析大型数据集的软件。如 Power BI 和 Tableau。
- 这个过程还包括数据验证，即确认数据已经完整正确的传输。

ETL 通过提供设计直观的数据提取软件程序——数据连接器，可以轻松地从数据库、电子表格、文本文件和许多其他数据源中提取数据。图 5.5（A）展示了一些可以在 Excel 中使用的数据连接器，图 5.5（B）中展示的是可以在 Power BI 中使用的数据连接器。

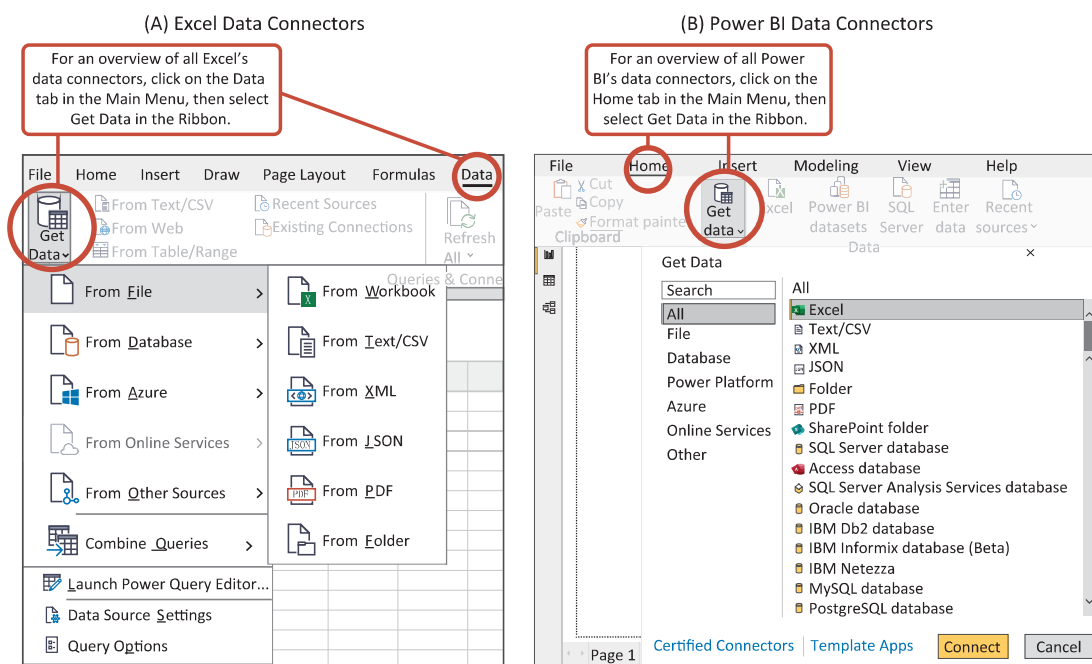


图 5.5 使用数据连接器提取数据

数据转换

原始数据在被提取之后通常并不能直接用于数据分析。而数据转换通过清洗、重组和集成提高了原始数据的可分析性。

数据清洗

数据可能是不正确的、无效的、不一致的或不完整的。因此，清洗数据是数据转换中最重要，也是最耗时的环节之一。清洗数据又名数据清洗或数据清理。它包括添加数据、修改数据和删除数据：

- 如果数据不完整，例如销售交易缺失，可能需要添加数据。处理不完整数据的具体策略是估算，即用估计的值来替代缺失的数据。
- 在数据不正确、无效、不一致或不完整情况下，要用新的数据来替换它们。在记录顾客状态的一栏中将值 NY 替换为 NJ 就是一个例子。
- 删除与分析无关的数据。冗余数据应该被删除，如重复的销售交易。

ETL 工具为一系列数据清洗任务提供了内置功能。如何从表格中删除行？如图 5.6 所示，Power Query、Excel 和 Power BI 的 ETL 工具具有删除重复行、空白行或包含错误行的命令。其他工具也提供类似的操作。

如何替换特定值？如图 5.7 所示，Power Query 的“查找和替换”选项可以解决不一致的问题。



图 5.6 在 Power Query 中删除行

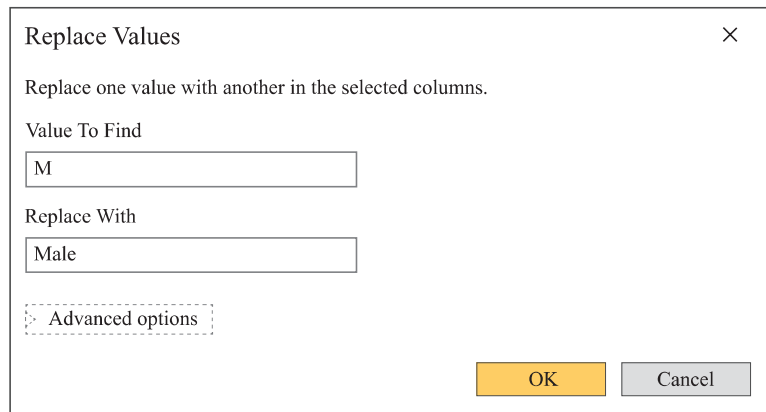


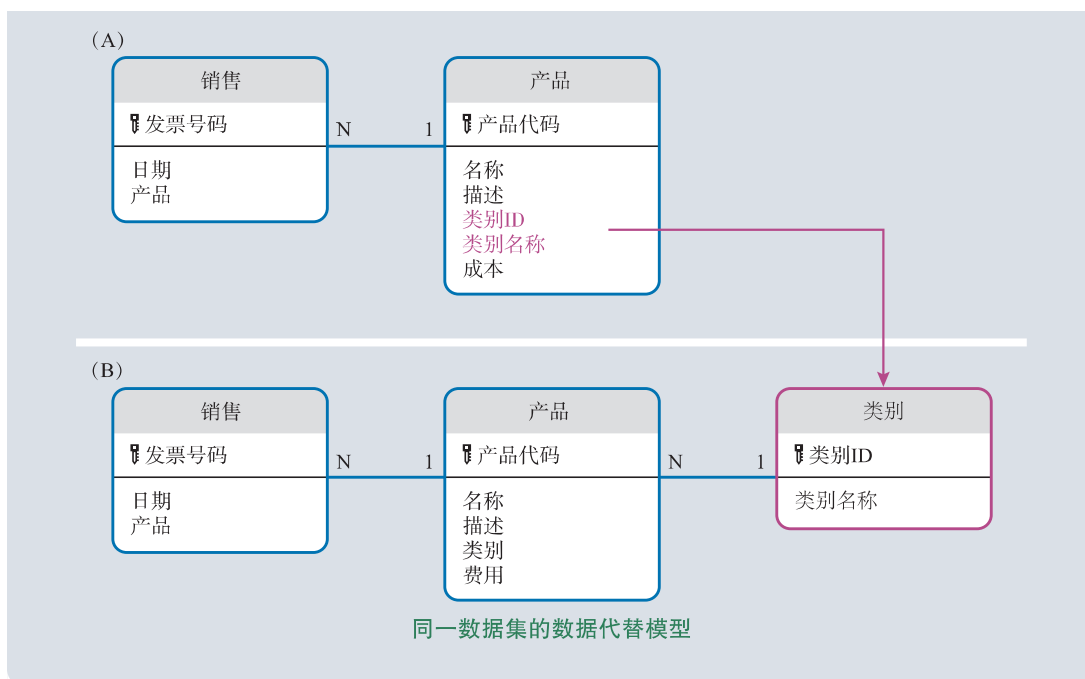
图 5.7 在 Power Query 中查询替换值

如果在数据集的性别一栏中有时使用“Male”，而有时用“M”来表示“Male”，那么性别栏中所有的“M”都可以被替换为“Male”。使用这个工具可以通过查找并替换不一致的值来快速解决数据不一致的问题。

批判性思维的运用 5.3：选择最佳方法

数据准备并不是一门精准的科学，因此判断是建立分析数据库的重要组成部分。始终要考虑好替代方案，分析它们的优劣势，并对它们进行排序。如下所述，面板（A）和面板（B）分别是组织方式不同的相同数据（替代方案）：

- 在面板（B）中，类别成为自己的实体，类别的名称只记录一次。
- 采用面板（A）中的数据模型则意味着同一类别中的所有产品都会重复类别名称。面板（B）节省了空间但模型更加复杂，因为需要在分析过程中额外考虑一种关系。



数据重组

“干净”数据不一定具备使分析变得简单高效的结构。数据重组，也称数据整理（data wrangling or data munging），不改变数据的值，但是会改变数据的组织方式。

ETL 提供了多种方式使数据重组更容易，具体如下：

- 添加和删除列。
- 重命名列和表格。
- 拆分和合并列。
- 拆分和组合表格。
- 换位表和逆透视表。

数据集成

大多数的数据分析项目涉及多个表，通常来自不同的数据源，因此进行分析之前必须集成数据。数据集成是连接相关数据的过程。以下是两种不同的集成方式：

- 通过定义两表之间的关系来连接。使用主键和外键来创建这种关系。必须被指定的其他方面的关系是基数。

- 合并两个或以上含有相同实体统一信息的表格。表可以用两种方式组合，联合（union）可以将数据结构不同的表格合并，得到一个有更多行的表。回顾数据分析基础章节内容，join 或 merge 则将不同表中的数据元素或列结合起来得到一个具有更多列的表。

数据匹配是集成中一个特有的挑战，这是一个比较数据并确定它们是否描述同一实体的过程。如表 5.4 所示的两个客户数据集。面板（A）包含财务信息，面板（B）包含有关客

户的人口统计信息。应该如何调节或匹配客户的姓名？需要解决的具体问题包括：

- 昵称：Jen Pollack versus Jenny Pollack。
- 错别字：Carlos Panetta versus Carlos Paretta。
- 反向名称：Margarita David versus David Margo。

面板 (C) 是合并后的表格。大多数 ETL 工具都为数据匹配提供高级支持。

表 5.4 数据匹配问题

(A) 顾客表：财务状况			(B) 顾客表：人口统计信息			(C) 合并后的顾客表				
姓名	等级	折扣	姓名	性别	年龄	姓名	等级	折扣	性别	年龄
William McCarthy	A	5	Carlos Paretta	M	44	Ann White	B	3	F	28
Margarita David	C	1	David Margo	M	22	Carlos Panetta	A	5	M	44
Ann White	B	3	Jenny Pollack	F	57	Jen Pollack	B	3	F	57
Carlos Panetta	A	5	Bill McCarthy	M	38	Margarita David	C	1	M	22
Jen Pollack	B	3	Ann White	F	28	William McCarthy	A	5	M	38

数据加载

完成清理和转换，数据就会被加载到软件中进行分析。数据加载是使分析数据库可供使用的处理过程。分析数据库通常发布在云端，许多用户可以同时使用。与数据提取一样，数据加载需要验证是否所有记录都已传输以及是否传输正确。



应用练习 5.2

合并表格进行分析

Stufan 首席执行官 Shanice 建议将产品数据列在三个单独的表格中，每个产品类别有一个表格。她的问题是，合并三个表格进行分析是否会出现问题？对此问题你会如何回答？

Aristocats			Dwarfs		Tintin		
编码	种类	现有数量	编码	种类	编码	种类	现有数量
BERL	Berlioz	23	BASHL	Bashful Large	CAPH	Captain Haddock	14
DUCH	Duchess	44	BASHS	Bashful Small	SNO	Snowy	22
MAR	Marie	34	DOCL	Doc Large	TINL	Tintin Large	23

续表

Aristocats			Dwarfs		Tintin		
编码	种类	现有数量	编码	种类	编码	种类	现有数量
THO	Thomas	19	DOCS	Doc Small	TINS	Tintin Small	44
TOU	Toulouse	32	DOPL	Dopey Large			
			DOPS	Dopey Small			
			GRUML	Grumpy Large			
			GRUMS	Grumpy Small			
			HAPL	Happy Large			
			HAPS	Happy Small			
			SLB	Sleepy Big			
			SLS	Sleepy Small			
			SNEL	Sneezy Large			
			SNES	Sneezy Small			
			SNOWL	Snow White Large			
			SNWS	Snow White Small			

答案

- 这三个表不能通过 union 进行组合。因为 Dwarfs 表格的结构与另外两表不同，Dwarfs 表格没有现有数量列。
- 产品的类别由表名决定，不能用于分析。需要在每个表中添加一个类别列。

5.3 数据提取模式**学习目标 ③**

将模式应用于提取数据

每个数据分析项目都有独特的数据准备挑战。虽然没有适用所有项目的单一通用方法，但一组结构化的数据准备模式可以解决大多数挑战。这些模式提示潜在问题，并为在数据集中查找并纠正这些问题提供指导。每种模式都可以识别出一种数据问题，讨论如何使用概要分析方法检测，并解释一种或多种 ETL 方法来纠正问题。将模式视为菜单，你可以选择最适合需求的模式。

Beans 是一家位于密歇根州奥克莫斯的会计公司，提供会计和税务服务。

佩特拉，他们的管理合伙人，已经把改善他们服务的分析工作作为新财年的优先事项，

佩特拉愿意做分析，但她认为很难整合不同的数据来源。假设你是帮助准备数据的会计人员之一。bean 的数据集由 Excel 文件中的四个工作表组成（见表 5.5）。

表 5.5 Beans 数据集中的工作表

工作表	描述
ClData	有关 Bean 客户的信息。
Employee	有关 Bean 员工的一般信息。
E - Dem	有关 Bean 员工的人口统计信息。
Service	有关 2025 年 1 ~ 7 月提供的服务的信息。

识别可用数据并创建数据字典——一个图表指出哪些为可用数据以及在哪里可以找到。数据字典记录每个字段的不同信息片段，包括名称、内容的简要描述、数据类型、字段为主键还是外键，以及字段是必填字段还是选填字段。数据字典通常称为元数据——关于数据的数据。它是在整个数据准备过程中逐步构建的，但创建列名称及描述是一个很好的起点。表 5.6 显示了 Beans 数据集的数据字典初稿。（有关已完成的数据字典的示例，请参阅表 5.19）。

表 5.6 Bean 数据字典

ClData	
名字	描述
ID	客户的唯一 ID
名字	客户的名称
行业名称	客户的行业
Employee	
名字	描述
ID	员工的 ID
姓名	员工的姓名
职位	员工的职务
费率	员工的小时费率
办公室	员工的办公室
婚姻情况	员工的婚姻状况
E - Dem	
名字	描述
名	员工的名字
姓	员工的姓氏
年龄	员工的年龄
证书列表	员工证书列表
大学	员工获得本科学位的大学

续表

Service	
名字	描述
ID	所提供服务的唯一 ID
日期	服务的发生日期
实际时间	在所提供的服务上花费的实际时间
预算时间	所提供服务的预算时间
ID	所执行任务的 ID
面积	任务区域
任务	执行任务的名称
员工 ID	执行任务的员工的 ID

以表 5.6 中的数据字典为起点，我们准备通过应用 20 种模式来准备数据。前两个是提取模式，用于确定是否已提取所有数据以及它们是否已正确传输。

数据准备模式 1：不完整的数据传输

提取将数据从源文件传输到 ETL 工具进行进一步处理。数据传输不完整和数据丢失会导致不可靠的结果。

比较行计数

比较行计数是检查完整性的一种方法。使用 Bean 数据集，通过在 ETL 工具中将 Excel 工作表的行计数与相应表的行计数进行比较来测试完整性。表 5.7 显示了“服务 Excel”工作表。选择 ID 列后，Excel 的状态栏显示工作表中有 4 632 行。

表 5.7 Excel 工作表的行计数

	A	B	C	D	E	F	G	H	I	J
1	ID	Date	ActualTime	BudgetedTime	ID	Area	Task	Employee	Client	
2	1	1/5/2023	3.75	3.5	2	Accounting	Review	18	42	
3	2	1/6/2023	2.65	2.5	1	Accounting	Prep	15	103	
4	3	1/6/2023	0.07	0.1	1	Accounting	Prep	12	178	
5	4	1/6/2023	0.35	0.3	1	Accounting	Prep	9	178	
6	5	1/6/2023	0.1	0.1	1	Accounting	Prep	2	2	
7	6	1/6/2023	0.35	0.4	1	Accounting	Prep	9	2	
8	7	1/6/2023	0.07	0.1	1	Accounting	Prep	12	2	
9	8	1/6/2023	0.25	0.3	1	Accounting	Prep	15	2	
10	9	1/6/2023	0.25	0.2	1	Accounting	Prep	17	2	

Workbook Statistics Caps Lock Average: 2316.3 Count: 4633 Numerical Count: 4633 Min: 1 Max: 4632 Sum: 10730028

Status Bar

Control Total: Average

Number of Rows

Control Total: Sum

接下来，在用于提取的 ETL 工具中确定服务表的行数。表 5.8 将该信息显示为 Power Query 中 ID 列配置文件的一部分。数字匹配表示所有行都已传输。如果数字不匹配，需要确定哪些数据未传输以及导致问题的原因。鉴于 ID 的顺序性质，差距分析为执行此操作的有用工具。在本案例中，数据从 Beans 数据集提取到 Power Query 时没有传输问题。

表 5.8 Power Query 的行计数

Column Statistics	...
Count	4 632
Error	0
Empty	0
Distinct	4 632
Unique	4 632
NaN	0
Zero	0
Min	1
Max	4 632
Average	2 316.5
Standard deviation	1 337.28...
Even	2 316
Odd	2 316

添加缺失行

如果行计数不匹配，可以将缺少的行添加到源数据、Beans 数据集文件中的“服务”工作表或 ETL 的数据集中。

模式 1 总结

问题	所有没有传输的数据。
检测（数据概要分析）	比较行计数。
纠正（ETL）	添加缺少的行。

数据准备模式 2：错误的数据传输

即使所有行都已传输，数据也有未得到正确传输的可能，这通常是由数据类型的差异引起的。

比较控制数量

可以使用控制数量来检测此问题。在 Bean 数据集中，将 Excel 工作表中的平均值（见表 5.7）与 ETL 工具中服务表的相同数字（见表 5.8）进行比较。匹配的数字表示正确传输“服务”表中 ID 列的值。可以对其他列运行类似的测试。

修改不正确的值

将数据从 Beans Excel 电子表格拉取到 Power Query 时没有不同的控制数量。但是，如果

数字不匹配，下一步是确定哪些数据传输不正确以及导致问题的原因。如果识别出这些问题，就可以修改 ETL 数据集中不正确传输的值。

模式 2

总结

问题	数据未正确传输。
检测（数据分析）	比较控制数量。
纠正（ETL）	修改不正确的值。

检查数据传输的完整性和正确性是审计人员的常见做法。客户通过电子邮件、云共享、USB 传输等方式提供数据。比较在发起方和接收方生成的数据统计信息有助于检测传输问题。



应用练习 5.3

使用模式提取数据

数据

Shanice 是 Stufan 公司的会计人员，她给你一个文本文件，其中包含 Stufan 的销售交易信息。她还提供了以下信息。

- 交易数量：28
 - 收取的平均价格：25.32
1. 将销售数据提取到 ETL 工具中。
 2. 确定是否传输了所有事务以及是否正确传输了所有数据。

答案

1. 将数据提取到 Excel 中。接下来，打开 Power Query 编辑器并检查“价格”列的配置文件。

Column Statistics	
Count	28
Error	0
Empty	0
Distinct	8
Unique	4
NaN	0

Column Statistics	
Zero	0
Min	10
Max	40
Average	25.3214...

2. 如可视化图所示，提供的信息表明所有交易都已转移—计数：28，并且已正确转移—平均值：25.32。

5.4 转换列的模式

学习目标 ④

将模式应用于转换列

当所有的数据都传输到 ETL 工具中以后，就可以转换它们了。转换有两个目的——通过修正值来清理数据，以及通过重组和集成数据进行分析。

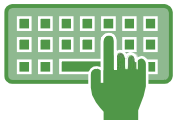
分析数据库是一个数据模型，它由一组集成的表组成，每个表有多个列。这意味着转换可以在列级、表级和模型级逐步进行。本节主要讨论列级的转换模式。这些模式会在单个列中查找数据问题，例如，名称不明确、数据类型问题以及不正确、不一致、不完整或无效的值。

数据准备模式 3：不相关和不可靠的数据

与决策无关的数据会使数据模型膨胀。避免将不可靠的数据集成到数据模型中也很重要。需要着重注意：从分析数据库中排除数据与删除数据是不同的。原始数据仍然存在，必要时可以进行集成。

浏览数据列查找不相关和不可靠的数据

如何做 5.1



不相关的列主要可以通过浏览数据可视化来识别。数据字典也可以是一个有用的工具。例如，员工表包含关于办公室的信息，如何使用这一列中的数据进行决策？

浏览数据还可以确定一列中是否包含不可靠的数据。在图 5.8 (A) 中，电子表中的年龄列混合了空值、数字、日期和文本，这使产生可靠的洞见变得困难。大多数 ETL 工具提供了关于错误、空值等的统计信息，可以帮助确定列的可靠性。

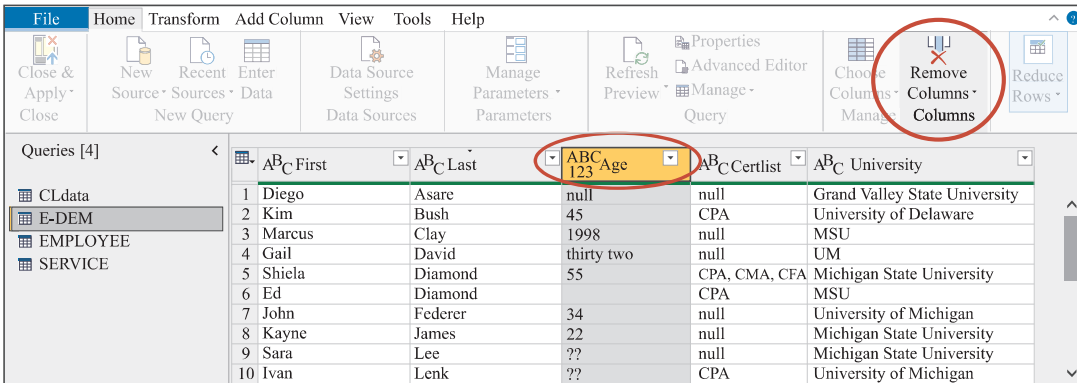
(**数据** 参见如何做 5.1 了解如何使用 Power Query 配置文件数据)。

删除含有不相关或不可靠数据的数据列

要纠正这个问题，需要从分析数据库中删除包含不相关和不可靠数据的列。图 5.8 (B) 展示了如何使用 Power Query 实现这一点。选择一个列，比如“年龄”，然后选择“删除列”。对于 Beans 数据集，删除了不相关的列“办公室”和不可靠的列“年龄”。

Age
45
1998
thirty two
55
34
22
??
??
51
9/2/1998

(A) 数据不可靠的列



(B) 删除数据不可靠的列

图 5.8 检测与纠正不可靠数据

模式 3

总结

问题	无关的数据使数据集膨胀。不可靠的数据增加了做出错误决定的风险。
检测（数据概要分析）	浏览可视化的各列，查找不相关和不可靠的数据。
纠正（ETL）	删除包含不相关或不可靠数据的列。

数据准备模式 4：不准确和模糊的列名

列名在数据探索 and 解释过程中成为变量。因为其他人可能会使用分析数据库，所以他们的名字很重要。从本质上，列名将成为数据库词汇表的一部分。回想一下，Beans 的管理合伙人想要进行必要的分析，但在数据准备方面却很困难。正确、直观和明确的列名使其他人

使用数据库和探索数据更容易。

浏览数据列查找不准确或不明确的列名

浏览可视化中列的内容和它的数据词典定义可以揭示列名是否准确地反映了它的内容。

以下是命名列的四条规则：

1. 应该准确地描述列的内容。
2. 对商业人士来说应该是直观的。
3. 使用大家都能理解的常见缩写，比如 YTD。
4. 删除空格、下划线或其他符号。例如，使用 CustomerName 而不是 Customer_Name。

重命名列

ETL 工具通过重命名列可以很容易地纠正这个问题。图 5.9 显示了如何在 Power Query 中重命名列。首先，右键单击 ClData 表中的 ID 列；其次，选择“重命名”，并输入新的名称：“ClientID”。请注意，目录名称更改会自动传播到所有公式和报告。

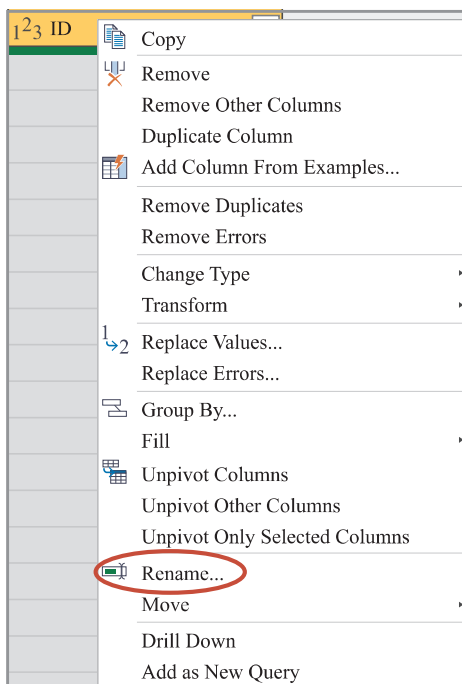


图 5.9 修改列名

在 Beans 数据集中，不是所有的列名称都是直观、容易理解的。表 5.9 显示了列的一些可能的名称更改。请注意，由于同一个表中的两列不能有相同的名称，Power Query 在加载数据时自动将 Service 表中的第二个 ID 列更改为 ID - 1。

表 5.9 对 Beans 的列名建议

Table	列名	修订后的列名
ClData	ID	ClientID
ClData	Industry Name	IndustryName
Employee	ID	EmployeeID
Employee	Job Title	JobTitle
Employee	MS	MaritalStatus
E - Dem	First	FirstName
E - Dem	Last	LastName
E - Dem	CertList	CertificationList
Service	ID	ServiceID
Service	Actual Time	ActualTime
Service	Budgeted Time	BudgetedTime
Service	ID - 1	TaskID
Service	Task	TaskName

模式 4 总结

问题	不准确或不明确的列名使数据集更难以理解和使用。
检测（数据概要分析）	浏览观察列中不正确或有歧义的名称。
纠正（ETL）	重命名列。

数据准备模式 5：不准确数据的类型

验证数据类型是数据准备的重要组成部分。数据类型是列定义的组成部分，因为它们决定了我们对列中的数据可以做什么，不可以做什么。例如，数学运算公式如求和及平均值需要数字字段，但是时间和日期函数需要一个具有日期数据类型的列。

检查数据类型

ETL 工具在提取过程中自动为每一列分配数据类型，但有时要么分配不正确，要么 ETL 工具无法确定数据类型。让我们用一个例子来说明最后的场景。图 5.10（A）显示了原始电子表格数据。面板（B）为 Power Query 中提取后的相同数据集。注意，数据类型是 ABC123，也称为任意数据类型。它表示 Power Query 无法识别数据类型，这意味着无法对列中的数据执行计算。

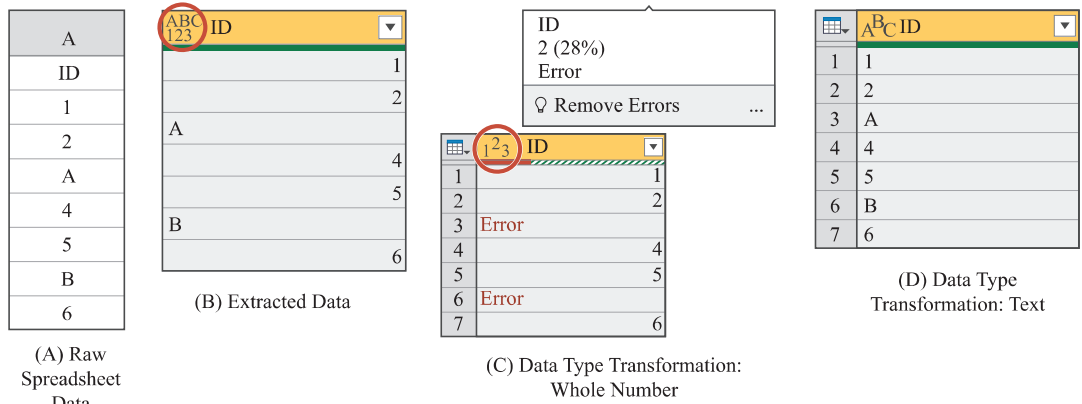


图 5.10 检查和更改数据类型

更改数据类型

通过使用 ETL 工具更改数据类型来纠正这个问题。图 5.11 显示了 Power Query 中可用的不同数据类型。图 5.10 (C) 显示了在数据设置面板 (B) 中选择的整数会呈现什么。有些值会被转换，而其他值会产生错误。这很有帮助，因为排查错误是数据准备工作的一部分。红色的条表示有错误，点击后会显示错误的百分比 (28%)。错误也可以被过滤掉，这对于大型数据集很有用。图 5.10 (D) 显示了当数据类型转换为文本时会发生什么。错误消失了，但是用数据能处理的事情就变得有限了。

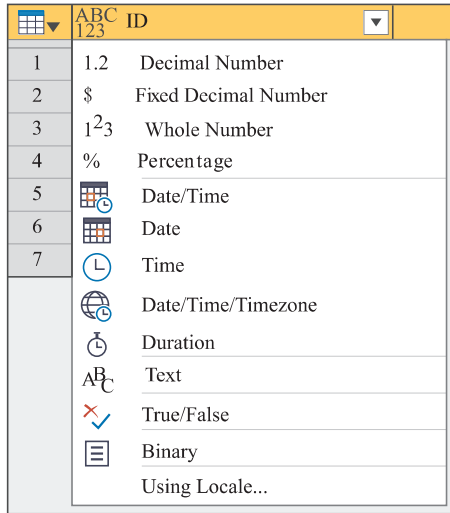


图 5.11 Power Query 数据类型

Beans 的数据类型没有问题，所以不需要进行修改。但是，需要向数据字典中添加指定数据类型的列。

模式 5

总结

问题	不准确的数据类型限制了对列中数据的处理。
检测（数据概要分析）	检查数据类型。
纠正（ETL）	改变数据类型。

数据准备模式 6：复合和多值列

每个单元格应该包含描述一个特征的一个值，因为同一个单元格中的两个或多个值会使分析难度增加。违反单值规则并使分析更加复杂的两种特定场景是复合列和多值列。

浏览查找复合和多值列

查找复合列或多值列的最佳方法是视觉扫描。用这种方式检查 Beans 的数据集可以发现以下问题：

- “员工”表中的“名称”列混合了员工的姓和名。虽然从分析的角度来看，这可能不是一个问题，但当将它们与 E - Dem 表中的名字和姓列进行匹配时会成为一个问题。
- E - Dem 表中的“证书列表”列是多值的。

重组数据

如何重组一个列取决于该列是复合的还是多值的。复合列的解决方案是将其拆分。在“Power Query”中，单击“名称”列。然后在主菜单中选择“主页”选项卡。单击功能区中的“拆分列”，然后选择按分隔符排列（见图 5.12）。

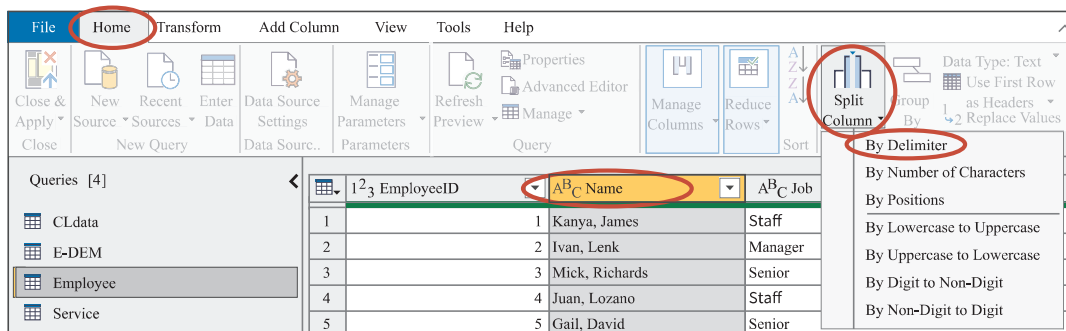


图 5.12 Power Query | 拆分列 | 设置

将出现如图 5.13 所示的窗口。Power Query 中选择了逗号作为分隔符，由于每个名称只有一个逗号，因此选择三个分割选项中的哪一个并不重要。在本例中，选择了最左边的分隔符选项。拆分后将 Name.1 重命名为“FirstName”，Name.2 重命名为“LastName”。

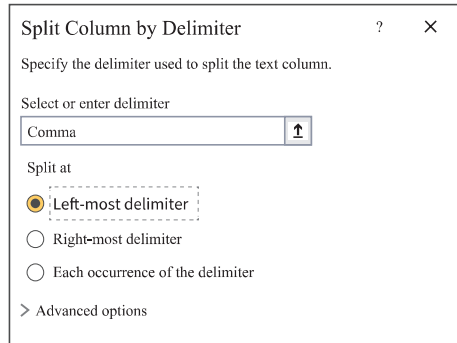


图 5.13 Power Query | 拆分列 | 选择分隔符

另外，“证书列表”列在 E - Dem 表中的多值特性需要创建一个新表。但是因为员工数据需要额外的工作，所以以后再做是有意义的。这个例子说明了应用模式的非顺序性，它们应用的顺序将取决于每个项目。Beans 数据集中的其他列是单值的，此模式不适用于它们。

模式 6

总结

问题	非单值列使分析变得困难。
检测（数据概要分析）	检查列中的值。
纠正（ETL）	拆分组合列。 为多值列创建一个单独的表。

数据准备模式 7：错误值

有时会将错误的值赋给实体的某个特征。员工的认证可以记录为 CMA 而不是 CPA，尽管两个都是有效的认证。不正确的值可能会产生严重的后果，包括运输和计费错误。

用离群值检测错误值

在单个列的内容中很难识别错误的值，因此寻找离群值，即在数值数据中突出的值很有帮助。如果一个价格为 19 美元的产品被记录为 91 美元，而如果大多数产品的价格在 10 ~ 30 美元，离群值就会凸显出来。离群值更正式的定义是，低于第一个四分位数或高于第三个四分位数的四分位数区间的 1.5 倍以上。

让我们分析服务表中的实际时间列。图 5.14 (A) 显示由 Power Query 生成的列配置文件统计信息。面板 (B) 按降序显示实际时间的值。列 [面板 (B)] 中的前三个值是离群值，很可能是不正确的。此外，ID 为“1325”的服务的实际时间为 25 小时，无效。应该是 2.5 小时。

Column Statistics		1.3 ID	1.2 ActualTime
Count	4 632	1325	25
Error	0	3277	22.5
Empty	0	4387	20
Distinct	190	2828	11.4
Unique	48	2452	10.75
NaN	0		
Zero	0		
Min	0.05		
Max	25		
Average	1.77672...		
Standard deviation	2.03039...		

(A) ActualTime Column: Statistics

(B) ActualTime Column:
Values in Descending Order

图 5.14 对错误数据的概要分析

修改错误值

一旦确定了一个有问题的值，就会有几个选项。首先，确定错误的根本原因并消除它。可能会警告用户正在将离群值输入系统。这将大大消除原始数据中的错误。其次，在源数据中更正值。最后，这个值可以在分析数据库中更正，但不能在源数据中更正。

表 5.10 显示 Beans 示例修改的实际时间值。

表 5.10 更正服务表中的实际时间字段

ID	IncorrectValue	CorrectValue
1325	25	2.5
3277	22.5	2.25
4387	20	2

模式 7 总结

问题	错误的数​​据可能会导致决策失误。
检测（数据分析）	使用离群值检测错误值。
纠正（ETL）	修改错误值。

数据准备模式 8：不一致的值

下一个模式处理数据不一致问题，当相同值的两个或多个不同表示在同一列中混合时，就会发生数据不一致。例如，当 MI 和密歇根都用作值时，以美元确定 MI 客户的总销售额可能会导致低估，这可能导致决策失误。

识别不一致的值

有两种概要分析技术可用于检测不一致的值：

- 不同的值：可视化浏览列中的不同值是识别不一致数据的有效方法。
- 频率：低频率的值可能表示数据不一致。

图 5.15 (A) 显示了员工表中工作职位列的不同值。此信息可用于“Power Query”中的所有列。图中显示高级经理的信息不一致。低频率也可能表明拼写错误导致了不一致。如图 5.15 (B) 所示，高级经理的频率为 1。图 5.15 (B) 中显示的值分布是 Power Query 列配置文件的一部分。Beans 数据集中另一个存在不一致问题的列是大学。

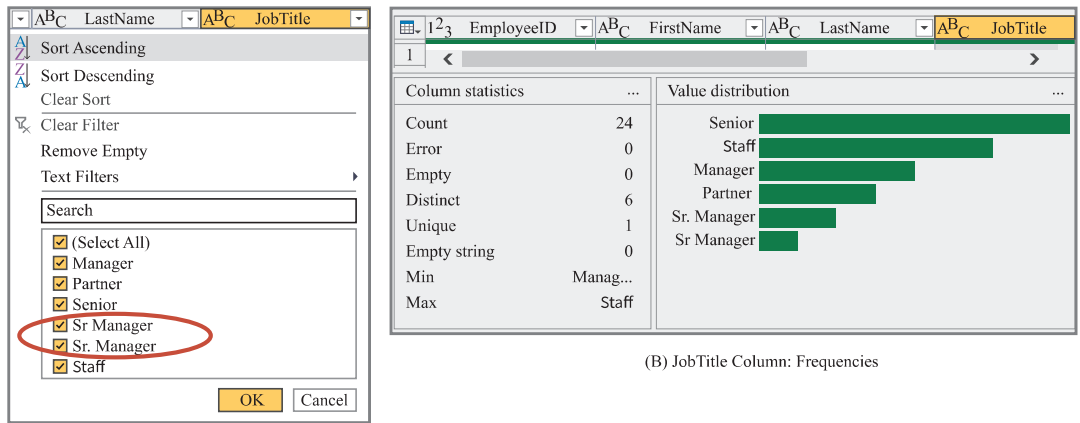


图 5.15 对不一致数据的概要分析

修改不一致的值

通过识别并消除根本原因或修改源数据、分析数据库中的值来纠正不一致的数据。修改值时，首先确定保留哪种表述。对于工作职位列，保留高级经理。表 5.11 中所示的变化是针对大学列进行的。

表 5.11 更改解决大学专栏中的不一致的问题

Current Value	New Value
MSU	Michigan State University
UM	University of Michigan

大多数 ETL 软件，包括 Power Query，都提供查找与替换工具来修改值。

模式 8

总结

问题	不一致的数据可能会导致决策失误。
检测（数据概要分析）	识别不一致的值。
纠正（ETL）	修改不一致的值。

数据准备模式 9：不完整的值

当前模式解决可能导致数据不可用和不可靠数据不完整的问题。例如，没有客户地址，企业就无法向他们发送营销材料。赋予空值的不同含义也可能导致数据不可靠。

不完整性有几个方面值得探讨：

- 是否应该允许空值，如果不允许，是否存在空值？
- 如果允许空值，值为空的百分比是多少？如果百分比很高，是否应该加载列？
- 如何表示不完整的值：空值？还是一个特定的代码？这些表示是一致的吗？

调查空值

ETL 工具逐列显示空值的百分比。使用 Power Query，表 5.12 显示了员工表中的“婚姻状态”列的信息。这些信息在许多方面都很有用：

- 对于每一个列，确定是否允许使用空值。主键不能包含空值。此外，员工表中的“费率”列不能包含任何空值，因为它是用于确定客户需要支付费用的信息。这个“费率”验证了收入情况是否符合。
- 高的或高于预期的百分比可能会使该列无法使用。

分析不完整性是如何“表达”的以及这如何影响数据的可靠性也很重要。“婚姻状态”列中的空值意味着什么？是另一种陈述员工为单身的方式？还是我们不知道该雇员的婚姻状况？

表 5.12

不完整性概要分析

ABC	MaritalStatus	
	Married	
	Single	
	Married	
	Single	
		null
	Married	
		null
		null
	Married	
	Single	
		null

MaritalStatus		
16 (67%)	0 (0%)	8 (33%)
Valid	Error	Empty
🔍 Remove Empty ...		

删除列或替换空值

修正方案视情况而定。如果不允许使用空值，但空值存在，则应该替换掉。如果空值的数量过多而无用，则从分析数据库中删除该列。如果在表示缺失值时存在不一致，则需要设

计一个一致的模式，并根据该模式修正这些值。对于婚姻状况列，值可以是已婚、单身和空白。空值表示该值是未知的。

模式 9

总结

问题	不完整性可能使数据无法使用和不可靠，可能会导致决策失误。
检测（数据概要分析）	研究空值。
纠正（ETL）	删除列或替换空值。

数据准备模式 10：无效值

为特定域创建规则，确定数据是否可以被大多数列接受。不符合这些预期的数据被认为是无效的。

创建和应用验证规则

对于某些验证规则，可以依赖由 ETL 工具自动生成的概要分析信息。对于强制性不能包含空值的列，可以使用 ETL 工具提供的关于空值的统计信息进行验证。但是，在大多数情况下，验证规则必须使用脚本语言来实现。表 5.13 显示了一些适用于 Beans 案例的验证规则的示例。因为没有使用特定的脚本语言，所以在示例代码列中使用的语法是通用的。

表 5.13

验证规则的设计与实施

Description	Sample Code
The actual hours for a service must be positive and can't exceed 14.	ACTUALHOURSVALID = IF SERVICE. ACTUALTIME > 0 AND SERVICEACTUALTIME <= 14, THEN "YES", ELSE "NO"
The minimum employee rate is \$150, and the maximum employee rate must be lower than \$500.	RATEVALID = IF EMPLOYEE. RATE >= 150 AND EMPLOYEE. RATE < 500, THEN "YES", ELSE "NO"
Valid job titles are: { Manager, Partner, Senior, Sr. Manager, Staff }.	JOBTITLEVALID = IF EMPLOYEE. JOBTITLE IN { "MANAGER", "PARTNER", "SENIOR", "SR. MANAGER", "STAFF" }, THEN "YES", ELSE "NO"

修改无效值

如果发现了有问题的值，消除根本原因，更改源数据库中的值，或更改分析数据库中的值。直到在 Beans 数据集中没有检测到无效的数据。

模式 10

总结

问题	无效的数据可能会导致决策失误。
检测（数据概要分析）	创建并应用验证规则。
纠正（ETL）	修改无效值。



应用练习 5.4

使用列转换模式

数据

一名助理准备了四份文件，分别是顾客、物品、销售和销售人员，然后交给你。这些文件有几个数据问题。请使用本节中的列转换模式识别它们。

你针对每个数据文件所发现的问题如下：

1. 描述问题。
2. 确定用于检测该问题的数据准备模式。
3. 解释如何纠正它。

答案

问题 1：

- (1) 客户表中的忠诚度评级列中的数据对分析没有用处。
- (2) 数据准备模式 3：不相关和不可靠的数据。
- (3) 从分析数据库中删除“忠诚度评级”列。

问题 2：

- (1) 销售表有一个标题不明确的列 - SP 的列。
- (2) 数据准备模式 4：列名不正确和不明确。
- (3) 用 SalesPerson 代替模糊的缩写 SP。

问题 3：

- (1) 客户表中的“州”列包含值 DN，该值可能不正确、不一致和（或）无效。
- (2) 数据准备模式 7、模式 8 和模式 10：错误、不一致和无效的值。
- (3) 将 DN 值替换为 DE。

5.5 转换表格的模式

学习目标 ⑤

将模式应用于转换表

表格层面的转换模式寻找单个表中的数据问题，如命名不明确的表、缺少主键和重叠的列。

数据准备模式 11：不直观的和模糊的表名

与列名一样，表名也是数据模型和数据集词汇的一部分，所以它们必须正确、直观、清晰。

浏览表格查找不准确或不明确的表名

检查一个表的内容和它的数据字典定义可以帮助确定名称是否准确反映了它的内容。命名表的规则与命名列的规则相同，应该是直观的，需要避免空格、下划线和特殊编码：

- 例如，使用 CashReceipt 而不是 Cash Receipt。
- 避免使用 DCustomer 这样的特殊编码（D 指的是维度表，但不是所有人都清楚）。

重命名表

由于名称的改变会自动传播到所有的公式，为了纠正这个问题，需要用 ETL 工具重新命名表格。以 Beans 为例，要做两个改动：首先，将 ClData 改为 Client；其次，将 E - Dem 改为 EmployeeDemographics。

模式 11

总结

问题	不准确或不明确的表名使人们更难理解和处理数据集。
检测（数据概要分析）	视觉上浏览表中不正确或不明确的名称。
纠正（ETL）	重新命名表。

数据准备模式 12：缺少主键

模式重点聚焦于主键。表是对于实体的描述，实体的每个实例都应该被唯一地识别。作为一个主键，一个列必须对每个实例有一个唯一的值，并且没有空值。从关系数据库中提取数据时，主键通常已经就位。然而，当从电子表格中提取数据时，主键并没有就位。比如在 Beans 案例中，为了建立一个主键，必须选择字段，而且以上两条规则都必须满足。

识别缺少主键的表格

每张表都需要至少一个列或列的组合，满足前面讨论的两个标准。符合这两个标准的列是候选键，ETL 工具可以帮助识别它们。Power Query 提供的列简介，先在表 5.8 中显示，并在表 5.14 中重复，提供了识别候选键的必要信息，即：

- 空值应该是零。
- Count、Distinct 和 Unique 的值应该相同。唯一值是指在列中只出现一次的值。

表 5.14 Power Query 中的列配置文件

列统计	
计数	4 632
错误	0
空	0
去掉重复的行	4 632
独立值	4 632
非数	0
零	0
最小值	1
最大值	4 632
平均值	2 316.5
标准差	1 337.28
偶数	2 316
奇数	2 316

创建主键

“雇员人口统计数据”表没有候选键。我们可以把名字和姓氏组合到同一个字段 (Name) 中，但是名字一般不应该被用作主键，因为它们很少是唯一的。对于这样的情况，可以创建一个人工密钥，例如数字。如图 5.16 所示，ETL 工具可以帮助我们做到这一点：

- 在“Power Query”中，点击“主菜单”中的“添加列”。
- 选择索引列。
- 点击“从 1 开始”。

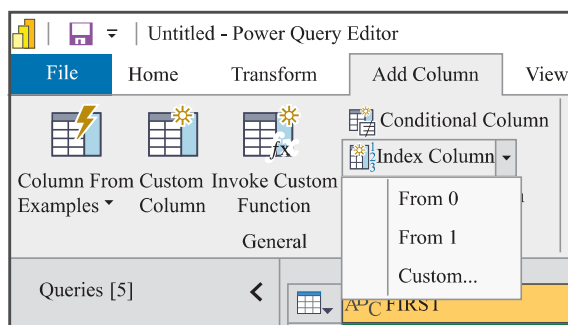


图 5.16 创建主键

模式 12

总结

问题	有些表没有主键。
检测（数据分析）	识别缺少主键的表。
纠正（ETL）	创建一个主键。

数据准备模式 13：跨列的冗余内容

这种模式寻找的是造成不一致的冗余。当相同的数据被记录了多次，并且在同一个地方发生了更改，但在另一个地方没有更改时，就会发生数据不一致，例如客户的电子邮件地址。这里有两种情况，一个表中的两个或多个列可能有相同的内容：

- 当存在重叠时，例如一个包含省份信息的地址和一个单独的省份字段。
- 当一个列的值依赖同一张表中另一个列的值时，就存在依赖关系。假设年龄和出生日期都有记录。但是随着时间的推移，年龄的值会发生变化，数据会变得不一致。因此应将其作为分析数据库的一部分进行计算，而不是数据源中提取年龄信息。

进行逐列比较

对重叠或依赖项进行逐列比较会发现这个问题。在 Beans 数据集中执行这类操作将显示当前没有冗余内容。

删除冗余的和从属列

可以删除包含冗余信息的列。当存在依赖项时，删除包含依赖项值的列。并使用公式在分析数据库中重新创建该列。

模式 13

总结

问题	表中各列的冗余内容可能导致不一致。
检测（数据概览分析）	进行逐列比较。
纠正（ETL）	删除冗余的和依赖的数据。

数据准备模式 14：用表内规则查找无效值

模式 14 与模式 10 类似，它也定义了一个列的可接受值。不同的是，模式 14 根据同一表中的一个或多个其他列的值来确定一列值的有效性。

创建并应用表内验证规则

验证规则的目标是识别无效的数据。创建验证规则需要对业务有深入的了解，而且要用脚本语言来实现。图 5.17 显示了一个 Beans 案例的表内验证规则。“费率”列中的值取决于“JobPosition”列中的值。将此验证规则应用于 Beans 的数据集，结果显示，对于五名员工（Alex Messi、Thibaut Martens、Paulo Lukaku、Ed Diamond 和 Molly McCarthy）来说，这个比率太低了。

描述

Beans 的员工“费率”是由他们的职位决定的。下表显示了每个职位的最低和最高工资，以美元为单位。

职位	≥	<
工作人员	150	200
高级岗位	200	250
经理	250	300
资深经理	300	350
合伙人	350	500

示例代码

```
RATEVALIDBASEDONJOBTITLE =
IF EMPLOYEE.JOBTITLE = "Staff" AND (EMPLOYEE.RATE >= 150 AND EMPLOYEE.RATE < 200), THEN "YES", ELSE,
IF EMPLOYEE.JOBTITLE = "Senior" AND (EMPLOYEE.RATE >= 200 AND EMPLOYEE.RATE < 250), "YES", ELSE,
IF EMPLOYEE.JOBTITLE = "Manager" AND (EMPLOYEE.RATE >= 250 AND EMPLOYEE.RATE < 300), "YES", ELSE,
IF EMPLOYEE.JOBTITLE = "Sr. Manager" AND (EMPLOYEE.RATE >= EMPLOYEE.RATE < 350), "YES", ELSE,
IF EMPLOYEE.JOBTITLE = "Partner" AND (EMPLOYEE.RATE >= 350 AND EMPLOYEE.RATE < 500), "YES", ELSE,
"NO"
)
```

图 5.17 表内验证规则的设计和实现

修改无效值

在 Beans 公司，修改员工费率或费率政策需要得到首席执行官的批准，因此在获得批准之前不可能有费率变化。

模式 14

摘要

问题	无效的数据可能导致决策失误。
检测（数据概要分析）	创建并应用表内验证规则。
纠正（ETL）	修改无效值。



应用练习 5.5

使用模式对表进行转换

在 Stufan，首席执行官 Shanice 想知道是否有可能整合销售并对采购信息进行分析。采购信息记录在一个名为 PTS 的 Excel 工作表中，样本显示如下。

	A	B	C	D	E
1	Invoice	Item	Price	Quantity	Vendor
2	1	DOCL	20	10	DITV
3	1	DOCS	15	10	DITV
4	2	TINL	10	10	BIC
5	2	SNOWL	7	10	BIC
6	3	DUCH	20	10	STWS
7	4	GRUML	20	10	DELT
8	4	SNOWL	25	10	DELT
9	5	TINL	9	10	STIMP

使用表格转换模式来识别 PTS 文件中的两个数据问题：

1. 描述数据问题。
2. 确定可以应用于检测的数据探索模式。

答案

问题 1：

- (1) 工作表的当前名称不明确。PTS 代表的 Purchase Transactions，但从标题上看并不明确。
- (2) 使用数据探索模式 11：非直观和模糊的表格名称。
- (3) 将工作表重命名为“purchasing”。

问题 2：

- (1) 该表目前没有主键，而且没有包含唯一值的列。表中的每一行代表一个发票行。表中的前两行代表发票 1 的两个行项，一个用于项目 DOCL；另一个用于项目 DOCS。
- (2) 使用数据探索模式 12：缺少主键。
- (3) 创建一个由唯一的顺序数字组成的人工密钥。

5.6 转换模型的模式

学习目标 ⑥

将模式应用于转换模型

模型层面的转换模式跨表搜索数据问题：如描述同一实体的数据分布在多个表中、结构难以理解的数据模型以及不支持高效处理的数据模型。

数据准备模式 15：数据跨表分布

描述同一实体的数据分布在多个表中会加大分析难度。表 5.15 显示了两种情况。

1. 在面板 (A) 中，两个表“一月销售额”和“二月销售额”具有相同的结构但行数不同。这两个表格也是面板 (C) 中自第四行起水平分割显示的销售表。所有数据都显示在同一个表中，回答销售额相关问题会更容易。

2. 在面板 (B) 中，两个表描述了同一实体——产品的不同特征。ID 为 1 的产品的部分信息在“产品描述”表中。相同产品 ID = 1 的其他信息位于“产品账簿”表中。在本例中，这两个表格也是面板 (D) 中所示的垂直拆分的产品表。

表 5.15

(A) Monthly Sales Data Across Tables			合并表 (B) Product Information Across Tables				
JanuarySales			ProductDescriptions		ProductAccounting		
Number	Date	Amount	ID	Type	Category	ID	QOH
189	1/10/2025	\$ 17 450	1	Stuffed Animals	Toys	1	84
190	1/15/2025	\$ 23 890	2	Radio	Electronics	2	133
191	1/24/2025	\$ 19 001	3	Soccer Balls	Toys	3	354
FebruarySales							
Number	Date	Amount					
192	2/9/2025	\$ 25 451					
193	2/10/2025	\$ 34 881					
194	2/14/2025	\$ 7 282					
195	2/23/2025	\$ 13 209					
(C) Monthly Sales Data Union			(D) Product Information Merge				
Sales			Product				
Number	Date	Amount	ID	Type	Category	QOH	
189	1/10/2025	\$ 17 450	1	Stuffed Animals	Toys	84	
190	1/15/2025	\$ 23 890	2	Radio	Electronic	133	
191	1/24/2025	\$ 19 001	3	Soccer Balls	Toys	354	
192	2/9/2025	\$ 25 451					
193	2/10/2025	\$ 34 881					
194	2/14/2025	\$ 7 282					
195	2/23/2025	\$ 13 209					

识别相似结构表格/描述同一实体的不同特征

要识别结构相似的表格，需要查找两个或多个具有相同结构的表格。这些表格具有相同的列和相似的数据。还有一种选择是搜索描述同一实体的不同特征的表格。在 Beans 案例中，“员工”和“员工人口统计”表就是这样的表格，任务目标是创建一个包含所有员工信息的表。

合并表格

回想一下本章内容，将两个具有相似结构的表组合在一起称为联合。在表 5.16 中，面板 (C) 中的表就是将面板 (A) 中的两个表组合在一起得出的。为同一实体组合具有不同特征的两个表是合并或连接。面板 (D) 中的表格合并了面板 (B) 中的两个表格。对于 Beans 案例，“员工”和“员工人口统计”表都包含员工信息，因此应将它们合并到一个表格中。（数据 参见如何做 5.2 了解如何使用 Power Query 合并这些表）。



表 5.16 显示了名为“员工”的合并表的结构。

表 5.16

合并员工表

员工
<ul style="list-style-type: none"> 员工编号 名字 姓氏 职位 费率 证书列表 婚姻状况 毕业院校

模式 15

总结

问题	一个实体的数据分布在多个表中，使分析复杂化。
检测（数据概要分析）	识别具有相似结构或描述同一实体的不同特征的表格。
纠正（ETL）	联合或合并表。

数据准备模式 16：数据模型不符合维度建模原则

维度建模是创建数据模型的一种技术，通过维度表对事实数据表进行校验。这些数据模型（如星型架构）易于理解，并能有效地处理数据。

分析数据模型对维度建模原则的遵从性

通过确定事实表和维度表并确保所有字段属于正确的表格来应用这些维度建模原则。在会计情境中，交易表对应于业务交易。另外，维度表描述了谁参与交易、何时发生的交易以及放弃或获得了什么。（关于信息建模的章节更详细地讨论了谁、什么以及何时分析会计交易。）

图 5.18 当前分析数据库构建为星型架构。服务表是事实数据表，员工表是主体维度，客户表也是主体（客体）维度。

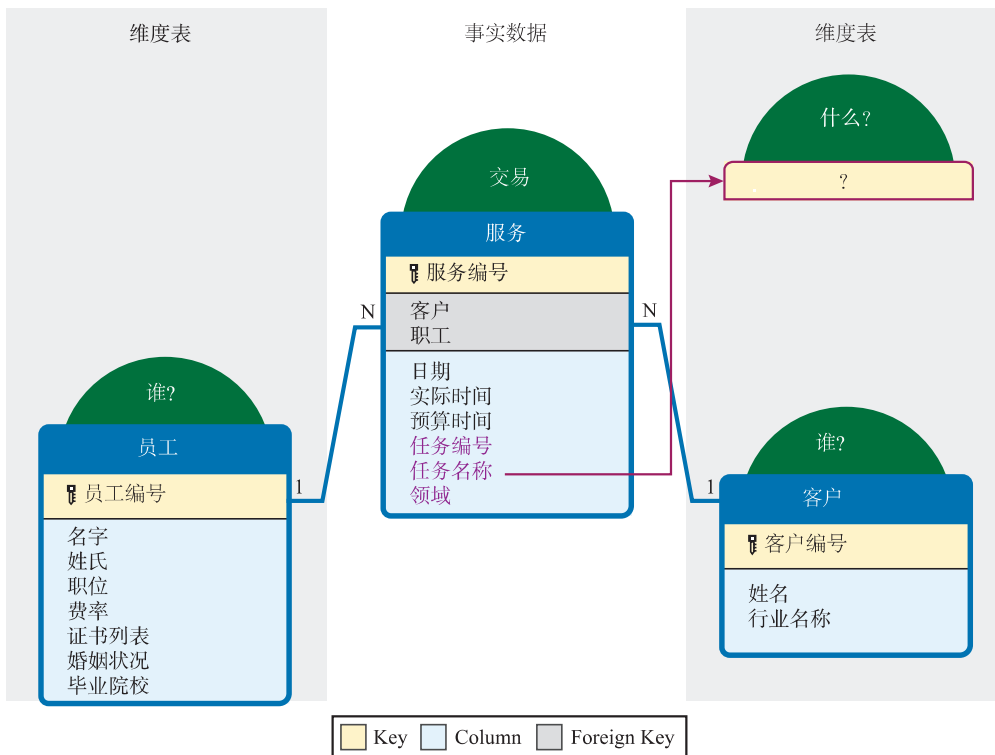


图 5.18 当前 BEAN 的星型架构

可以注意到缺少了时间和内容维度。服务表中只有数据字段，我们选择不指定单独的时间维度。任务编码、任务名称和领域都是描述的内容——Beans 向客户销售什么？因此，可以创建名为任务的新维度表。此外，回顾一下模式 6，员工表中的资格证书列是一个多值列，创建新表会将其转换为单值列。图 5.19 显示了我们要创建的数据模型。

为资格证书列创建一个单独的表，会产生雪花架构。在雪花架构中，维度的信息分布在多个表中。可以使用 Power Query 作为 ETL 工具创建图 5.19 中的架构。

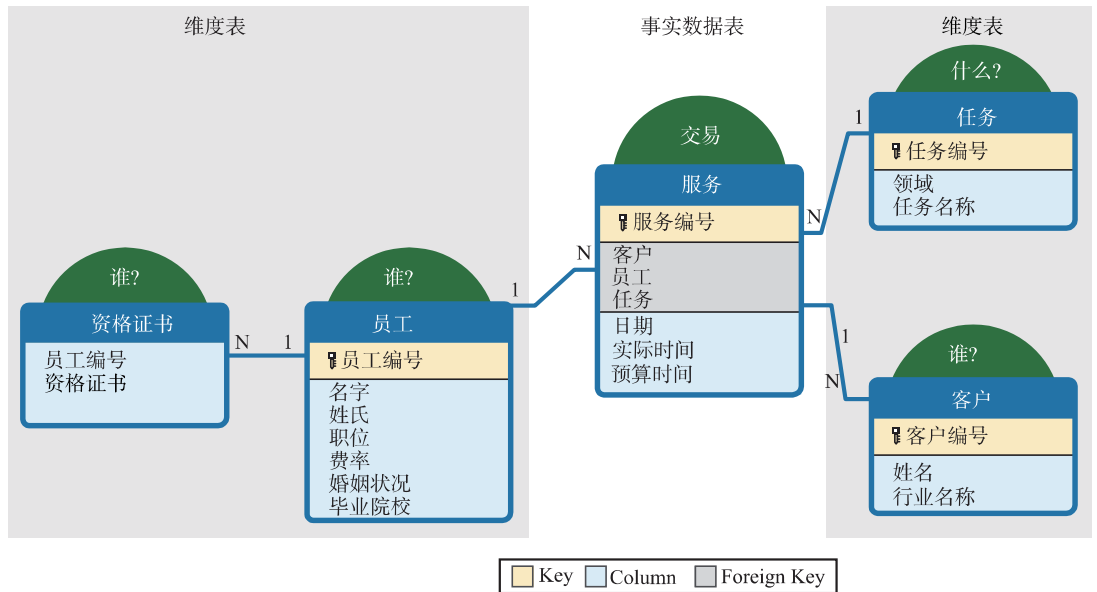


图 5.19 理想的星型/雪花型架构

将数据模型重新配置为星型/雪花型架构

首先，创建一个获取服务描述的任务维度表。

步骤 1：复制服务表并将复制的表重命名为“任务”（见图 5.20）。在服务表和任务表中只保留图 5.19 中显示的列。

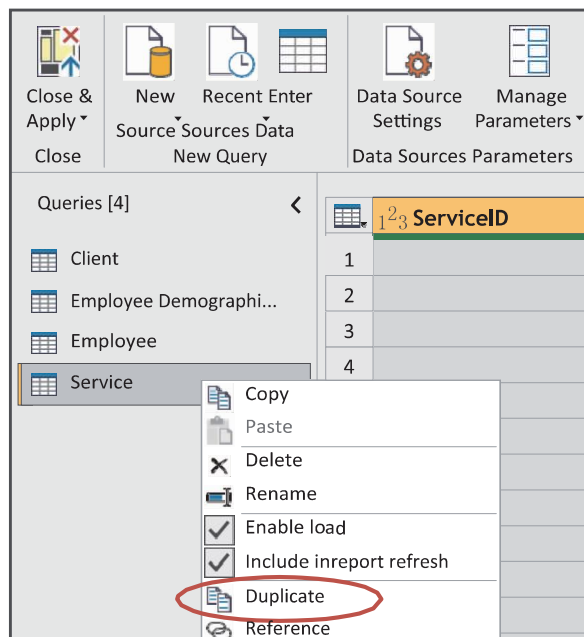


图 5.20 重复的服务表

确保任务编号字段保留在服务表中，并重命名为“任务”。服务表现在已拆分为两个表。“服务”是事实数据表，“任务”是维度表。

步骤2：通过选择 Power Query 主菜单中的 Home 选项卡，删除任务表中的重复行。选择“删除行”，然后选择“删除重复项”。现在，任务表应与表 5.17 所示的一致。

表 5.17 任务维度表

TaskID	Area	TaskName
1	Accounting	Prep
2	Accounting	Review
3	Accounting	Admin
4	Accounting	Other
5	Tax	Prep
6	Tax	Review
7	Tax	Admin
8	Tax	Other

其次，解决 CertificateList 的问题，CertificateList 是 Employee 表中的一个多值列：

步骤 1：使用与之前相同的过程创建 CertificateList 表：

- 复制 Employees 表并将重复项命名为“Certificates”。
- 从 EmployeeTable 中删除 CertificateList 列。
- 在新的 Certificates 表中，只保留 EmployeeID 和 CertificateList 列。

步骤 2：将 CertificateList 从多值列转换为单值列：

- 在 Power Query 中，拆分 CertificateList 列并选择“每次出现”选项。
- 单击 EmployeeID 列，然后在主菜单中选择 Transform。
- 选择功能区中的“取消透视列 (Unpivot Columns)”，然后选择“取消透视其他列 (Unpivot Other Columns)” (见图 5.21)。

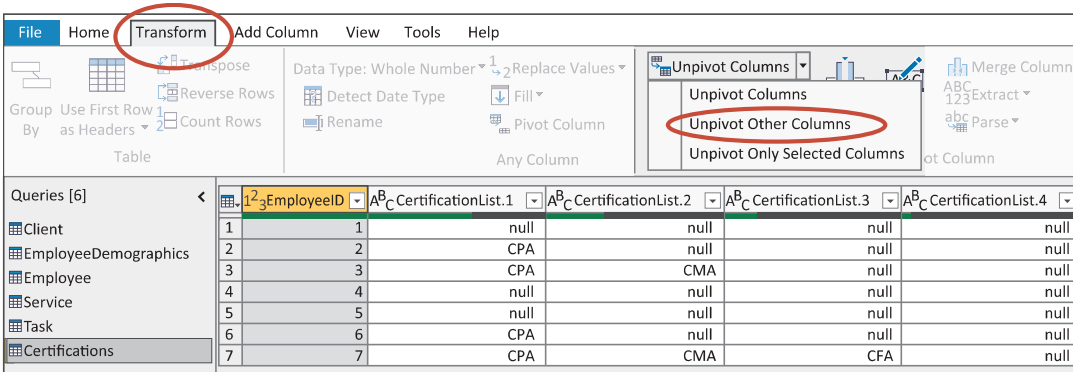


图 5.21 取消透视列

步骤 3：在生成的表中，删除“属性”列并将“值”列的名称更改为“证书”。员工的证书现在记录在一个单值列中。

这完成了 Bean 数据集到星型/雪花型架构的转换。

模式 16

总结

问题	不遵从维度建模原则的数据模型通常更难分析。
理想的星型/雪花型架构	分析数据模型是否符合维度建模原则。
纠正 (ETL)	将数据模型重新配置为星型/雪花型架构。

数据探索模式 17：使用表间规则查找无效值



如何做 5.3

模式 10、模式 14 和模式 17 类似，因为它们定义了一个列的可接受值。模式 17 基于一个或多个其他表格中的值来确定列值的有效性。广泛使用表间验证规则的例子是参照完整性，它指的是外键中的所有值也应作为相应主键中的值而存在。（如何做 5.3 解释了如何使用 Microsoft Access 实现引用完整性）。

创建和应用表间验证规则

表间验证规则识别无效数据。规则的创建需要对业务的深入了解。表 5.18 显示了应用于 Beans 案例的表间验证规则。它是服务表的一部分，它指定只有经理、高级经理和合作伙伴才能审核业务。该规则检测到，3971 号服务由一名高级员工审核，4193 号服务由员工审核。

表 5.18

表间验证规则的设计与实现

题面描述	示例代码
只有经理、高级经理和合作伙伴才能审核业务	<pre>NOAUTHORITY = IF TASK.TASKNAME = "REVIEW" AND EMPLOYEE [JOBTITLE] IN { "MANAGER", "SR. MANAGER", "PARTNER" }, THEN "OK", ELSE "ISSUE"</pre>

修改无效值

如果你正在处理 Beans 案例，可以咨询首席执行官，确定是否因违反规则、数据输入错误或其他原因导致错误。如果确定了错误的存在，可以通过设置控制避免进一步违反规则、更正源数据中的值或更正分析数据库中的值来纠正问题。

模式 17

总结

问题	无效数据可能导致决策失误。
检测 (数据分析)	创建并应用表间验证规则。
修正 (ETL)	修改无效规则。



应用练习 5.6

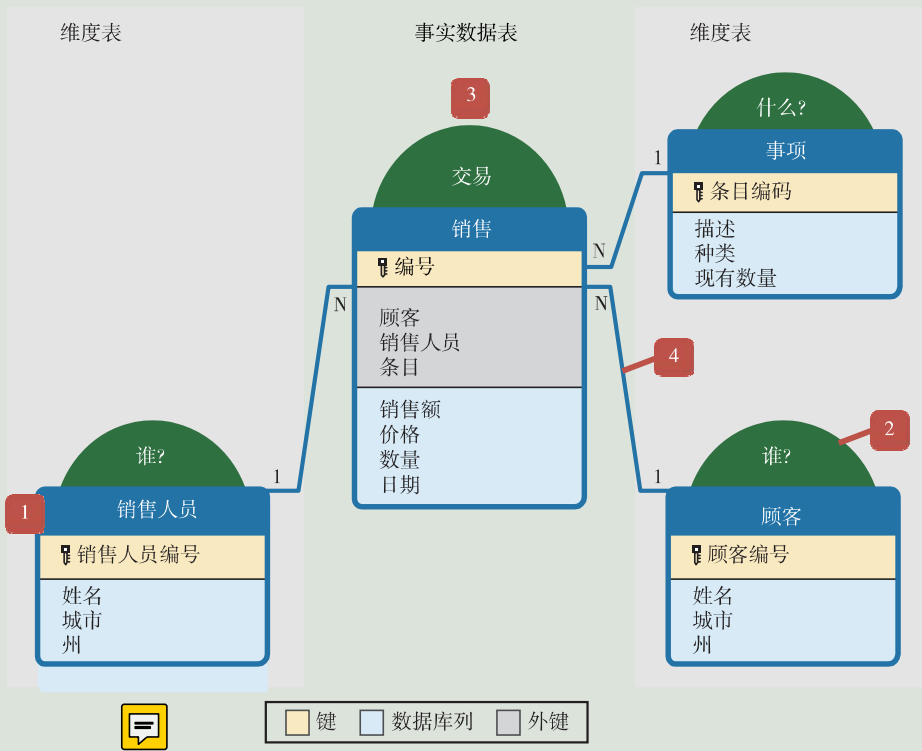
绘制星型架构

数据

Stufan 公司的助理为你提供客户、项目、销售和销售人员的修订版本。使用这些文件中的数据绘制星型架构：

1. 绘制不同的表及其字段。
2. 将每个表标记为事实或维度，并将每个维度表标记为谁、何时或什么。
3. 把事实表放在中间。
4. 通过连接表并定义它们的基数来完成星型架构。

答案



5.7 哪些模式适用于数据加载？

学习目标 7

将模式应用于数据加载问题

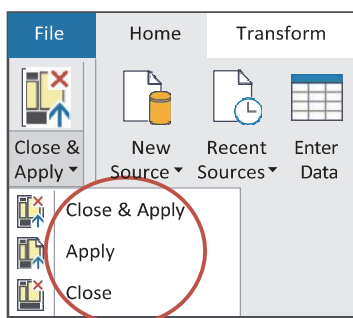
数据被清理和转换完毕，就可以加载到软件中进行分析了。数据加载是使分析数据库可供使用的过程。由于提取和加载是传输过程，因此在传输数据的完整性和正确性方面存在相似问题。验证分析数据库的数据模型也很重要，即所有关系都已确定。

数据准备模式 18：不完整的数据加载

加载将数据从 ETL 工具移动到分析数据库。表 5.19 显示了如何将数据从 ETL 工具（在本例中为 Power Query）传输到分析数据库。

表 5.19

ETL 工具传输到分析数据库



有三个选项：

- **(close & apply)** 关闭并应用：关闭 Power Query 并将所有转换应用于分析数据库。
- **(apply)** 应用：将所有转换应用于分析数据库，但保持 Power Query 打开。
- **(close)** 关闭：关闭 Power Query，不对分析数据库应用任何转换。

如果选择了前两个选项中的一个，确认所有数据都已经传输了是非常重要的。

比较行计数

与模式 1 类似，分析数据库的行计数可以与 ETL 工具中数据集的行计数进行比较。如果转换应用分析数据库过程中发生错误，ETL 工具还将发出警报。

添加缺失的行

如果数字不匹配，需要确定哪些行未传输以及出现的原因。找到后，将缺少的行添加到分析数据库中。

模式 18

总结

问题	在加载过程中未传输所有数据。
检测（数据概要分析）	比较行计数。
修正（ETL）	添加缺失数据。

数据准备模式 19：数据加载不准确

即使所有行都传输完整，也可能存在没有传输正确的数据。模式 19 将解决这个问题。

比较控制量

与模式 2 类似，验证数据正确传输的有效方法是比较总和、平均值或任何其他控数量。

修改不准确的值

如果数字不匹配，确定哪些数据传输错误以及问题的原因。找出后修改分析数据库中错误传输的值。

模式 19

总结

问题	加载过程中未传输正确的数据。
检测（数据概要分析）	比较控制数量。
修正（ETL）	修改不正确的值。

数据准备模式 20：丢失或不准确的数据关系

分析在很大程度上依赖基础数据模型。错误定义的数据关系可能会导致分析失败或不能实现，因此必须在加载后验证数据模型的完整性和准确性。

调查数据模型的完整性和精确性

一个完整精确的数据模型应该是所有关系都正确的模型。图 5.22 显示了使用 Power Query 创建的 Beans 分析数据库的最终数据模型。与你的数据模型进行比较，确定没有丢失任何关系，没有非必要的关系，并且所有关系都定义正确。

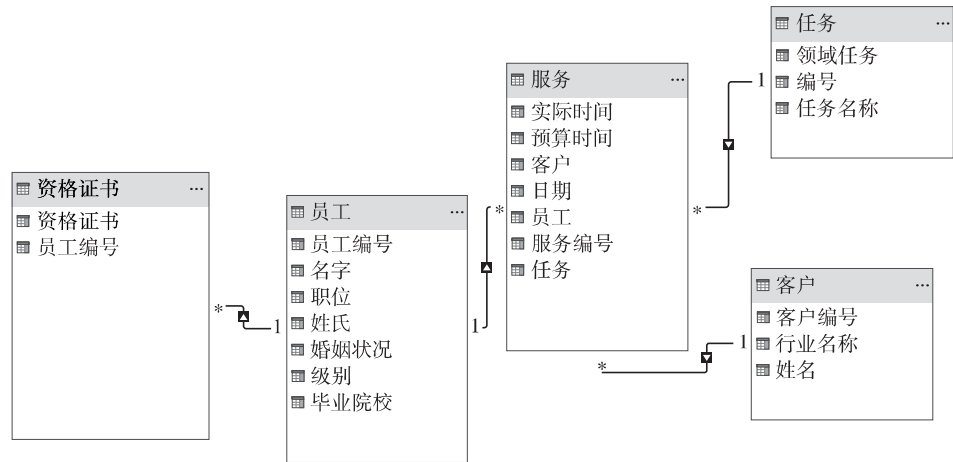


图 5.22 Beans 数据模型

修改数据模型

要在 Power BI 中执行此项操作，请选择主菜单中的“主页”选项卡，然后单击功能区中的“管理关系”。将出现图 5.23 所示的窗口。选择窗口底部的按钮以创建、编辑或删除关系。

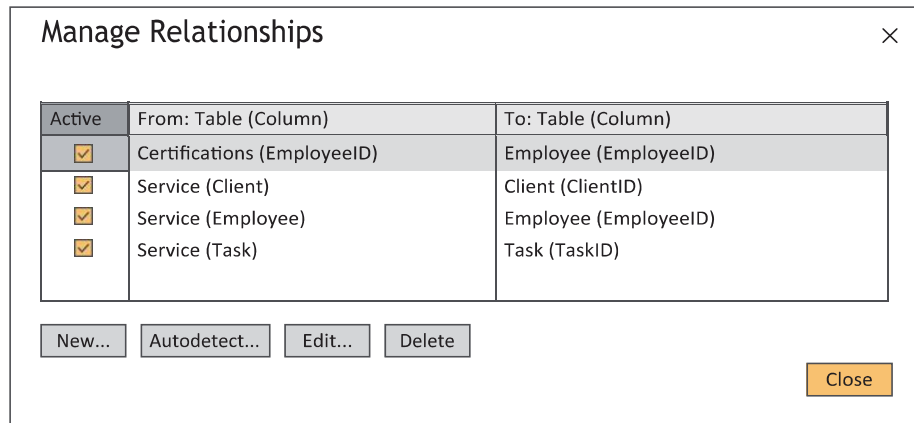


图 5.23 管理关系

图 5.24 列示了可以定义的关系的一些方面，包括以下三项：

- 指定了关系的数据字段。
- 应用于关系的基数。
- 用于聚合数据的导航方向。

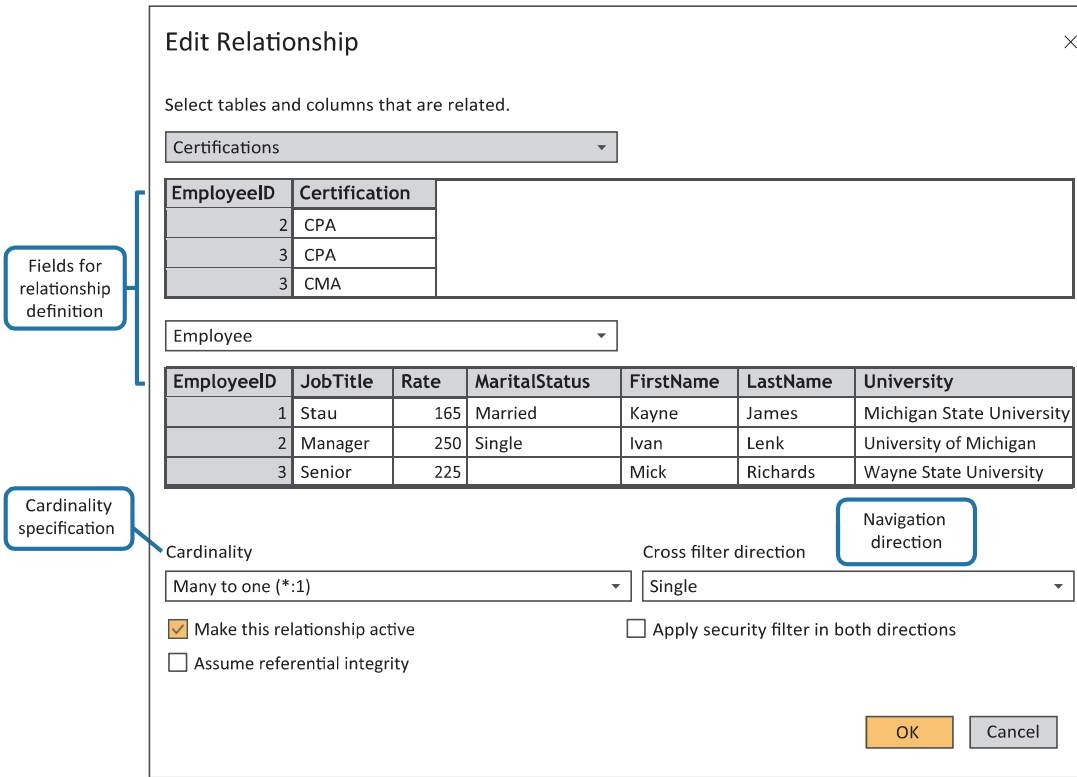


图 5.24 定义关系

模式 20

总结

问题	缺失或错误定义的数据关系会使数据关系具有迷惑性，甚至可能关系不存在。
检测（数据概要分析）	研究数据模型的完整性和准确性。
修正（ETL）	修改数据模型。

Beans 数据集的转换及其分析准备工作现已完成。

表 5.20 显示了图 5.22 中分析数据库的修订数据字典。

表 5.20

Beans 修订数据字典

服务				
姓名	描述	数据类型	键	命令
客户	客户的唯一编号	整数	外键	是
员工	员工的唯一编号	整数	外键	是
任务	任务的唯一编号	整数	外键	是
任务编号	服务的唯一编号	整数	主键	是
日期	服务发生的日期	日期		是

续表

服务				
姓名	描述	数据类型	键	命令
实际时间	所提供服务的实际时间	小数		是
预算时间	所提供服务的预算时间	小数		是
任务				
名称	描述	数据类型	键	命令
任务编号	任务的唯一编号	整数	主键	是
任务名称	任务名称	文本		是
领域	任务领域	文本		是
客户				
姓名	描述	数据类型	键	命令
客户编号	客户的唯一编号	整数	主键	是
姓名	客户姓名	文本		是
行业名称	客户行业	文本		是
员工				
姓名	描述	数据类型	键	命令
员工编号	员工的唯一编号	整数	主键	是
名字	员工的名字	文本		是
姓氏	员工的姓氏	文本		是
职位	员工的职位	文本		是
级别	为员工收取的小时费率	整数		是
婚姻状况	员工的婚姻状况	文本		否
毕业院校	员工获得本科学位的大学	文本		否
资格认证				
姓名	描述	数据类型	键	命令
员工编号	员工的唯一编号	整数	外键	是
资格证书	员工所获资格证	文本		是



应用练习 5.7

评估表之间的关系

以下两幅插图显示了 Stufan 数据集的 Sale 和 Customer 表之间的关系。有两个问题会严重扭曲与这种关系相关联的分析。识别并描述为什么它们会成为问题。

Edit Relationship

Select tables and columns that are related.

Customer

CustomerCode	Name	City	State
1	Cruella De Vil	Phoenix	AZ
2	Scar LeRoi	Orlando	FL
3	Winnie Pooh	Wilmington	DE

Sale

ID	Sale	Item	Price	Quantity	Customer	SalesPerson	Date
1	1	DOCL	30	10	1	111223333	Friday, February 21, 2025
2	1	GRUML	30	10	1	111223333	Friday, February 21, 2025
3	2	SNES	20	20	6	null	Saturday, February 22, 2025

Cardinality: One to one (1:1) Cross filter direction: Both

Make this relationship active
 Assume referential integrity

OK Cancel

答案

1. 在错误的字段之间定义了关系：CustomerCode 和 ID。ID 是分配给 Sale 表中每一行的序列号。结果，错误的客户被分配给了销售人员。应该在 Customer 表中的 Customer Code 字段和 Sales 表中的 Customer 字段之间指定关系。

2. 1 - N 基数模式被预期为一种维度表（客户）和事实数据表（销售）之间的一种关系。客户可以进行多个销售，但每次销售只能指定一个客户。

章节回顾与练习

学习目标回顾

① 解释数据概要分析的过程

数据分析是调查数据质量和结构的过程。它有三个部分：

- 调查数据质量：确定数据中是否存在异常，即数据是否有问题。
- 调查数据结构：改进数据组织，以便进行分析。
- 决定和通知：决定是否解决已确定的问题、这样做的成本以及不解决问题的后果。

② 描述提取—转换—加载（ETL）过程

提取、转换、加载（ETL）是一个纠正问题的过程：

- 提取是将需要转换的数据移动到临时区域。
- 转换涉及在三个子过程中改进数据：清理、重组和集成。
- 加载是将数据移动到即将用于分析的区域。

③ 将模式应用于提取数据

数据准备模式是数据准备项目的强大工具。它们有助于识别数据问题，并为检测和纠正这些问题提供指导。数据提取模式处理：

- 数据的不完整传输。
- 数据传输不准确。

④ 将模式应用于转换列

转换通过修正值（质量）、重组和整合数据结构以进行分析（结构）来清理数据。它可以在列、表和模型级别逐步进行。列转换模式始终在一个列中发布数据。

面向结构的列模式强调：

- 包含不可靠数据的相关列。
- 列名不正确且不明确。
- 数据类型不正确。
- 组合或多值列。

面向质量的列模式处理：

- 数据不正确的列。
- 列中的数据不一致。
- 数据不完整的列。
- 数据无效的列。

⑤ 将模式应用于转换表

表转换模式处理单个表中的数据问题。面向结构的表格模式处理：

- 不直观且不明确的表格名称。
- 无主键的表格。
- 具有两列或更多列且内容重叠的表格。

面向质量的表模式解决：

- 可由表内规则确定的无效值。

⑥ 将模式应用于转换模型

模型转换模式处理跨表间的数据问题。结构导向的模型模式处理：

- 将具有类似结构的表进行联合。合并同一实体的具有不同特征的表格。
- 符合维度建模原则。

面向质量的表格模式处理：

- 可由表格间规则确定的值无效。

⑦ 将模式应用于数据加载问题

提取并重新加载传输过程，并解决以下相同的数据问题：

- 数据传输不完整。
- 数据传输不准确。

此外，加载数据后，重要的是：

- 调查数据模型是否完整准确。

表 5.21 提供了 20 种数据准备模式的概览。第一列显示 ID 便于查阅，第二列中的代码指示模式是聚焦于值（V）还是重组数据（S）。表中还包括数据的问题如何检测及纠正。

表 5.21 数据准备模式概述

提取				
编号	代码	问题	检测（数据分析）	修正 t（ETL）
1	V	所有数据都没有传输	比较行计数	添加缺少的行
2	V	数据未准确传输	比较控制数 修改不正确的值	
转化				
编号	代码	问题	检测（数据分析）	修正（ETL）
3	S	不相关或不可靠的数据	可视化浏览列查找不相关和不可靠的数据	删除包含不相关和不可靠数据的列
4	S	不正确或不明确的数据	浏览列以查找不正确或不明确的名称	重命名列
5	S	数据类型不正确	检查数据类型	更改数据类型
6	S	列不是单值	检查列值	拆分组合列为多值列创建分隔符
7	V	值不正确	检测带有离群值的不准确值	修订不准确值
8	V	值不一致	确定不一致的值	修订不一致的值
9	V	值不完整	调查空值	删除列或替换空值
10	V	无效值	创建并应用验证规则	修订无效值

续表

转化				
编号	代码	问题	检测 (数据分析)	修正 (ETL)
表				
11	S	表名不正确或不明确	浏览表查找不正确或不明确的名称	重命名表
12	S	缺少主键	查找出缺少主键的表	创建主键
13	S	冗余列内容	执行逐列比较	删除冗余和从属数据
14	V	使用表内规则检测到无效值	创建并应用表内验证规则	修改无效数据
模型				
15	S	数据跨表分布	识别具有相似结构的表或描述同一实体不同特征的表	连接或合并表
16	S	符合维度建模原则	分析数据模型是否符合维度建模原则	重组模型为星型/雪花型架构
17	V	使用表间规则检测到无效值	创建并应用表间验证规则	修改无效规则
加载				
编号	代码	问题	检测 (数据分析)	修正 (ETL)
18	V	数据加载不完整	比较行计数	添加缺少的数据
19	V	数据加载不准确	比较控数	修改不正确的值
20	S	数据关系缺失或不正确	调查数据模型的完整性和准确性	创建、修改和删除关系修改数据模型

关键词回顾

分析数据库
基数
复合列
数据连接器
数据字典
数据提取
数据集成
数据加载
数据匹配

数据分析
数据重组
数据转换
数据仓库
维度
“脏数据”
提取转换加载 (ETL)
事实
平面表

离群值
参照完整性
单值列
切片
雪花型架构
星型架构
联合
验证规则

选择题

1. (LO 1) 以下哪个术语描述了在处理和之前对数据进行概要分析、清理、重组和整合的过程? ()。

A. 数据分析

B. 数据准备

C. 数据查询

D. 数据分析

2. (LO 1) 郝先生是一名应收账款分析师，正在审查客户用于分析项目的主文件。他注意到，在地址字段中，一个客户所在的州列为北美洲；另一个客户所属的州则列为新墨西哥州。以下哪项是正确的？（ ）。

- A. 发现了一个不正确的数据实例
- B. 发现了一个无效数据实例
- C. 发现了一个不一致的数据实例
- D. 发现了一个规模不大的数据实例

3. (LO 1) Spencer 是一家体育用商店的财务分析师，负责为公司所有体育经纪人编制主文件。利用表格信息确定以下哪项是正确的（ ）。

发言人	姓名	运动
巴拉特 - 阿伦, 保龄球	巴拉特	保龄球
拉马克里什南 - 斯里达尔, 田径运动	维拉特 - 科利	田径运动
维拉特 - 科利, 击球手	维拉特 - 科利	击球手
罗赫特 - 夏尔马, 击球手	罗赫特 - 夏尔马	击球手

- A. 第一列包含错误数据，而第二列是效数据
- B. 第一列是一个单值列，而第二列是一个复合列
- C. 第一列是一个复合列，而第二列是一个单值列
- D. 第一列是会计列，而第二列和第三列为切片列

4. (LO 1) 以下关于星形模式的陈述中哪一项是不正确的？（ ）。

- A. 事实数据表展示商业交易
- B. 与事实数据表相比，维度表通常具有更少的实例和更多的列
- C. 为了使切片数据更容易，大多数度量值都定义为维度表的一部分
- D. 维度表用于什么、何时和何人分析
- E. 粒度表示事实数据表的粒度级别

5. (LO 2) 以下关于数据转换的陈述中哪一项是正确的？（ ）。

- (1) 数据转换有三个子过程：清理、重组和集成
- (2) 数据清理旨在纠正数据异常
- (3) 数据重组不会改变任何数据，只是改变数据的组织方式
- (4) 数据集成通过定义关系将数据链接在一起

- A. (1) 和 (2)
- B. (1) 和 (3)
- C. (1)、(2)、(4)
- D. (1)、(2)、(3)、(4)

6. (LO 2) 数据清理可以包括添加、修改和删除数据。以下选项中需要修改数据的是？（ ）。

- A. 采购订单文件缺少一个月的采购订单
- B. 在准备分析超过 500 美元的采购订单数量时，待处理的文件中在每个新月份的开头

都有一行，日期列中列出了月份名称。该行中没有其他具有月份名称的数据


C. 在检查采购订单数据时，供应商的名称、街道地址、城市和州列在单独的列中。在某些情况下，州列显示州名，在其他情况下显示州缩写

D. 在分析采购订单时，有些订单包含供应商电话号码

7. (LO 3) 通过将源数据与 ETL 工具中的数据进行比较来验证数据传输。以下关于数据传输的选项中不正确的是? ()。

- A. 行数有助于验证数据传输的完整性
- B. 序列间隙分析有助于验证数据传输的完整性
- C. 比较数值字段的平均值有助于验证数据传输的准确性
- D. 比较列标题有助于验证完整性和准确性
- E. 比较数字字段的总和有助于验证数据传输的准确性

8. (LO 3) 您正在准备对供应商和采购订单的分析，并已从 ERP 系统传输过去一年中所有采购订单的数据和所有供应商的数据。您已获得有关数据的以下信息 ()。

(1)	采购订单总数	15 786	
(2)	所有采购订单的总金额	\$ 1 567 679	
(3)	平均采购订单金额	\$99. 31	
(4)	供应商总数	672	

哪些信息与确保所有数据传输相关? ()。

- A. (1)、(2)
- B. (2)、(3)
- C. (1)、(4)
- D. (1)、(2)、(3)、(4)

9. (LO 4) Lizelle 是一名金融分析师，正在准备分析数据库分析制造数据以便确定最有效的流程。她目前正在表中的列中识别不相关和不可靠的数据。当她发现不相关或不可靠的数据时，以下哪项是最合适的对策? ()。

- A. 删除包含不相关或不可靠数据的列
- B. 放弃项目，因为数据无效且无法分析
- C. 取消透视列，以便将数据分隔为不同的字段
- D. 将相关、不相关、可靠和不可靠的数据放在单独的列中

10. (LO 4) Hofflak 是美国西北部的一家大型建筑公司。它们的 ERP 系统跟踪所有资产。资产表的数据字典如下所示。

资产

姓名	描述
ID	资产的唯一编号
Description	资产的描述
AssetCategory	资产的类别

续表

姓名	描述
Price	为资产支付的价格
Salvam	资产使用寿命结束时的估计账面价值
EstimatedLifetime	资产的估计寿命
DprMethod	适用于资产的折旧方法

如果要进行数据分析，应更改哪些列名？（ ）。

- A. ID, Salvam, DprMethod
- B. ID, Price, EstimatedLifetime, DprMethod
- C. Price, Salvam, EstimatedLifetime, DprMethod
- D. AssetCategory, Price
- E. Salvam, EstimatedLifetime, DprMethod
- F. EstimatedLifetime, DprMethod

11. (LO 4) 如果要重构复合列，则应该（ ）。

- A. 删除不需要的列数据
- B. 将列数据拆分为单独的列
- C. 创建一个单独的表
- D. 将列移动到另一个表

12. (LO 4) Courtiera 是一家高端奢侈品公司，设计、制造手表和首饰并销往全球。下表提供了它们跟踪的客户信息的一小部分样本。公司有一个非常简单的客户忠诚度计划。如果客户购买了价值超过 1 000 0000 美元的产品，他们将获得“钻石”状态；如果他们购买的价格超过 1 000 000 美元，则获得“黄金”状态；而如果购买的价格低于 1 000 000 美元，将获得“无”状态。

顾客

编号	顾客忠诚度计划	国家
1	钻石	美国
2	黄金	意大利
3	N/A	英国
4	黄金	

根据提供给您的信息，此数据集存在什么样的问题？（ ）。

- A. 不正确、不一致、不完整、无效
- B. 不一致、不完整、无效
- C. 不一致，无效
- D. 不完整，无效

13. (LO 5) 关于在表中建立主键的陈述，以下哪个选项是正确的？（ ）。

- A. 主键必须在所有相关表中使用相同的列标题
- B. 主键的每一行必须具有唯一的值，并且没有空值

- C. 主键和外键是在同一表中重复的相同信息
 - D. 如果表没有主键，则不能将其包含在分析数据库中
14. (LO 5) 您正在检查正在分析的数据集中的列。数据集中包含以下列。

HotelPropertyID	包含每个酒店的唯一标识号
HotelAge	酒店开业多少年了
OpeningDate	酒店开业日期
RoomsAvailable	酒店的房间数量
RoomsRented	全年入住的房间数量
Revenue	从房间收取的金额

在回顾了专栏标题和描述之后，你会关注以下哪一项问题？（ ）。

- A. 列之间没有重叠
 - B. HotelAge 和 OpeningDate 之间存在重叠
 - C. RoomsAvailable 和 RoomsRented 之间存在冗余
 - D. HotelAge 和 OpeningDate 之间存在依赖关系
15. (LO 5) 考虑以下验证规则应用于所示员工表的场景：

VALID =

IF EMPLOYEE. AGE < 24 AND EMPLOYEE. DEGREE = "COLLEGE", THEN "YES",
ELSE "NO"

员工

编码	年龄	学历
1	23	高中
2	23	大学
3	25	大学
4	21	

以下哪个选项是不正确的？（ ）。

- A. 代码为“1”的员工的“有效”列的值为“否”
- B. 代码为“2”的员工有效列的值是“否”
- C. 代码为“3”的员工“有效”栏的值是“否”
- D. 代码为“4”的员工有效栏的值为“否”

16. (LO 6) 金融分析师 Mithali 正在准备数据用于数据分析项目，根据环境因素预测下一时期的收入。她目前正在检查表以验证表间引用完整性规则。以下哪项是关于引用完整性的问题？（ ）。

- A. 主键字段和外键字段必须具有相同的名称

- B. 主键必须是单值的，但外键可以是多值的
- C. 主键字段和外键字段可以具有不同的数据类型
- D. 外键中的所有值也应作为相应主键中的值存在

17. (LO 6) Skyline 是一家国有的管道和供热公司。Skyline 的 CFO 正在为分析准备员工绩效数据。下面的表格为其中四个文件的示例。你建议如何合并这四个文件? ()。

1 月购买记录

购买编号	日期	数量	员工
14399	1/1/2025	1 432.24	7
14400	1/1/2025	799.99	14
14401	1/1/2025	320	12
14402	1/1/2025	822.21	22

员工统计资料

编号	姓名	出生日期
14399	Ben Jespers	3/3/1975
14400	Bei Shi	7/17/1985
14401	Michael Rodman	5/20/1984
14402	Benita Alvarez	11/11/1979

员工工资

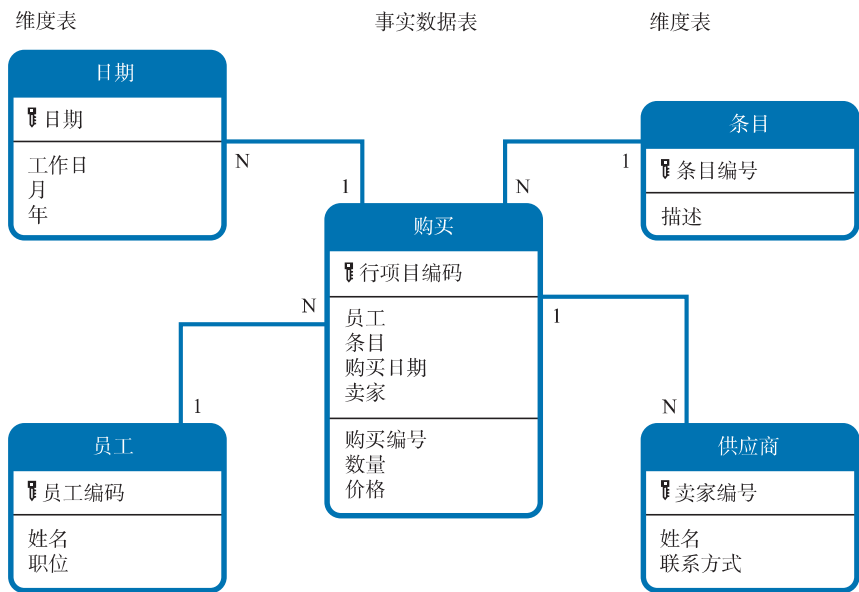
编号	小时工资
14399	25
14400	38
14401	36
14402	45

2 月购买记录

购买编号	日期	数量	员工
15021	2/1/2025	123.89	4
15022	2/1/2025	540.99	1
15023	2/1/2025	183.12	14
15024	2/1/2025	744.41	21

- A. 合并 1 月购买记录和员工工资，合并 2 月购买记录和员工工资
- B. 合并 1 月购买记录和 2 月购买记录以及员工统计资料和员工工资

- C. 合并 1 月购买记录和 2 月购买记录, 合并员工统计资料和员工工资
 D. 合并 1 月购买记录和员工工资, 合并 2 月购买记录和员工工资
 18 (LO6) 以下哪项将导致将星型架构转换为雪花型架构?()。
- A. Emp dyeeName 是一个组合字段, 包含员工的姓氏和姓氏
 B. 对于一些员工来说, 他们的职位是未知的
 C. 可以为一个供应商建立多个联系人
 D. 工作日、月和年被广泛用于分析目的



19. (LO7) Hamza 正在将数据传输到分析数据库中进行分析。他应该采取哪些步骤来确定是否传输了所有数据? ()。

- A. 查看列以获得相应的命名规则
 B. 建立数据验证规则并确定哪些数据不符合数据验证规则
 C. 检查不同的值并分析频率
 D. 将分析数据库的行计数与 ETL 工具中数据集的行计数进行比较

20. (LO7) 你正在为一家名为 Dine At Home 的食品配送公司准备分析项目。“在家吃饭” (Dine At Home) 从餐厅领取食物订单并送到顾客手中。客户可以多次订购, 但订单只针对一个客户。餐厅可能有很多订单, 但一份订单只针对一家餐厅。以下哪个基数规格不正确? ()。



本章更多习题可通过扫描本书封底二维码注册获得；相关答案和专业应用案例解答，教育工作者可通过向中图数字科技（北京）有限公司电子邮箱（待添加）发送邮件申请获得。

普通高等教育经管类专业系列教材

彭飞 白默 主编

微 课 版

会计信息化实训教程

—— 财务业务一体化（用友U8 V10.1）

第2版

教学资源

U8 V10.1安装程序

实验账套

微课视频

PPT课件



清华大学出版社

内 容 简 介

本书以简明原理、突出实训为主导思想,以一个企业单位的经济业务为原型,重点介绍了用友 U8 财务管理与供应链管理系统核心子系统的主要功能和业务处理方法,兼顾原理阐释与实务训练。书中为读者定制了几十个实训项目并提供了实训准备账套和结果账套,每个实验既环环相扣,又可以独立运作,适应不同层次教学的需要。

本书共分为 10 个项目,分别介绍了企业建账、基础设置以及用友 U8 软件中最重要和最基础的总账、固定资产、薪资管理、采购管理、应付款管理、销售管理、应收款管理、库存管理、存货核算、UFO 报表共 10 个模块的基本功能,并以系列实训的形式详解了 10 个模块的主要业务的处理。

本书可以用作高等院校会计学及经济管理类其他专业的教材,也可用作全国大学生会计信息化技能大赛的辅导教材,还可用作在职会计人员的用友 U8 自学读本。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

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

会计信息化实训教程:财务业务一体化:用友 U8 V10.1

微课版/彭飞,白默主编.—2版.—北京:清华大学出版社,2025.8.

(普通高等教育经管类专业系列教材).—ISBN 978-7-302-69589-9

I . F232

中国国家版本馆 CIP 数据核字第 20254NZ911 号

责任编辑:刘金喜

封面设计:常雪影

版式设计:思创景点

责任校对:成凤进

责任印制:宋 林

出版发行:清华大学出版社

网 址: <https://www.tup.com.cn>, <https://www.wqxuetang.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:三河市龙大印装有限公司

经 销:全国新华书店

开 本:185mm×260mm 印 张:21.5 字 数:550千字

版 次:2021年2月第1版 2025年9月第2版 印 次:2025年9月第1次印刷

定 价:68.00元

产品编号:105700-01

第 二 版 前 言

在当今这个信息化飞速发展的时代，财务人员若不能熟练掌握和运用信息化管理工具，其职业道路将面临严峻挑战。随着信息技术的不断进步，传统的会计工作方式已经无法满足现代企业的需求。因此，“会计信息化”不仅成为了会计专业的核心课程之一，更是财务人员必须掌握的关键技能。

会计信息化涉及会计数据的收集、处理、存储和传输等多个环节，是企业财务管理现代化的重要标志。本书的编写初衷，是帮助会计专业的学生和在职财务人员通过对会计信息化的系统学习，深入理解会计信息化的基本原理，掌握业务财务一体化核心子系统的业务操作流程，提升专业技能和职业竞争力。书中不仅涵盖了会计信息化的理论基础，还包括了实际操作技巧和案例分析，使读者能够将理论知识与实践相结合，更好地适应信息化时代对财务工作的要求。

考虑到随着云计算、大数据、人工智能等新兴技术的发展，会计信息化也在不断地更新和升级，因此有必要对第一版教材进行修订。本次修订主要体现在以下四个方面：

- (1) 根据最新修订的企业会计准则，对相关业务的操作步骤做了部分修订和改写。
- (2) 对涉及新修订的税收法律相关内容进行了调整与修订。
- (3) 结合同行教学反馈，增加了部分业务描述与操作步骤截图。

第2版教材由天津商业大学会计学院彭飞、白默担任主编。由于水平所限，书中难免存在疏漏和不当之处，期待使用本书的教师和同学不吝指正，以便今后不断完善教材内容。

服务邮箱：476371891@qq.com

编 者
2025年1月

项目一 企业建账 1

- 实训一 增加用户 1
 - 实训任务 1
 - 任务解析 1
 - 实训指引 2
- 实训二 建立账套 4
 - 实训任务 4
 - 任务解析 4
 - 实训指引 5
- 实训三 为用户设置权限 9
 - 实训任务 9
 - 任务解析 9
 - 实训指引 10
- 实训四 账套输出/引入 16
 - 实训任务 16
 - 任务解析 16
 - 实训指引 16
 - 拓展任务 17

项目二 基础设置 18

- 实训一 机构人员设置 18
 - 实训任务 18
 - 任务解析 19
 - 实训指引 19
- 实训二 设置客商信息 22
 - 实训任务 22
 - 任务解析 23
 - 实训指引 24
- 实训三 设置存货信息 25
 - 实训任务 25
 - 任务解析 26
 - 实训指引 27
- 实训四 设置财务信息 30
 - 实训任务 30
 - 任务解析 34
 - 实训指引 35

实训五 设置收付结算 41

- 实训任务 41
- 任务解析 42
- 实训指引 43

实训六 系统服务 45

- 实训任务 45
- 任务解析 45
- 实训指引 46
- 拓展任务 46

项目三 总账管理 49**实训一 总账管理系统初始化设置 49**

- 实训任务 49
- 任务解析 51
- 实训指引 51

实训二 凭证管理 55

- 实训任务 55
- 任务解析 56
- 实训指引 56

实训三 出纳管理 74

- 实训任务 74
- 任务解析 76
- 实训指引 76

实训四 期末结转 79

- 实训任务 79
- 任务解析 80
- 实训指引 80

实训五 账务管理系统对账与结账 86

- 实训任务 86
- 任务解析 86
- 实训指引 86
- 拓展任务 88

项目四 固定资产 92**实训一 固定资产系统初始化设置 92**

- 实训任务 92

任务解析·····	94	实训指引·····	94	实训指引·····	136
实训二 固定资产增加业务·····	100	实训五 缴纳社保业务·····	139	实训任务·····	139
实训任务·····	100	任务解析·····	140	实训指引·····	140
任务解析·····	101	实训六 薪资管理系统结账·····	142	实训任务·····	142
实训指引·····	101	任务解析·····	142	实训指引·····	142
实训三 固定资产变动业务·····	108	项目六 供应链管理系统初始化	144	实训一 业务信息设置·····	144
实训任务·····	108	实训任务·····	144	任务解析·····	146
任务解析·····	108	实训指引·····	146	实训二 单据设置·····	149
实训指引·····	108	实训二 单据设置·····	149	实训任务·····	149
实训四 计提当月折旧·····	111	任务解析·····	150	实训指引·····	150
实训任务·····	111	实训三 设置业务子系统凭证科目·····	152	实训三 设置业务子系统凭证科目·····	152
任务解析·····	111	实训任务·····	152	任务解析·····	154
实训指引·····	112	实训指引·····	154	实训四 录入业务子系统期初数据·····	156
实训五 固定资产减少业务·····	113	实训四 录入业务子系统期初数据·····	156	实训任务·····	156
实训任务·····	113	任务解析·····	158	任务解析·····	158
任务解析·····	114	实训指引·····	159	实训指引·····	159
实训指引·····	114	实训五 设置业务子系统参数·····	163	实训五 设置业务子系统参数·····	163
实训六 固定资产管理系统对账		实训任务·····	163	任务解析·····	164
与结账·····	116	实训指引·····	164	实训指引·····	164
实训任务·····	116	实训四 录入业务子系统期初数据·····	156	拓展任务·····	168
任务解析·····	116	实训任务·····	156	项目七 采购与应付款管理	170
实训指引·····	116	任务解析·····	158	实训一 普通采购业务·····	170
拓展任务·····	117	实训指引·····	159	实训任务·····	170
项目五 薪资管理	119	实训一 薪资管理系统初始化设置·····	119	任务解析·····	172
实训一 薪资管理系统初始化设置·····	119	实训任务·····	119	实训指引·····	172
实训任务·····	119	任务解析·····	121	实训二 采购现结业务·····	182
任务解析·····	121	实训指引·····	122	实训任务·····	182
实训二 工资数据变动及计算·····	130	实训二 工资数据变动及计算·····	130	任务解析·····	184
实训任务·····	130	实训任务·····	130		
任务解析·····	131	任务解析·····	131		
实训指引·····	131	实训指引·····	131		
实训三 工资分摊设置·····	133	实训三 工资分摊设置·····	133		
实训任务·····	133	实训任务·····	133		
任务解析·····	134	任务解析·····	134		
实训指引·····	134	实训指引·····	134		
实训四 工资分摊相关账务处理·····	135	实训四 工资分摊相关账务处理·····	135		
实训任务·····	135	实训任务·····	135		
任务解析·····	136	任务解析·····	136		

实训指引·····	184	实训任务·····	252
■ 实训三 采购退货业务·····	188	任务解析·····	253
实训任务·····	188	实训指引·····	254
任务解析·····	189	■ 实训六 委托代销业务·····	258
实训指引·····	190	实训任务·····	258
■ 实训四 含现金折扣的采购业务·····	193	任务解析·····	260
实训任务·····	193	实训指引·····	261
任务解析·····	195	■ 实训七 受托代销业务·····	269
实训指引·····	196	实训任务·····	269
■ 实训五 代垫运费二次分摊·····	201	任务解析·····	272
实训任务·····	201	实训指引·····	273
任务解析·····	202	■ 实训八 代垫运费业务·····	279
实训指引·····	202	实训任务·····	279
■ 实训六 采购业务手工结算·····	206	任务解析·····	281
实训任务·····	206	实训指引·····	282
任务解析·····	208	■ 实训九 直运销售业务·····	286
实训指引·····	208	实训任务·····	286
■ 实训七 采购暂估业务·····	213	任务解析·····	288
实训任务·····	213	实训指引·····	289
任务解析·····	214	■ 实训十 票据贴现业务·····	293
实训指引·····	215	实训任务·····	293
拓展任务·····	220	任务解析·····	294
		实训指引·····	294
		■ 实训十一 坏账处理·····	295
		实训任务·····	295
		任务解析·····	296
		实训指引·····	296
项目八 销售与应收款管理 ·····	222		
■ 实训一 普通销售业务·····	222	项目九 库存管理与存货核算 ·····	300
实训任务·····	222	■ 实训一 库存盘点业务·····	300
任务解析·····	224	实训任务·····	300
实训指引·····	225	任务解析·····	300
■ 实训二 销售现结业务·····	232	实训指引·····	301
实训任务·····	232	■ 实训二 计提存货跌价准备·····	304
任务解析·····	234	实训任务·····	304
实训指引·····	234	任务解析·····	305
■ 实训三 销售退货业务·····	237	实训指引·····	305
实训任务·····	237	■ 实训三 期末结账·····	306
任务解析·····	238	实训任务·····	306
实训指引·····	238	任务解析·····	306
■ 实训四 销售折让业务·····	242	实训指引·····	307
实训任务·····	242		
任务解析·····	245		
实训指引·····	245		
■ 实训五 分期收款业务·····	252		

项目十 UFO报表管理系统 313

实训一 利用模板生成“资产

负债表” 313

实训任务 313

任务解析 313

实训指引 314

实训二 利用模板生成“利润表” 319

实训任务 319

任务解析 319

实训指引 319

实训三 利用模板生成“现金

流量表” 321

实训任务 321

任务解析 321

实训指引 321

实训四 自定义“财务指标

分析表” 324

实训任务 324

任务解析 325

实训指引 325

项目九

库存管理与存货核算

实训一 库存盘点业务



实训任务

2025年6月30日，进行期末库存盘点，结果如图9-1所示。

库存商品实存账存对比表											
盘点日期：2025年6月30日											
盘点单位：仓管部各仓库											
单位：元											
商品名称	单 价	单价	账面结存		实际盘存		升溢		损耗		升溢损耗原因
			数量	金额	数量	金额	数量	金额	数量	金额	
床笠 (秋天)	件	80.00	490	39 200.00	510	40 800.00	20	1 600.00			收发计量错误
被芯 (泰国产)	件	120.00	300	36 000.00	290	34 800.00			10	1 200.00	系6月8日山西春天家居用品制造有限公司少发，对方同意补货

单位主管：(略) 会计：(略) 复合：(略) 监盘：(略) 物资负责人：(略)

图 9-1 库存商品实存账存对比表

任务解析

1. 背景知识

为了保证企业库存资产的安全和完整，做到账实相符，企业必须对存货进行定期或不定期的清查，查明存货盘盈、盘亏、损毁的数量及造成的原因，并据以编制存货盘点报告表，按规定程序，报有关部门审批。

库存管理系统提供了盘点单，用来定期对仓库中的存货进行盘点。存货盘点报告表，是证明企业存货盘盈、盘亏和毁损并据以调整存货实存数的书面凭证，经企业领导批准后，即可作为原始凭证入账。对于存货的盘盈、盘亏，应及时办理存货的账务手续，按盘盈、盘亏存货的计划成本或估计成本，调整存货账面数，记入“待处理财产损溢”科目。

(1) 存货的盘盈

借：存货(如库存商品、原材料等)

贷：待处理财产损溢/待处理流动资产损溢

(2) 存货的盘亏

企业对于盘亏的存货，根据存货盘点报告表所列金额，作如下处理。

借：待处理财产损溢/待处理流动资产损溢

贷：存货(如库存商品、原材料等)

盘点前应将所有已办理实物出入库但未录入计算机的出入库单，或销售发货单、销售发票均录入计算机中。同一时刻不能有两张相同仓库、相同存货的盘点表未记账。库存盘点业务处理流程如图9-2所示。

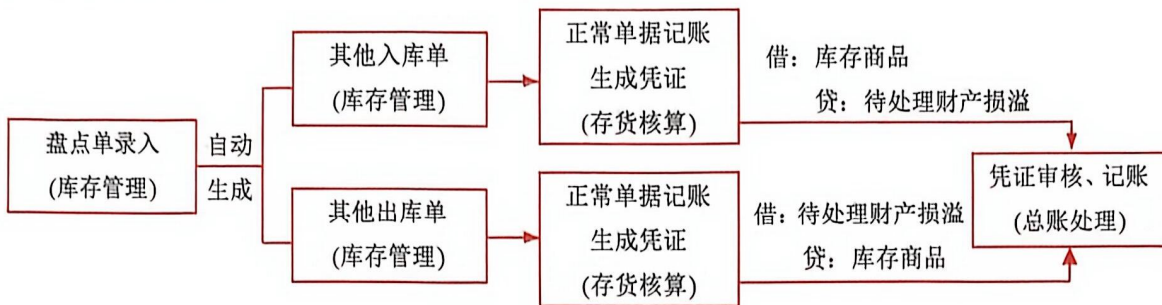


图 9-2 盘点业务处理流程图

2. 岗位说明

以业务员C01身份完成库存管理系统业务处理；以会计W02身份完成存货核算系统、总账管理系统业务处理。

实训指引

以系统管理员身份在系统管理中引入“8-1销售与应收”账套作为基础数据。

(1) 在库存管理系统中录入盘点单

① 以仓管人员“C01 陈晨”身份登录，修改“注册日期”为“2025年6月30日”。

② 在库存管理系统中，执行“盘点业务”命令，单击“增加”按钮，选择表头项目“盘点仓库”为“床上用品仓库”，“出库类别”选择“盘亏出库”，“入库类别”选择“盘盈入库”，“部门”为“仓管部”。

③ 单击“盘库”按钮，弹出“盘库将删除未保存的所有记录，是否继续？”信息提示框，单击“是”按钮，弹出“盘点处理”窗口，默认选择“盘点方式”为“按仓库盘点”。

④ 单击“确定”按钮，在“盘点单”表体中根据实验资料图9-1所示录入“单价”“盘点数量”等信息，“账面金额”“盘点金额”信息自动生成。

⑤ 依次单击“保存”“审核”按钮，如图9-3所示。盘点单审核后，会自动生成相应的其他出、入库单。

特别提醒：

◇ 盘点表中的账面数，为增加盘点表中的存货的那一时刻该仓库、该存货的现存量，它是库存管理系统中该仓库、该存货的账面结存数减去销售管理系统中已开具发货单或发票但未生成出库单的数量之差。

(2) 在库存管理系统中审核其他出、入库单

① 在库存管理系统中，执行“入库业务”|“其他入库单”命令，单击“末张”按钮，查看

盘盈生成的其他入库单，单击“审核”按钮，如图9-4所示。

盘点单

表体排序
 普通仓库盘点 合并显示
 倒冲仓库盘点

盘点会计期间 _____ 盘点单号 0000000001 盘点日期 2025-06-30
 盘点日期 2025-06-30 盘点仓库 床上用品仓库 出库类别 盘亏出库
 入库类别 盘盈入库 部门 仓管部 经手人 _____
 备注 _____

	存货编码	存货名称	规格型号	主计量单位	账面数量	单价	账面金额	调整入库数量	调整出库数量	账面调整数量	盘点数量	盘点金额	溢
1	0101	床垫(秋天)	1.8M	件	100.00			0.00	0.00	100.00	100.00		
2	0102	空调被(夏季)	1.8M	件	50.00			0.00	0.00	50.00	50.00		
3	0103	床笠(秋天)	1.8M	件	490.00	80.00	39200.00	0.00	0.00	490.00	510.00	40800.00	
4	0104	蕾丝被(夏季)	1.8M	件	100.00			0.00	0.00	100.00	100.00		
5	0105	被芯(泰国产)	1.8M	件	300.00	120.00	36000.00	0.00	0.00	300.00	290.00	34800.00	
6	0106	印花床上四件套...	1.8M	套	600.00			0.00	0.00	600.00	600.00		
7	0107	儿童床上四件套...	1.5M	套	546.00			0.00	0.00	546.00	546.00		
8													
	合计				2186.00		75200.00	0.00	0.00	2186.00	2196.00	75600.00	

制单人 陈晨 审核人 陈晨

图 9-3 盘点单

其他入库单

表体排序
 蓝字 合并显示
 红字

入库单号 0000000002 入库日期 2025-06-30 仓库 床上用品仓库
 入库类别 盘盈入库 业务类型 盘盈入库 业务号 0000000001
 部门 仓管部 审核日期 2025-06-30 备注 _____

	存货编码	存货名称	规格型号	主计量单位	数量	单价	金额
1	0103	床笠(秋天)	1.8M	件	20.00	80.00	1600.
2							
3							
4							
5							
6							
7							
8							
9							
10							
11							
12							
13							
14							
	合计				20.00		1600.

制单人 陈晨 审核人 陈晨

图 9-4 盘盈生成的其他入库单

② 在库存管理系统中，执行“出库业务”|“其他出库单”命令，单击“末张”按钮，查看盘亏生成的其他出库单，单击“审核”按钮，如图9-5所示。

(3) 在存货核算系统中完成存货记账并根据出、入库单结转成本记账凭证

① 以会计人员“W02 张文华”身份登录，修改“注册日期”为“2025年6月30日”。

② 在存货核算系统中，完成对其他入库单、其他出库单的记账。

③ 执行“财务核算”|“生成凭证”命令，在“查询条件”对话框中，单击“全消”按钮，选中“(08)其他入库单”和“(09)其他出库单”复选框，生成相应记账凭证，如图9-6和图9-7所示。

特别提醒:

“盘亏”原因若无法查明或为人为保管不当,则还需在“盘亏出库记账凭证”中补录“应交税费——应交增值税(进项税额转出)”科目。

(4) 在总账管理系统中录入盘盈、盘亏批准后的记账凭证

在总账管理系统中,根据实验资料图9-1中的批准结果录入盘盈、盘亏批准后的相关会计凭证,如图9-8、图9-9所示。

转账凭证				附单据数: _____	
转 字 0029	制单日期: 2025.06.30	审核日期:			
摘要	科目名称	借方金额	贷方金额		
盘盈处理	待处理财产损益/待处理流动资产损益	160000			
盘盈处理	管理费用/存货盘点	160000			
票号		数量	合计		
日期	单价				
备注	项目	部门			
	个人	客户			
	业务员				
记账	审核	出纳	制单 张文华		

图 9-8 填制盘盈处理凭证

转账凭证				附单据数: _____	
转 字 0030	制单日期: 2025.06.30	审核日期:			
摘要	科目名称	借方金额	贷方金额		
盘亏处理	在途物资	120000			
盘亏处理	待处理财产损益/待处理流动资产损益		120000		
票号		数量	合计		
日期	单价				
备注	项目	部门			
	个人	供应商 山西春天			
	业务员 林群				
记账	审核	出纳	制单 张文华		

图 9-9 填制盘亏处理凭证

实训二 计提存货跌价准备



实训任务

2025年6月30日,接销售部通知,部分库存商品期末可变现净值低于成本,如图9-10所示。由财务部按要求计提存货跌价准备。

存货编码	商品名称	型号	可变现净值单价/元
0101	床垫(秋天)	1.8M	130.00 120.00

图 9-10 库存商品期末可变现净值一览表

任务解析

1. 背景知识

企业的存货应当在会计期末时,按照账面成本与可变现净值孰低法的原则进行计量,对于可变现净值低于存货账面成本的差额,计提存货跌价准备。对于存货跌价准备,企业应设置存货跌价准备账户核算企业提取的存货跌价准备。存货可变现净值低于成本的差额,借记资产减值损失账户,贷记存货跌价准备账户。如果计提跌价准备后的存货的价值又得以恢复,应按恢复增加的数额(其增加数应以补足以前入账的减少数为限),借记存货跌价准备账户,贷记资产减值损失账户。

2. 岗位说明

以会计W02身份完成存货核算系统业务处理。

实训指引

(1) 在存货核算系统中计提存货跌价准备

① 以会计人员“W02 张文华”身份登录,修改“注册日期”为“2025年6月30日”。

② 在存货核算系统中,执行“跌价准备”|“计提跌价准备”命令,打开“计提跌价处理单”窗口,单击“增加”按钮,根据实验资料图9-10录入相关信息,依次单击“保存”“审核”按钮,如图9-11所示。

	存货编码	存货名称	规格型号	计量单位	结存数量	结存单价	结存金额	可变现价格	可变现金额	应计提金额	已计提金额	本次计提金额
1	0101	床垫(秋天)	1.8M	件	100.00	180.00	18000.00	130.00	13000.00	5000.00	0.00	5000.00
2												
3												
4												
5												
6												
7												
8												
	合计				100.00		18000.00		13000.00	5000.00	0.00	5000.00

图 9-11 计提跌价处理单

(2) 在存货核算系统中生成计提存货跌价准备记账凭证

执行“跌价准备”|“跌价准备制单”命令，单击“选择”按钮，在打开的“查询条件”对话框中，单击“全消”按钮，选中“(47)跌价准备单”复选框，生成计提存货跌价准备记账凭证，如图9-12所示。

已生成		转账凭证		附单据数: 1	
凭证字: 0031	制单日期: 2025.06.30	审核日期:			
摘要	科目名称	借方金额	贷方金额		
跌价准备	资产减值损失	500000			
跌价准备	存货跌价准备		500000		
合计		500000	500000		
制单日期	数量 单价	部门		客户	
备注	个人 业务员				
记账	审核	出纳	制单 张文华		

图 9-12 计提存货跌价准备凭证

实训三 期末结账



实训任务

2025年6月30日，业务部门与财务部门完成期末结账。

任务解析

1. 背景知识

期末结账主要包括两项内容：期末处理和月末结账。在采购与应付、销售与应收、库存管理、存货核算集成应用模式下，期末处理只涉及存货核算子系统，月末结账则涉及所有系统。

(1) 期末处理

当日常业务全部完成后，存货核算系统需要进行期末处理。对于采用全月平均法核算的存货，期末处理时计算全月平均单价及本月出库成本；对于按计划价/售价核算的存货，期末处理时计算差异率并分摊差异；对于按照移动平均法、先进先出法、后进先出法、个别计价法核算的存货，期末处理仅作为本月业务完结的标志。

(2) 月末结账

结账表示本期业务处理终结。在财务业务一体化各系统集成应用时，期末结账要遵从一定的顺序。按照子系统之间的数据传递关系，各子系统结账的先后顺序如下：

- 采购管理、销售管理
- 应付款管理、应收款管理、库存管理

- 存货核算
- 总账

2. 岗位说明

以业务员G01身份完成采购管理系统结账；以业务员X01身份完成销售管理系统结账；以业务员C01身份完成库存管理系统结账。以会计W02身份完成存货核算系统、应收应付款管理系统结账，以及总账管理系统内期间损益结转与凭证记账处理；以财务经理W01身份完成凭证审核及总账结账；以出纳W03身份完成凭证出纳签字处理。

实训指引



(1) 采购管理系统结账

- ① 以采购人员“G01 林群”身份登录，修改“注册日期”为“2025年6月30日”。
- ② 在采购管理系统中，执行“月末结账”命令，打开“结账”对话框，单击6月所在行，选中该条记录。单击“结账”按钮，系统弹出“月末结账”信息提示框，提示“是否关闭订单？”，单击“否”按钮，完成采购管理系统结账，如图9-13所示。

❖ 特别提醒：

- ❖ 采购管理系统若需取消结账，可以在“结账”对话框中单击“取消结账”按钮实现。
- ❖ 若应付款管理系统或库存管理系统或存货核算系统已结账，则采购管理系统不能取消结账。

(2) 销售管理系统结账

- ① 以销售人员“X01 王大国”身份登录，修改“注册日期”为“2025年6月30日”。
- ② 在销售管理系统中，执行“月末结账”命令，打开“结账”对话框，单击6月所在行，选中该条记录。单击“结账”按钮，系统弹出“月末结账”信息提示框，提示“是否关闭订单？”，单击“否”按钮，完成销售管理系统结账，如图9-14所示。

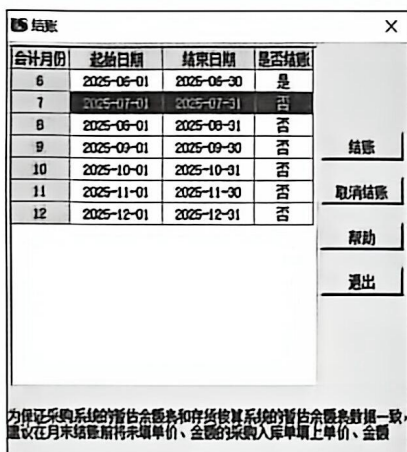


图 9-13 采购管理系统结账



图 9-14 销售管理系统结账

❖ 特别提醒：

- ❖ 销售管理系统与采购管理系统期末结账不存在先后顺序。
- ❖ 销售管理系统若需取消结账，可以在“结账”对话框中单击“取消结账”按钮实现。
- ❖ 若应收款管理系统或库存管理系统或存货核算系统已结账，则销售管理系统不能取消结账。

(3) 库存管理系统结账

① 以仓管人员“C01 陈晨”身份登录，修改“注册日期”为“2025年6月30日”。

② 在库存管理系统中，执行“月末结账”命令，打开“结账”对话框，单击6月所在行，选中该条记录。单击“结账”按钮，系统弹出“库存管理”信息提示框，提示“库存启用月份结账后将不能修改期初数据，是否继续结账？”，如图9-15所示。

③ 单击“是”按钮，完成库存管理系统结账，如图9-16所示。

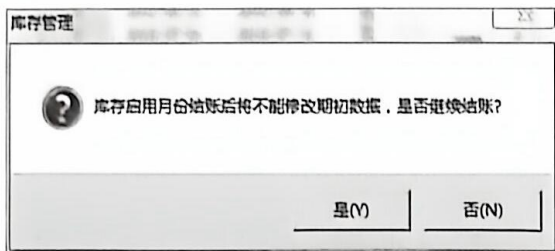


图 9-15 库存管理结账提示框

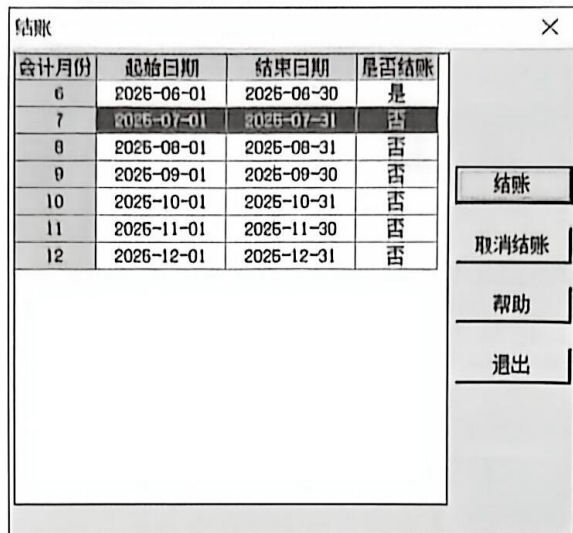


图 9-16 库存管理系统结账

(4) 存货核算系统结账

① 以会计人员“W02 张文华”身份登录，修改“注册日期”为“2025年6月30日”。

② 在存货核算系统中，执行“财务核算”|“与总账系统对账”命令，进入“与总账对账”窗口。

③ 选中“包含未记账凭证”复选框，查看对账结果，白色显示记录表示对账结果相平，蓝色显示记录表示对账结果不平，如图9-17所示。单击“退出”按钮返回。

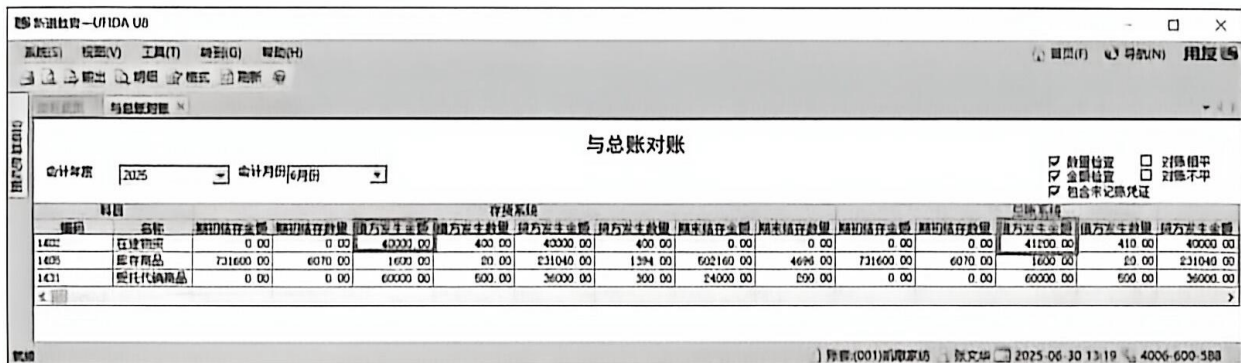


图 9-17 与总账对账

❖ 特别提醒：

本例中“1402 在途物资”科目余额与总账对账存在的差额系盘点业务中山西春天家居用品制造有限公司补发货物尚未运达所致。

④ 同理，完成“发出商品与总账对账”，如图9-18所示。

图 9-18 发出商品与总账对账

⑤ 执行“业务核算”|“期末处理”命令，在打开的对话框中单击“处理”按钮，对所有仓进行期末处理，如图9-19所示。

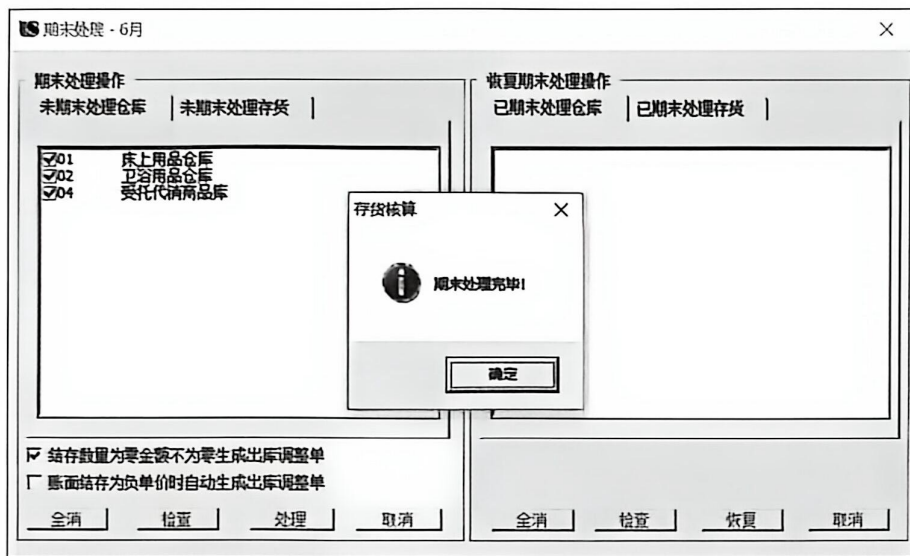


图 9-19 期末处理

⑥ 执行“业务核算”|“月末结账”命令，对存货核算系统结账，如图9-20所示。

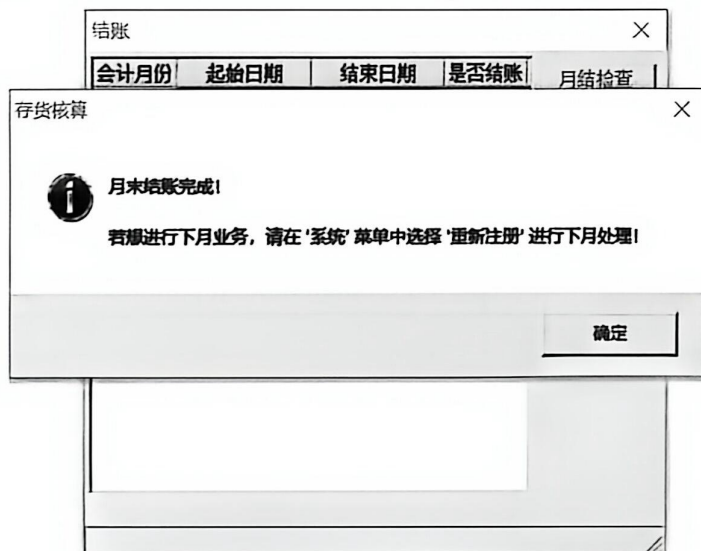


图 9-20 存货核算系统结账

❖ 特别提醒:

❖ 存货核算系统若需取消结账,需以“下月”任意日期登录系统服务平台,在“结账”对话框中单击“取消结账”按钮实现。

(5) 应付款管理系统结账

① 在应付款管理系统中,执行“期末处理”|“月末结账”命令,打开“月末处理”对话框,双击六月份的“结账标志”栏,如图9-21所示。

② 单击“下一步”按钮,屏幕显示所有“处理类型”的处理情况,如图9-22所示。

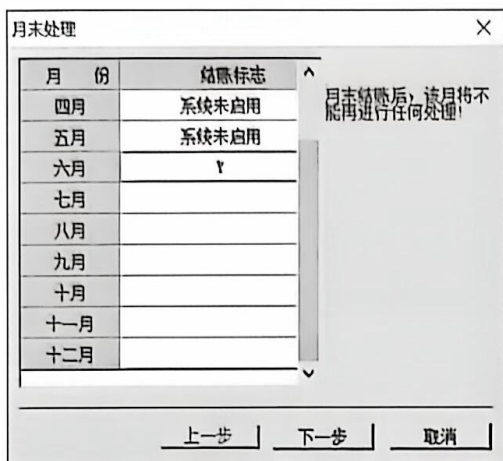


图 9-21 月末处理

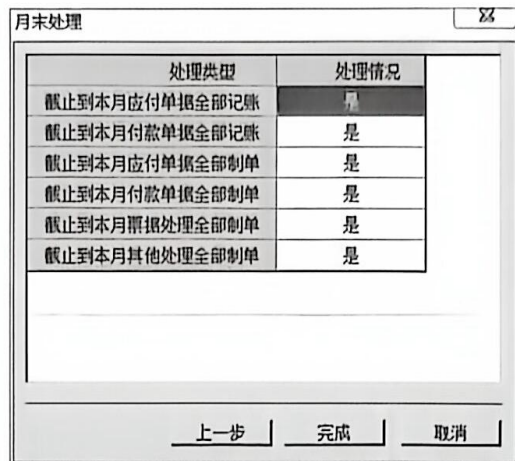


图 9-22 处理情况

③ 在处理情况都是“是”的情况下,单击“完成”按钮,结账后,系统弹出“6月份结账成功!”信息提示对话框,如图9-23所示。

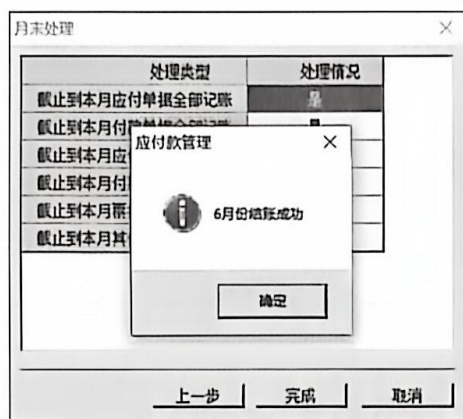


图 9-23 应付款管理系统结账

④ 单击“确定”按钮,系统自动在对应的结账月份的“结账标志”栏中显示“已结账”字样。

❖ 特别提醒:

- ❖ 本月的单据在结账前应该全部审核;本月的结算单据在结账前应全部核销。
- ❖ 若应付款管理系统与采购管理系统集成使用,应在采购管理系统结账后,才能对应付款管理系统进行结账处理。
- ❖ 应付款管理系统若需取消结账,可通过执行“期末处理”|“取消月结”命令实现。
- ❖ 如果当月总账管理系统已经结账,则应付款管理系统不能取消结账。

(6) 应收款管理系统结账

同理，完成应收款管理系统结账。

❖ 特别提醒：

- ❖ 应收款管理系统与应付款管理系统期末结账不存在先后顺序。
- ❖ 应收款管理系统若需取消结账，可通过执行“期末处理”|“取消月结”命令实现。

(7) 总账系统结账

① 以“W01 汪小菲”身份登录完成凭证批量审核。请参考前述业务操作指导，在此不再赘述。

② 以出纳人员“W03 黄宁”身份登录，完成凭证出纳签字。请参考前述业务操作指导，在此不再赘述。

③ 以会计人员“W02 张文华”身份登录，完成凭证记账，如图9-24所示。请参考前述业务操作指导，在此不再赘述。

记账

记账选择
 2025.06月份凭证 其他月份调整期凭证

期初试算平衡表

资产 - 借 11,593,640.00	负债 - 贷 2,427,007.50
共同 - 平	权益 - 贷 9,166,632.50
成本 - 平	损益 - 平
合计 - 借 11,593,640.00	合计 - 贷 11,593,640.00

试算结果平衡

确定 打印

记账 记账报告 退出

图 9-24 总账记账

④ 完成期间损益结转，并对生成的凭证审核、记账，如图9-25、图9-26所示。请参考前述业务操作指导，在此不再赘述(注意更换操作员)。

已生成		转账凭证		附单据数: 0	
转 字 0032	制单日期: 2025.06.30	审核日期:			
摘要	科目名称	借方金额	贷方金额		
期间损益结转	本年利润				55182000
期间损益结转	主营业务收入	55182000			
票号	数量	1834.00件(套个)	合计	55182000	55182000
日期	单价	300.89			
备注	项目	部门	客户		
	个人				
	业务员				
记账	审核	出纳	制单	张文华	

图 9-25 期间损益结转凭证——收入

已生成		转账凭证			
转 字 0033 - 0001/0002		制单日期: 2025.06.30		审核日期: 附单据数: 0	
摘要	科目名称	借方金额	贷方金额		
期间损益结转	本年利润	34558249			
期间损益结转	主营业务成本		25484000		
期间损益结转	销售费用/委托代销手续费		429400		
期间损益结转	管理费用/存货盘点		160000		
期间损益结转	财务费用		270949		
票号	数量	合计	34558249	34558249	
日期	单价				
备注	项目	部门			
	个人	客户			
	业务员				
记账	审核	出纳	制单	张文华	

图 9-26 期间损益结转凭证——支出

⑤ 以账套主管“W01 汪小菲”身份登录审核凭证，以会计人员“W02 张文华”身份登录完成凭证记账，最后由账套主管“W01 汪小菲”查看2025年6月工作报告，如图9-27所示，完成总账结账。

结账		2025年06月工作报告	
开始结账	1. 本月损益类未结转为零的一级科目	无	
核对账簿	2. 本月账面试算平衡	资产 = 借 11930375.89 负债 = 贷 2557585.88 共同 = 平 0.00 权益 = 贷 9372790.01 成本 = 平 0.00 损益 = 平 0.00	
月度工作报告	合计借 11930375.89 = 贷 11930375.89	试算结果平衡	
完成结账			
打印月度工作报告	上一步	下一步	取消

图 9-27 结账——月度工作报告

⑥ 将账套输出至“D:\凯撒家纺\9-1库存与存货”。

读者对象

本书可用作高等院校会计学及经济管理类其他专业的教材，也可用作全国大学生会计信息化技能大赛的辅导教材，还可用作在职会计人员的用友U8自学读本。

本书特色

实训主导，理论简明 每章内容按实训任务、任务解析、实训指引和拓展任务展开，先实践再理论后进阶，并以原始凭据为牵引，加深学生对企业真实业务的了解及对基本原理的把握。

内容独立，自由组合 实验业务按系统模块设置为“拼板”方式，既可以由上至下顺序进行，也可以任意选择其中的若干实验进行组合，给予教学较大的自由空间。

资源丰富，便于自学 本书提供丰富的教学资源——用友U8 V10.1教学版安装程序、实验账套、教学用PPT、微课视频，极大地方便了课堂教学和自学。

业务典型，实用性强 本书以企业财务工作实际业务流程为主线，按照财务人员的日常处理业务类型谋篇布局，业务典型，可操作性强，具有较强的实用性。

系列书目

会计信息化实训教程——供应链（用友U8 V10.1）（第3版）（云实训）

会计信息化实训教程——财务链（用友U8 V10.1）（第3版）（云实训）

会计信息化实训教程——财务业务一体化（用友U8 V10.1）（第2版）（微课版）

清华大学出版社



富方微借号

ISBN 978-7-302-69589-9



9 787302 695899 >

定价：68.00元



智能会计人才培养新形态系列教材

教学资源

- 教学课件
- 案例文件
- 微课视频

Excel财务会计 与管理会计应用教程

微课版

李田 彭飞 宫苗苗 主编

金杰海 于淼 副主编

内 容 简 介

本书立足于业务情景导入与 Excel 软件实训应用, 共分为 10 章, 分别为 Excel 应用基础, 会计原始单据制作, 财务报表数据分析, 薪酬管理与分析, 固定资产管理与分析, 采购、销售、库存数据分析, 成本管理数据分析, 营运管理数据分析, 投融资决策分析与预测, 预算管理数据分析与预测, 基本涵盖了管理会计模块的主要内容。

本书既可作为高等院校会计学专业及经济管理类相关专业的教学用书, 也可作为在职会计人员学习 Excel 财务应用的自学读本。

本书封面贴有清华大学出版社防伪标签, 无标签者不得销售。

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

Excel 财务会计与管理会计应用教程: 微课版 /
李田, 彭飞, 宫苗苗主编. -- 北京: 清华大学出版社,
2025. 7. -- (智能会计人才培养新形态系列教材)
ISBN 978-7-302-69561-5

I . F234.4-39

中国国家版本馆 CIP 数据核字第 2025BD2871 号

责任编辑: 刘金喜

封面设计: 何凤霞

版式设计: 思创景点

责任校对: 成凤进

责任印制: 宋 林

出版发行: 清华大学出版社

网 址: <https://www.tup.com.cn>, <https://www.wqxuetang.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: 三河市龙大印装有限公司

经 销: 全国新华书店

开 本: 185mm×260mm 印 张: 16 字 数: 380 千字

版 次: 2025 年 9 月第 1 版 印 次: 2025 年 9 月第 1 次印刷

定 价: 68.00 元

产品编号: 090474-01

第1章	Excel应用基础	1	实训三	库存汇总分析	146
	实训一	Excel 窗口介绍			1
	实训二	Excel 数据管理基本操作			2
第2章	会计原始单据制作	14	第7章	成本管理数据分析	161
	实训一	财务数据录入与管理		实训一	建立“静态的本量利分析模型”
	实训二	图表制作与函数应用			161
	实训三	记账凭证录入及查询		实训二	建立“动态的本量利分析模型”
	实训四	会计账簿			171
	实训五	会计报表		实训三	分析“单因素下的盈亏平衡销量”
					176
第3章	财务报表数据分析	74		实训四	分析“多因素下的盈亏平衡销量”
	实训一	偿债、营运与盈利能力财务指标计算			180
				实训五	标准成本系统分析模型设计
	实训二	杜邦分析模型设计			185
	实训三	目标利润规划分析模型设计		实训六	作业成本系统分析模型设计
					186
第4章	薪酬管理与分析	95	第8章	营运管理数据分析	189
	实训一	薪资项目设置		实训一	最佳现金持有量模型设计
					189
	实训二	薪资项目查询		实训二	应付账款管理与分析
					195
	实训三	薪资项目汇总与分析		实训三	应收账款管理与分析
					201
	实训四	编制工资费用分配表		实训四	信用政策调整模型设计
					207
第5章	固定资产管理与分析	113		实训五	建立“经济订货批量模型”
	实训一	固定资产计提折旧方法及模型设计			216
				第9章	投融资决策分析与预测
	实训二	不同折旧计算方法对比应用			221
				实训一	货币时间价值及函数应用
	实训三	固定资产更新决策模型设计			221
				实训二	债券融资模型设计
					224
第6章	采购、销售、库存数据分析	124		实训三	长期借款模型设计
	实训一	采购成本分析			225
				实训四	建立无风险项目投资决策模型
	实训二	销售利润分析			230
				第10章	预算管理数据分析与预测
					235
				实训一	经营预算模型设计
					237
				实训二	专门预算模型设计
					244

投融资决策分析与预测

实训一 货币时间价值及函数应用

【实训任务】

小张今年30岁，购买了某种养老保险，从今年起每年年初交纳保费1600元，交费期为20年。从60岁开始，保险公司每年年初向小张支付养老金6000元，一直到小张去世为止，小张去世时还能获得一次性抚恤金5000元。假定保险公司的投资收益率为5%，请利用Excel相关函数计算，如果仅就投资收益分析而言，小张至少要活到多少岁投保才划算。

❖ 特别提醒：

货币的时间价值，是指货币经过一段时间的投资和再投资所增加的价值。货币时间价值的存在使得不同时点上相同数量货币的价值并不相等。与货币时间价值有关的概念如下。

1. 复利终值

复利终值是指在复利计息方式下，现在的一笔资金，经过若干期后的本利和。复利终值的计算公式如下。

$$FV = PV \times (1+i)^n$$

其中，FV是资金的复利终值；PV是资金的复利现值； i 是利率； n 是计息期数。

2. 复利现值

复利现值是指在复利计息方式下，未来若干期的一笔资金折算到现在时点的价值。复利现值的计算公式如下。

$$TC = \frac{FV}{(1+i)^n}$$

3. 年金终值

年金是指定期、等额的系列收支，即在某一期限内，每隔一定相同的时期，收入或支

出相等金额的款项。年金分为普通年金、先付年金(预付年金)和永续年金。年金终值是指年金按复利计算,在若干期后的期末可得到的本利和。永续年金只有现值,没有终值。

(1) 普通年金终值。普通年金终值的计算公式如下。

$$PV = A \times \frac{1 - (1 + i)^n}{i}$$

其中, A 是年金; i 是利率; n 是期数; FV 是年金的终值; 年金终值系数简记为($F/A, i, n$)。

(2) 预付年金终值。预付年金曲线图值的计算公式如下。

$$FV = A \times \frac{(1 + i)^{n+1} - 1}{i} - 1$$

4. 年金现值

年金现值是指为了在每期取得相等金额,现在需要投入的金额。

(1) 普通年金现值。普通年金现值的计算公式如下。

$$PV = A \times \frac{1 - (1 + i)^{-n}}{i}$$

其中, A 是年金; i 是利率; n 是期数; PV 是年金的现值; 年金现值系数简记为($P/A, i, n$)。

(2) 预付年金现值。预付年金现值的计算公式如下。

$$PV = A \times \left[\frac{1 - (1 + i)^{-(n-1)}}{i} + 1 \right]$$

5. 永续年金

如果年金定期等额支付一直持续到永远,称为永续年金。永续年金现值的计算公式为 $PV = \frac{A}{i}$ 。

【任务解析】

如果仅就投资收益分析而言,小张至少要活到多少岁投保才划算?该问题实际上是求年龄问题,即求年金期间的问题。小张30~50岁时,每年年初交1600元,这20年每年初的1600元是20年预付年金的形式,而从60岁起至去世,保险公司每年年初支付6000元,同样也是年金的形式,但由于去世时间未知,年金时间待求。小张前20年的保费可以换来从60岁到去世的养老金和一次性抚恤金,也就是说,前20年年金到60岁的终值可以作为60岁后年金的现值,因此我们选择小张60岁作为解题的出发点,计算60岁到去世年金期间。在进行货币时间价值的Excel函数应用时,需要了解以下函数。

1. 年金终值函数FV(rate,nper,pmt,pv,type)

该函数基于固定利率和等额分期付款的方式,返回某项投资的未来值。各参数的含义如下。**rate**: 复利利率; **nper**: 年金期数; **pmt**: 每期固定支付或收入的金额; **pv**: 投资开始计算时已经入账的价值,默认值为0; **type**: 年金类型,当取1时表示预付年金,当取0或默认时表示普通年金。

2. 年金现值函数FV(rate,nper,pmt,fv,type)

该函数用于返回某项投资的一系列等额分期偿还额的当前值之和，即年金的现值。fv是指在最后一次付款期后获得的一次性偿还额，默认值为0。其余参数的含义和年金终值函数对应参数的含义相同。

3. 年金函数PMT(rate,nper,pv,fv,type)

该函数用于返回固定利率下投资或贷款的等额分期偿还额(包括本金和利息)。参数的含义和年金终值函数对应参数的含义相同。

4. 年金本金函数PPMT(rate,per,nper,pv,fv,type)

该函数用于返回在定期偿还、固定利率条件下给定期次内某项投资回报(或贷款偿还)的本金部分。参数的含义和年金终值函数对应参数的含义相同。

5. 年金利息函数IPMT(rate,per,nper,pv,fv,type)

该函数用于返回在固定利率、期数下某项投资回报或贷款偿还的利息部分。参数的含义和年金终值函数对应参数的含义相同。

6. 期数函数NPER(rate,pmt,pv,fv,type)

该函数用于返回每期付款金额及利率固定的某项投资或贷款的期数。参数的含义和年金终值函数对应参数的含义相同。

7. 利率函数RATE(nper,pmt,pv,fv,type,guess)

该函数在已知期数、每期付款及现值或终值的条件下，返回年金的每期利率。参数的含义和年金终值函数对应参数的含义相同。

【实训指引】

1. 计算前20年现金流的终值之和

第一步：分析现金流的形式。由于小张30~50岁每年年初需要交1600元保费，因此这20期的现金流是20年预付年金的形式。

第二步：利用年金终值函数FV(rate,nper,pmt,pv,type)计算20年预付年金的终值。保险公司的投资收益率为5%，因此函数的第一个参数rate为5%。第二个参数nper为年金期数，即为20。第三个参数pmt为每期固定支付或收入的金额，即为1600。第四个参数pv指投资开始计算时已经入账的价值，即为0。第五个参数type为年金类型，当取1时表示预付年金，当取0或默认时表示普通年金，即为1。因此，小张30~50岁每年年初交的1600元保费在50岁那年年末的终值之和为FV(5%,20,1600,0,1)。

2. 计算前20年现金流到60岁的终值

由于我们选择小张60岁作为解题的出发点，因此在计算前20年现金流的终值之和之后，还需要计算50~60岁的保费价值。

利用复利终值的原始公式 $FV = PV \times (1+i)^n$ 进行计算。投资收益率为5%，时间为10年，因此50~60岁的保费价值为 $FV(5\%,20,1600,0,1) \times (1+5\%)^{10}$ 。

3. 计算60岁到去世期间

利用期数函数 $NPER(rate,pmt,pv,fv,type)$ 计算60岁到去世期间。保险公司的投资收益率为5%，因此第一个参数 $rate$ 为5%。第二个参数 pmt 就是小张每年年初领取的养老保险6000元；第三个参数是现值 p_v ，实际上是小张30~50岁每年年初缴纳的保费计算到60岁时的价值。 $FV(5\%,20,1600,0,1)$ 计算的是小张30~50岁每年年初缴纳的保费到50岁时的价值； $FV(5\%,20,1600,0,1) \times (1+5\%)^{10}$ 是每年所交保费从50~60岁时的价值，也即第三个参数的值。第四个参数可以将抚恤金5000元看作终值 f_v ；第五个参数 $type$ 由于是年初付，因此为1。

图9-1可以帮助大家很好地理解期数函数 $NPER()$ 的意义和Excel的计算过程。

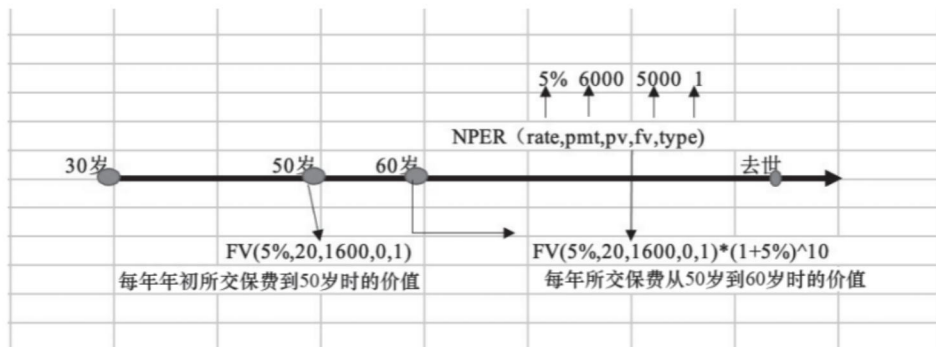


图 9-1 期数函数的计算

因此，所求年龄 = $NPER(5\%,6000,FV(5\%,20,1600,0,1) \times (1+5\%)^{10},5000,1)+60=85.13$ 岁。

实训二 债券融资模型设计

【实训任务】

某公司欲发行面值为1000元的债券若干，债券期限为5年，票面利率为10%，每年年末支付利息，到期偿还本金，若市场实际利率为12%，请问该债券的发行价格是多少？

❖ 特别提醒：

企业债券是企业为筹措长期资金而发行的债券。一般而言，债券包括以下基本要素。

- (1) 债券面值。债券面值包括币种和票面金额两项基本内容。
- (2) 债券期限。债券期限是指发行日至到期日的时间。
- (3) 债券利率。债券上标明的利率一般是年利率，且多为固定利率，近年来也有浮动利率。也有的债券票面利率为零，到期按面值偿还。
- (4) 债券价格。债券价格可分为发行价格与市场交易价格。发行价格是指投资者在发行市场(一级市场)上购买债券时实际支付的价格。债券可以平价、溢价或折价发行，因此，债券发行价格既可能等于面值，也可能高于或低于面值。市场交易价格是指债券发行后，投资者在流通市场(二级市场)上交易债券的价格。

【任务解析】

债券估价就是对债券的价格进行估计。投资者进行债券投资时预期在未来一定时期内

会收到包括本金和利息在内的现金流入，债券价格则是投资者为了取得未来的现金流入而愿意投入的资金。理论上，债券发行价格是债券的面值和支付的年利息按发行当时的市场利率折现所得到的现值。因此选用年金现值函数PV来解决这个问题。

年金现值函数PV(rate,nper,pmt,fv,type)用于返回某项投资的一系列等额分期偿还额的当前值之和，即年金的现值。各参数的含义如下。

rate: 复利利率; nper: 年金期数; pmt: 每期固定支付或收入的金额; fv: 最后一次付款期后获得的一次性偿还额，默认值为0; type: 年金类型，当取1时表示预付年金，当取0或默认时表示普通年金。

【实训指引】

1. 计算票面利息

债券的利息=票面面值×票面利率。该债券面值为1 000元，票面利率为10%，因此票面利息=1000×10%=100元。也就是说，该债券5年内每年年末需要支付利息100元。

2. 利用年金现值函数计算债券价值

利用年金现值函数PV(rate,nper,pmt,fv,type)计算债券价值。市场实际利率为12%，因此函数第一个参数rate为12%。第二个参数nper为年金期数5。第三个参数pmt为每期固定支付或收入的金额，即为年利息-100。第四个参数fv是最后一次付款期后获得的一次性偿还额，即为本金-1000。第五个参数type为年金类型，当取1时表示预付年金，当取0或默认时表示普通年金，即为0。因此，债券的发行价格=PV(12%,5,-100,-1000,0)=927.9元。

实训三 长期借款模型设计

【实训任务】

假设企业贷款200万元，贷款利率可能在6%~12%之间变动(每次变动0.25%)，贷款年限3~10年。请利用Excel设计一个长期借款分析模型，分析当贷款金额一定时，在不同贷款利率和贷款年限组合下每期需要偿还的金额。

❖ 特别提醒:

长期借款是指企业向银行或其他金融机构借入的，期限在一年以上(不含一年)或超过一年的一个营业周期以上的各项借款。企业长期借款主要用于固定资产投资或更新改造、科技开发和新产品试制等。

【任务解析】

若要建立“长期借款分析模型”，即计算不同贷款利率和贷款年限组合下每期需要偿还的金额，则可以利用年金函数PMT(rate,nper,pv,fv,type)来计算。该函数用于返回固定利率下投资或贷款的等额分期偿还额(包括本金和利息)。各参数的含义如下。rate: 复利利率; nper: 年金期数; pv: 投资开始计算时已经入账的价值，默认值为0; fv: 最后一次付款期后获得的一次性偿还额，默认值为0; type: 年金类型，当取1时表示预付年金，当取0或默

认时表示普通年金。

接下来，可以通过Excel的辅助工具来更直观地观察不同数值下的情况。此时可以利用双变量模拟运算表和窗体控件来设计长期借款分析模型。

【实训指引】

1. 利用双变量模拟运算表设计长期借款分析模型

第一步：新建工作簿和工作表。新建一个工作簿并以“长期借款”为名进行保存，将其中的工作表Sheet1重命名为“长期借款分析模型”，在其中输入表题与相应的表格项目，并设置单元格格式，如图9-2所示。其中，B列为从6%到12%不同的贷款利率，每次变动0.25%，第2行为从3年到10年不同的贷款年限。

	A	B	C	D	E	F	G	H	I	J
1					长期借款分析模型					
2			3	4	5	6	7	8	9	10
3		6%								
4		6.2500%								
5		6.5000%								
6		6.7500%								
7		7.0000%								
8		7.2500%								
9		7.5000%								
10		7.7500%								
11		8.0000%								
12		8.2500%								
13		8.5000%								
14		8.7500%								
15		9.0000%								
16		9.2500%								
17		9.5000%								
18		9.7500%								
19		10.0000%								
20		10.2500%								
21		10.5000%								
22		10.7500%								
23		11.0000%								
24		11.2500%								
25		11.5000%								
26		11.7500%								
27		12.0000%								

图 9-2 输入表格项目并设置单元格格式

第二步：输入年金函数。在B2单元格输入计算每期还款额的函数“=PMT(A3,A2,-200000)”，如图9-3所示。本例中贷款期限和利率均为不确定的数据，因此PMT公式中使用了两个变量A2和A3，分别代表期限和利率。按Enter键确认，B2单元格中出现“#NUM!”错误提示，可不予理会。

	A	B	C	D	E	F	G	H	I	J
1					长期借款分析模型					
2		#NUM!	3	4	5	6	7	8	9	10
3		6%								
4		6.2500%								
5		6.5000%								
6		6.7500%								
7		7.0000%								
8		7.2500%								
9		7.5000%								
10		7.7500%								
11		8.0000%								
12		8.2500%								
13		8.5000%								
14		8.7500%								
15		9.0000%								
16		9.2500%								
17		9.5000%								
18		9.7500%								
19		10.0000%								
20		10.2500%								
21		10.5000%								
22		10.7500%								
23		11.0000%								
24		11.2500%								
25		11.5000%								
26		11.7500%								
27		12.0000%								

图 9-3 输入函数

第三步：建立模拟运算表。选择B2:J27单元格区域，选择“数据”|“模拟分析”|“模拟运算表”命令，如图9-4所示。

第四步：打开“模拟运算表”对话框，在“输入引用行的单元格”文本框中，选择A2单元格，同样在“输入引用列的单元格”文本框中，选择A3单元格，完成后单击“确定”按钮，如图9-5所示。最终结果如图9-6所示。

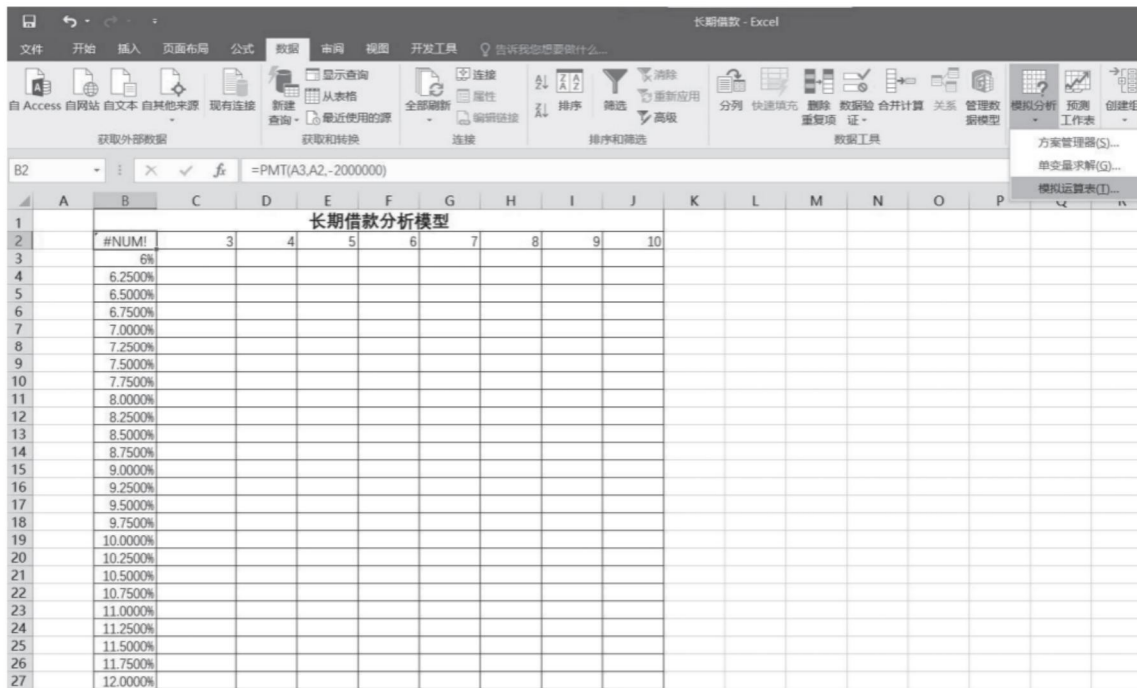


图 9-4 建立模拟运算表

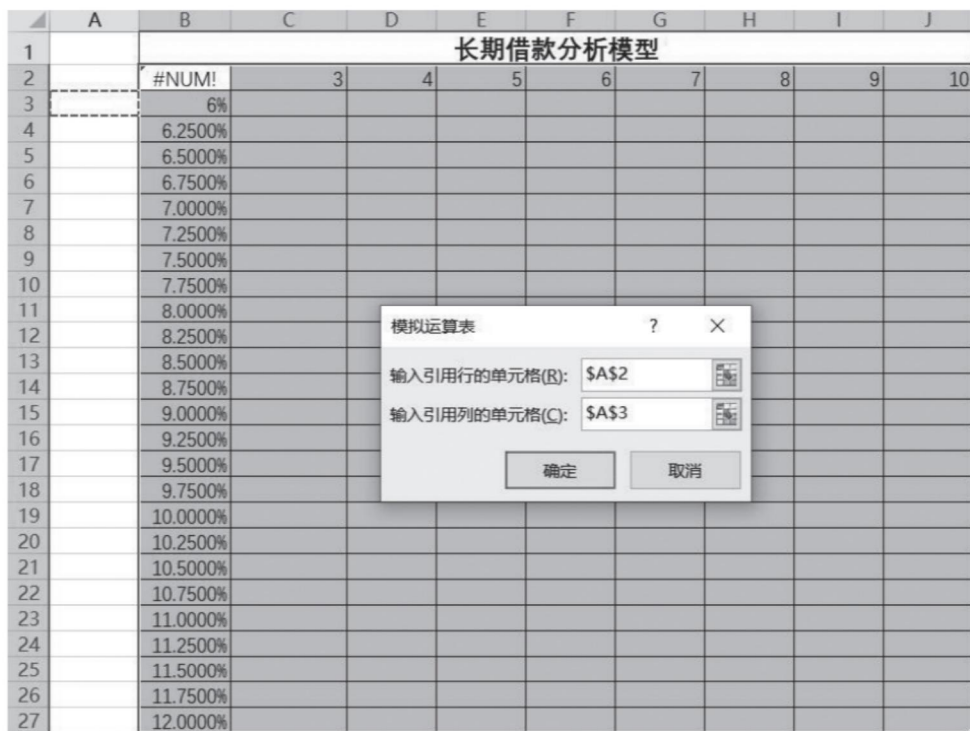


图 9-5 引用单元格

	A	B	C	D	E	F	G	H	I	J
1		长期借款分析模型								
2		#NUM!	3	4	5	6	7	8	9	10
3		6%	¥748,219.63	577183	474792.8	406725.3	358270	322071.9	294044.5	271735.9
4		6.2500%	751682.987	580490.7	478026.4	409925.5	361460	325265.9	297252.1	274963.6
5		6.5000%	755151.404	583805.5	481269.1	413136.6	364662.7	328474.6	300476.1	278209.4
6		6.7500%	758624.858	587127.3	484520.7	416358.7	367878.2	331697.8	303716.4	281473.2
7		7.0000%	762103.331	590456.2	487781.4	419591.6	371106.4	334935.5	306972.9	284755
8		7.2500%	765586.804	593792.1	491051	422835.3	374347.3	338187.6	310245.6	288054.6
9		7.5000%	769075.256	597135	494329.4	426089.8	377600.6	341454	313534.3	291371.9
10		7.7500%	772568.671	600484.9	497616.8	429355	380866.5	344734.7	316839	294706.7
11		8.0000%	776067.028	603841.6	500912.9	432630.8	384144.8	348029.5	320159.4	298059
12		8.2500%	779570.31	607205.3	504217.8	435917.2	387435.5	351338.4	323495.6	301428.6
13		8.5000%	783078.497	610575.8	507531.5	439214.2	390738.4	354661.3	326847.4	304815.4
14		8.7500%	786591.572	613953.1	510853.9	442521.6	394053.7	357998.1	330214.8	308219.3
15		9.0000%	790109.515	617337.3	514184.9	445839.6	397381	361348.8	333597.6	311640.2
16		9.2500%	793632.308	620728.3	517524.6	449167.9	400720.5	364713.2	336995.7	315077.9
17		9.5000%	797159.934	624126	520872.8	452506.6	404072.1	368091.2	340409.1	318532.3
18		9.7500%	800692.373	627530.5	524229.6	455855.5	407435.6	371482.9	343837.6	322003.3
19		10.0000%	804229.607	630941.6	527595	459214.8	410811	374888	347281.1	325490.8
20		10.2500%	807771.619	634359.4	530968.8	462584.2	414198.3	378306.6	350739.5	328994.6
21		10.5000%	811318.391	637783.9	534351	465963.7	417597.3	381738.6	354212.8	332514.6
22		10.7500%	814869.903	641215	537741.6	469353.4	421008.1	385183.7	357700.7	336050.8
23		11.0000%	818426.139	644652.7	541140.6	472753.1	424430.5	388642.1	361203.3	339602.9
24		11.2500%	821987.081	648097	544547.9	476162.8	427864.6	392113.6	364720.4	343170.8
25		11.5000%	825552.71	651547.8	547963.5	479582.5	431310.1	395598	368251.9	346754.4
26		11.7500%	829123.009	655005.1	551387.4	483012	434767.1	399095.4	371797.8	350353.6
27		12.0000%	832697.961	658468.9	554819.5	486451.4	438235.5	402605.7	375357.8	353968.3

图 9-6 最终结果

2. 利用窗体控件设计长期借款分析模型

第一步：新建工作表Sheet2并将其重命名为“长期借款分析模型2”，在其中输入表题与相应的表格项目，并设置单元格格式，如图9-7所示。

第二步：先在C3单元格输入函数“=D3/400”，并设置C3单元格的格式为百分数，保留两位小数。其次利用年金函数计算每期还款额，即在C5单元格输入函数“=PMT(C3,C4,C2)”，最终结果如图9-8所示。

	A	B	C	D
1		长期借款分析模型		
2		借款额	-2000000	
3		利率		24
4		期数		
5		每期还款		
6				
7				

图 9-7 长期借款分析模型 2

	A	B	C	D	E
1		长期借款分析模型			
2		借款额	-2000000		
3		利率	6%	24	
4		期数			
5		每期还款	#NUM!		
6					
7					

图 9-8 利用年金函数计算每期还款额

第三步：利用滚动条控件模拟变化的利率。选择“文件”|“选项”|“自定义功能区”|“开发工具”命令，如图9-9所示。

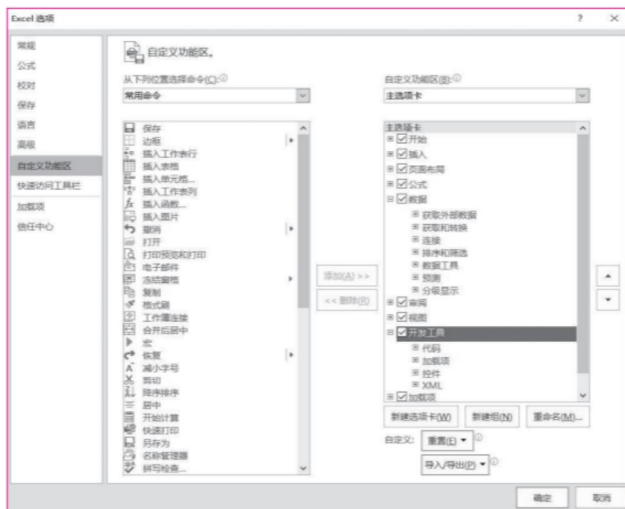


图 9-9 调出开发工具选项

第四步：选择“开发工具”|“控件”|“插入”|“滚动条”命令，在D3单元格中插入一个滚动条控件，如图9-10所示。



图 9-10 插入滚动条控件

第五步：设置控件格式。右击滚动条，在弹出的快捷菜单中选择“设置控件格式”命令，打开“设置对象格式”对话框。按照图9-11所示操作设置当前值、最小值、最大值、步长、页步长和单元格链接。在滚动条控件格式设置中，步长最小只能设为1，而利率的变化步长为0.25，因此可以通过设置中间变量D3来实现，令 $C3=D3/400$ ，当D3在24到48之间滚动时，C3就在6%到12%之间变化。

第六步：利用数值调节钮控件模拟期数。选择“开发工具”|“控件”|“插入”|“数值调节钮”命令，在C4单元格中插入一个数值调节钮控件，如图9-12所示。



图 9-11 设置对象格式



图 9-12 插入数值调节钮控件

第七步：设置控件格式。右击数据调节钮，在弹出的快捷菜单中选择“设置控件格式”命令，打开“设置对象格式”对话框。按照图9-13所示操作设置当前值、最小值、最大值、步长、页步长和单元格链接。

第八步：单击滚动条箭头可以观察利率的变化对每期还款额的影响，单击数值调节钮箭头可以观察还款期数的变化对每期还款额的影响，如图9-14所示。

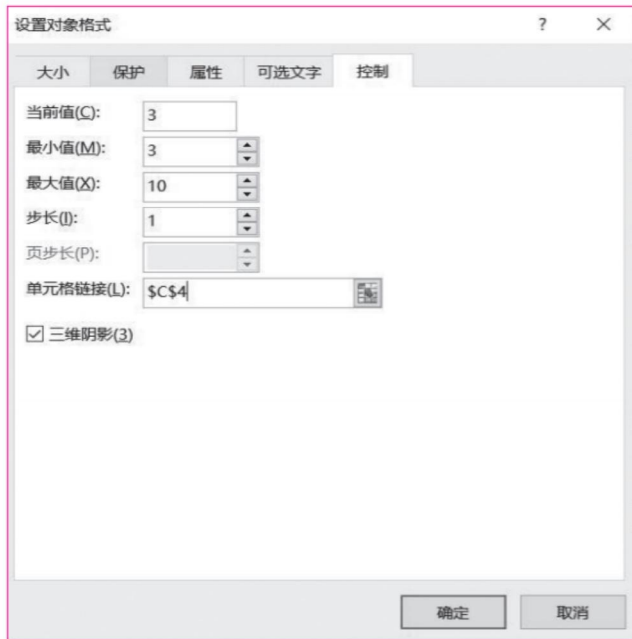


图 9-13 设置对象格式

	A	B	C	D
1		长期借款分析模型		
2		借款额	-2000000	
3		利率	6.50%	< > 26
4		期数	5	
5		每期还款	¥481,269.08	
6				
7				

图 9-14 最终结果

实训四 建立无风险项目投资决策模型

【实训任务】

某企业有A、B、C3个投资项目，3个项目的现金流量表如表9-1所示。假设资本成本率和再投资的资本成本率都为10%，请用Excel函数计算3个项目的净现值、现值指数、内含报酬率和修正内含报酬率。

表9-1 现金流量表

单位：元

期间	A项目	B项目	C项目
0	-100 000	-200 000	-300 000
1	50 000	20 000	100 000
2	40 000	40 000	90 000
3	30 000	60 000	80 000
4		80 000	70 000
5		100 000	60 000
6			50 000
7			40 000
8			30 000
9			20 000
10			10 000

❖ 特别提醒:

投资是指投放财力于某个对象以期在未来获取收益的经济行为。投资决策是指投资者为了实现其预期的投资目标,运用一定的科学理论、方法和手段,通过一定的程序对投资的必要性、投资目标、投资规模、投资方向、投资结构、投资成本与收益等经济活动中重大问题所进行的分析、判断和方案选择。投资决策是企业所有决策中最关键、最重要的决策。

评价投资方案时使用的指标有两类:一类是非贴现指标,它没有考虑资金的时间价值,主要有回收期、会计收益率等;另一类是贴现指标,它考虑了资金的时间价值,包括净现值、现值指数和内含报酬率。根据这两类指标,投资项目评价方法也相应地分为非贴现的评价方法和贴现的评价方法。

【任务解析】

若要建立“投资决策分析模型”,则应了解相应的投资决策指标和Excel函数表达。

1. 净现值

净现值法以项目的净现值作为评价方案优劣的指标。净现值是指项目未来的现金流入按照预定贴现率折算的现值与项目未来的现金流出按照预定的贴现率折算的现值之差。这里的贴现率既可以是企业的资金成本率,也可以是企业要求的最低报酬率。项目的净现值大于零,说明该项目的报酬率大于预定的贴现率,则项目可行。

净现值函数 $NPV(\text{rate}, \text{value1}, \text{value2} \dots)$ 。其中,rate为贴现率,value1,value2...分别代表未来第1期、第2期...的期末现金流量。初始投资现金流不出现在参数中。

2. 现值指数

现值指数法以项目的现值指数作为评价方案优劣的指标。现值指数是项目未来的现金流入按照预定贴现率折算的现值与项目未来的现金流出按照预定的贴现率折算的现值之商。现值指数大于1,说明该项目的报酬率大于预定的贴现率,则项目可行。当计算项目的净现值时,只要计算未来各期的NPV与初始投资的商即可。

绝对值函数 $ABS(\text{Number})$ 。该函数用于返回给定数值的绝对值,即不带符号的数值。由于现值指数的分母为初始投资额,而一般初始现金流量为负,因此在计算现值指数时一般会用到绝对值函数,对所求数据取绝对值。

3. 内含报酬率

内含报酬率法以方案本身的内含报酬率作为评价方案优劣的指标。内含报酬率是使项目未来现金流入现值恰好等于项目未来现金流出现值的贴现率。内含报酬率是项目本身的投资报酬率,它同现值指数一样,也是一个相对指标,可以用于独立投资项目的评价。内含报酬率法在用于项目评价时不必事先选择贴现率,只要最后将资金成本率或最低报酬率与此内含报酬率进行比较,即可确定方案是否可行。而现值指数法则要求事先必须确定一个贴现率,才能确定每个项目的现值指数大小,然而贴现率的大小势必会影响方案的现值指数,进而影响对项目的评价。

内含报酬率函数 $IRR(\text{values}, \text{guess})$ 。该函数用于返回连续期间现金流量的内含报酬率。

其中，values是连续期间现金流量，guess是用户所猜想的接近IRR结果的数值，一般可省略。

values中数据的先后顺序代表了现金流量的期间顺序，因此初始投资应作为现金流出出现在values的第一个数据位置。

4. 修正内含报酬率

内含报酬率虽然考虑了时间价值，但是未考虑现金流入的再投资机会。根据再投资的假设，提出了修正内含报酬率。

修正内含报酬率函数MIRR(values,finance_rate,reinvest_rate)。该函数用于返回在考虑投资成本及现金再投资利率的情况下一系列分期现金流的内部报酬率。其中，values是连续期间的现金流量，finance_rate是现金流中使用的资金支付的利率，reinvest_rate是将现金流再投资的收益率。

【实训指引】

第一步：新建工作簿和工作表。新建一个工作簿并以“投资决策”为名进行保存，将其中的工作表Sheet1重命名为“投资决策模型”，在其中输入表题与相应的表格项目，并设置单元格格式，如图9-15所示。

	A	B	C	D
1	无风险项目投资决策			
2	期间	A项目	B项目	C项目
3	0	-100000	-200000	-300000
4	1	50000	20000	100000
5	2	40000	40000	90000
6	3	30000	60000	80000
7	4		80000	70000
8	5		100000	60000
9	6			50000
10	7			40000
11	8			30000
12	9			20000
13	10			10000
14	无风险项目投资决策模型			
15			贴现率	10%
16	评价指标	A项目	B项目	C项目
17	净现值			
18	现值指数			
19	内含报酬率			
20	修正内含报酬率			

图 9-15 输入数据并设置单元格格式

第二步：选择B17单元格，输入公式“=NPV(\$D\$15,B4:B13)+B3”，完成后按Ctrl+Enter组合键，计算A项目净现值，如图9-16所示。

❖ 特别提醒：

初始投资现金流不出现在净现值函数参数中。

第三步：选择B18单元格，输入公式“=ABS(NPV(\$D\$15,B4:B13)/B3)”，完成后按Ctrl+Enter组合键，计算A项目现值指数，如图9-17所示。

B17 : X ✓ fx =NPV(\$D\$15,B4:B13)+B3				
	A	B	C	D
1	无风险项目投资决策			
2	期间	A项目	B项目	C项目
3	0	-100000	-200000	-300000
4	1	50000	20000	100000
5	2	40000	40000	90000
6	3	30000	60000	80000
7	4		80000	70000
8	5		100000	60000
9	6			50000
10	7			40000
11	8			30000
12	9			20000
13	10			10000
14	无风险项目投资决策模型			
15			贴现率	10%
16	评价指标	A项目	B项目	C项目
17	净现值	¥1,051.84		
18	现值指数			
19	内含报酬率			
20	修正内含报酬率			

图 9-16 计算 A 项目净现值

B18 : X ✓ fx =ABS(NPV(\$D\$15,B4:B13)/B3)				
	A	B	C	D
1	无风险项目投资决策			
2	期间	A项目	B项目	C项目
3	0	-100000	-200000	-300000
4	1	50000	20000	100000
5	2	40000	40000	90000
6	3	30000	60000	80000
7	4		80000	70000
8	5		100000	60000
9	6			50000
10	7			40000
11	8			30000
12	9			20000
13	10			10000
14	无风险项目投资决策模型			
15			贴现率	10%
16	评价指标	A项目	B项目	C项目
17	净现值	¥1,051.84		
18	现值指数	1.010518407		
19	内含报酬率			
20	修正内含报酬率			

图 9-17 计算 A 项目现值指数

第四步：选择B19单元格，输入公式“=IRR(B3:B13)”，完成后按Ctrl+Enter组合键，计算内含报酬率，如图9-18所示。

B19 : X ✓ fx =IRR(B3:B13)				
	A	B	C	D
1	无风险项目投资决策			
2	期间	A项目	B项目	C项目
3	0	-100000	-200000	-300000
4	1	50000	20000	100000
5	2	40000	40000	90000
6	3	30000	60000	80000
7	4		80000	70000
8	5		100000	60000
9	6			50000
10	7			40000
11	8			30000
12	9			20000
13	10			10000
14	无风险项目投资决策模型			
15			贴现率	10%
16	评价指标	A项目	B项目	C项目
17	净现值	¥1,051.84		
18	现值指数	1.010518407		
19	内含报酬率	11%		
20	修正内含报酬率			

图 9-18 计算 A 项目内含报酬率

❖ 特别提醒：

初始投资应作为现金流出出现在内含报酬率函数values中第一个数据位置。

第五步：选择B20单元格，输入公式“=MIRR(B3:B13,\$D\$15,\$D\$15)”，完成后按Ctrl+Enter组合键，计算修正内含报酬率，如图9-19所示。

第六步：选择B17:B20单元格区域，拖动该区域的填充柄，将公式复制到C17:D20单元格区域，最终结果如图9-20所示。由图可知，应选择C项目。

中国注册会计师协会
CHINA INSTITUTE OF CERTIFIED ACCOUNTANTS
CICPA

中国注册会计师
CHINA INSTITUTE OF CERTIFIED ACCOUNTANTS

2020年编
美国注册会计师协会出版
CMA 中文教材系列

CMA

财务规划、绩效与分析

PART 1 - FINANCIAL PLANNING, PERFORMANCE AND ANALYTICS

DAVID H. STOLTZ

中国注册会计师协会
中国财政经济出版社


 中华会计网校
www.chinaacc.com
正保远程教育旗下品牌网站
美国纽交所上市公司(代码:DL)

梦想成真
系列辅导丛书

2020年版美国注册管理会计师(CMA)认证考试教材

财务规划、绩效与分析

白默 中华会计网校 编

 中国商业出版社

图书在版编目(CIP)数据

财务规划、绩效与分析 / 白默, 中华会计网校编. —北京:
中国商业出版社, 2019. 12 (2020. 9 重印)

ISBN 978-7-5208-1044-9

I. ①财… II. ①白…②中… III. ①企业管理-财务管理-
资格考试-教材 IV. ①F275

中国版本图书馆 CIP 数据核字 (2019) 第 271782 号

责任编辑: 朱丽丽

中国商业出版社出版发行
010-63180647 www.c-cbook.com
(100053 北京广安门内报国寺1号)

新华书店经销
北京市群英印刷有限公司印刷

*

787 毫米×1092 毫米 16 开 30 印张 827 千字
2019 年 12 月第 1 版 2020 年 9 月第 2 次印刷

定价: 118.00 元

* * * *

(如有印装质量问题可更换)

前 言

2011年一次偶然的机会有幸受邀在北京讲授英文CMA课程,开始与CMA考试结缘。彼时,我刚刚在南开大学会计系拿到博士学位,自恃通读过财务和管理学的英文文献,然而现实是CMA的备课和讲授过程让我倍感压力。因为它的课程体系不仅有理论深度,还非常重视贴近实务。题目中有很多情景或案例的分析,考查多个知识点的衔接和具体应用。课堂上,我也曾有过被学员追问知识点如何落地的尴尬。所谓知耻后勇,九年期间,我有幸与国家电网、北京联想股份有限公司、北汽集团等公司合作交流,并利用自己受聘天津市科技特派员的机会,深入一些制造业、服务业公司,认真了解、学习管理会计应用的现状和趋势,并开始参与一些企业具体的管理会计项目的实施。经过不断充实积累,我已经在北京、上海、南京、天津、珠海、大连等十二个城市讲授过CMA课程,教授学员超过3万人,取得了良好的教学效果,我也逐渐可以与实务管理经验丰富的学员深入交流管理会计的落地应用。我非常珍视和感激这段教学相长的经历。

2018年年末,正保教育集团邀请我撰写CMA教材。此时,我也正有把讲授CMA课程的收获、经验系统总结出来的想法。而且在与正保教育长期合作的过程中,我也感受到公司员工专业、敬业、一丝不苟的工作态度,所以我们一拍即合开始工作。期间经历了2020年开始实施的新考试大纲,对相应知识点做了仔细修改,终于成稿。

这套CMA教材的特点,或者说“差异化”之处,主要体现在以下三点:

第一,作为一本应试考试的教材,笔者的第一要务当然是帮助考生顺利通过考试。依据2020年开始实施的新考试大纲,我对所有考点进行了全面细致的梳理、对重点难点进行深入分析和总结、对经典例题加以解析、精选了章节习题帮助考生熟悉考试思路 and 模式,相信考生读后会有所受益。

第二,我将不同章节、包括P1(财务规划、绩效与分析)和P2(战略财务管理)两本教材中相关联的知识点加以提示和标注,目的不仅在于帮助读者在复习时注意知识点的前后衔接,还希望读者能够以战略和全局的角度来掌握管理会计的知识体系,真正将其学以致用。

第三,我将自己在实务领域学习和掌握的管理会计知识与考纲中的知识点相融合,希望帮助读者认识到:管理会计的知识点不是冰冷枯燥的计算公式和模型,而是能够帮助企业实现价值增值的工具。

2019年,随着中国经济增速步入“新常态”,原来一些企业的野蛮扩张模式必然遭遇更大的挑战,而向管理要效益、重视决策的科学性、使用恰当的分析工具和方法进行价值判断和预测,会被越来越多的企业所重视。因此,管理会计会有更广阔的发展空间。

在讲授CMA课程近十年之际,我把授课过程的所学、所思、所想总结出来奉于读者。“妆罢堂前问夫婿,画眉深浅入时无”,读者是最好的评委,请“孜孜焉唯进修是急”的各位来评判!

白默

目 录

第一章 外部财务报告决策	1
第一节 财务报表	2
第二节 确认、计量、估值与披露	21
课后习题	122
第二章 规划、预算和预测	132
第一节 战略管理	133
第二节 预算概念和方法	146
第三节 预测技术	152
第四节 年度利润计划及附表	160
课后习题	176
第三章 绩效管理	203
第一节 成本和差异核算	204
第二节 责任中心和报告分部	223
第三节 绩效考核	232
课后习题	239
第四章 成本管理	257
第一节 成本的计量	258
第二节 吸收成本法与变动成本法	269
第三节 成本体系	280
第四节 间接成本	308
第五节 供应链管理	327
第六节 业务流程改进	336
课后习题	341

第五章 内部控制	377
第一节 公司治理与合规性	378
第二节 系统控制与安全措施	393
课后习题	401
第六章 科技与分析	410
第一节 信息系统	411
第二节 数据管控	419
第三节 技术支持下的财务转型	425
第四节 数据分析	430
课后习题	443
附录 1 考试大纲	451
附录 2 系数表	464
参考文献	473



中华会计网校 CMA 学习交流群

第六章 科技与分析

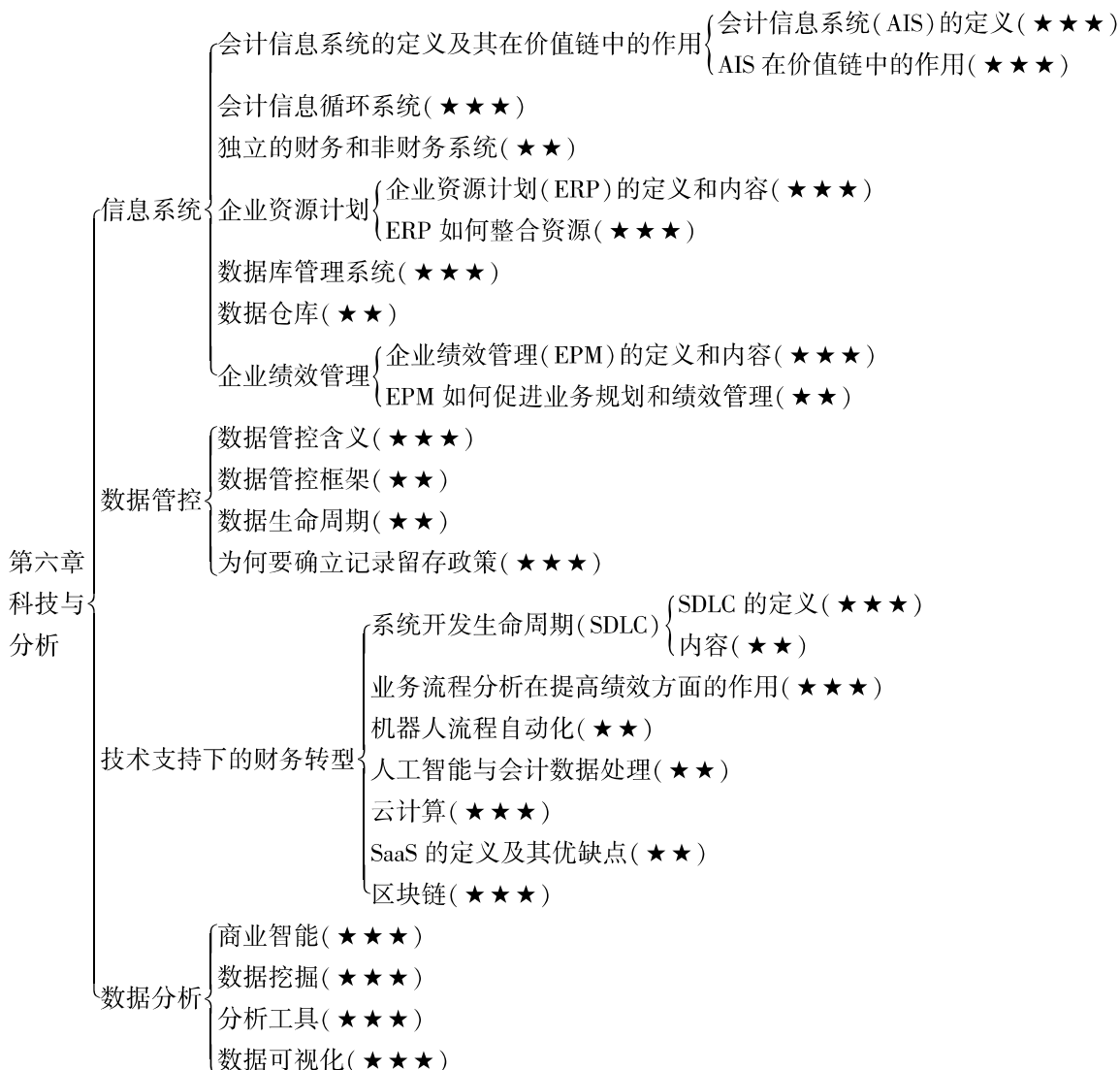
本章概述

本章探讨了科技与分析在提供和使用信息中的角色。包含信息系统、数据管控(包括 COSO 和 COBIT)的角色和使用,怎样利用科技把财务数据转换成信息,以及商业智能、数据挖掘、分析工具和数据可视化的相关知识。

本章学习指导

本章涉及的考点为基本文字表述的定性考点,包括重要概念、相关方法的优缺点等。考点主要集中于:企业资源计划、数据仓库、数据管控、数据挖掘、时间序列分析等基本概念,商业智能、分析工具等的适用范围和优缺点等。

知识框架



考情分析

(一)本章考试内容占比:15%。

(二)本章重点难点:

本章为2020年新考纲新增的章节,没有真题可供借鉴,具体的考试重点尚待观察。

(三)简答题考试方向:

企业资源计划、数据仓库、数据管控、数据挖掘、时间序列分析等基本概念;识别数据生命周期的各阶段、软件即服务的优缺点、时间序列分析的优缺点等。

第一节 信息系统

为了给决策制定者提供相关的可靠信息,公司需要一种方法来获取、处理数据,然后报告信息,这一方法被称为信息系统。本小节对运行有效的信息系统的基本要素进行论述。

一、会计信息系统(AIS)的定义及其在价值链中的角色

一个设计合理、执行有效的**会计信息系统**(Accounting Information System, AIS)的最终目的是为决策制定者提供相关和可靠的信息。AIS通过创造、记录、报告和总结公司的财务往来而达到该目的。从本质上来说,AIS最基本的模型是“投入—处理—产出—报告”。交易被创造或者投入到AIS中,它们经记录或者处理,然后被报告或者输出。AIS里的数据能够以对决策制定者最有力的方式被总结归纳。总结归纳的信息可能会像日常销售报告那样简单或者像提供给证券交易委员会(SEC)的10-K表格^①公司财务报告那样复杂。AIS的关键因素是数据的质量和可靠性,以及信息的及时性。计算机科学的谚语“无用信息输入=无用信息输出”用在这里恰到好处。为了确保只有干净的数据被输入AIS并一直保持干净,且在处理和报告的过程中是可得到的,一个设计合理的AIS会对数据进行输入控制、处理控制、输出和报告控制。

二、会计信息系统循环

设计合理的AIS会根据它们的相关循环创造、记录、报告、总结所有的财务数据或者交易。循环只是根据它们的目的组织财务往来的一个简单方法,共有七个主要循环:

- ◆收入;
- ◆支出;
- ◆生产;
- ◆人力资源和工资;
- ◆融资;
- ◆固定资产(物业、厂房和设备);
- ◆总账和记账系统。

每一个公司都会把它的交易通过循环进行分配,但并不是所有的公司都需要使用每一个循环。比如,制造业公司很可能会使用所有的循环,广告代理机构不太可能使用生产循环,因为生产循环是用来追踪制造产品的成本的。现在大多数交易几乎完全在线上完成。下文呈现的手工

^① 10-K适用于美国上市公司。在每个财政年度末的90天之内(拥有7500万美元资产以上的公司必须在60天内),向美国证券交易委员会(SEC, security and exchange commission)递交10-k表格,内容包括公司历史、结构、股票状况以及盈利等信息。

系统会帮助大家更容易理解这一过程。

(一) 收入循环

收入循环(Revenue to Cash)或销售循环是致力于处理公司销售的,是与向客户提供商品和服务以及为这些销售收取现金有关的一系列业务活动和相关的信息处理操作,见图 6-1-1 所示。主要的外部信息交换是与客户进行的。有关收入循环活动的信息也将流向其他会计循环。例如,支出和生产循环使用有关销售交易的信息来发起购买或生产其他库存以满足需求。人力资源管理/工资核算循环使用有关销售的信息来计算销售佣金和奖金。总分类账和报告功能使用收入循环产生的信息来准备财务报表和绩效报告。

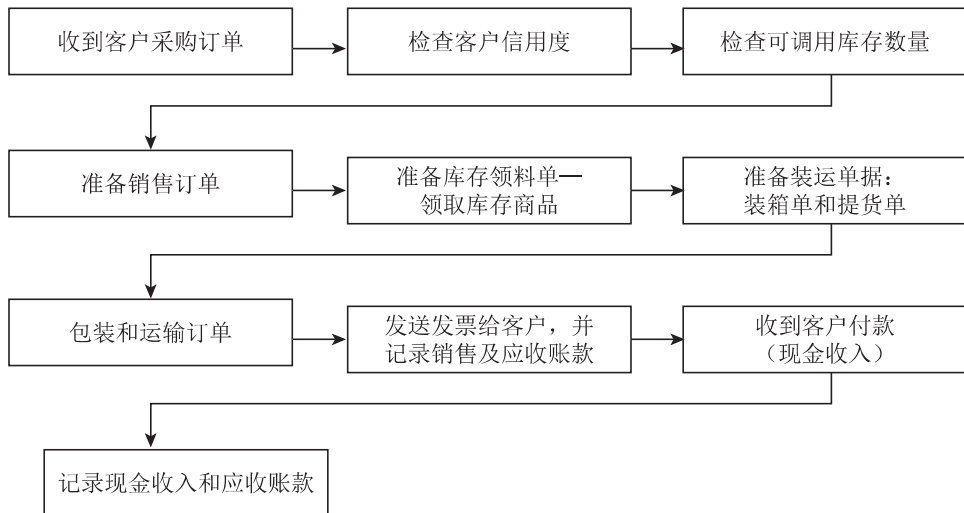


图 6-1-1 收入循环

收入循环的主要目标是在正确的时间、正确的地点、以合适的价格提供正确的产品。为了实现该目标,管理层必须做出以下六个关键决定:

- ◆ 每个产品或服务的最佳价格是多少?
- ◆ 是否应将信贷扩展到客户? 如果是,应提供哪些信用条件? 应向个人客户提供多少信用?
- ◆ 产品在多大程度上可以并且应该根据个人客户的需求和愿望进行定制?
- ◆ 应储存多少库存,应将库存放在何处?
- ◆ 应如何将商品交付给客户? 公司应该自行执行运输功能还是将其外包给专门从事物流的第三方?
- ◆ 如何处理客户付款以使现金流量最大化?

这些问题的答案指导组织如何执行图 6-1-1 所示的收入循环活动。

典型的收入或销售过程是从公司收到顾客的订购单时开始的。由于订购需要赊账,因此必须对顾客的信用状况进行评估,由公司的信用管理者决定是否批准。接下来,存货管理决定了顾客所需商品是否还有库存且有足够的数量来交付订单。然后内部文件会记录该销售订单。存货管理部门会准备一个挑选单,使得仓库的全体员工都能够领取或者挑选出顾客所订购的商品。在这些商品被挑选出来之后,货运单据,例如领料单和提货单,会和商品一起寄出。而后商品被包装送至购买者手中。公司在货物运输的同时,也会把发票寄给客户,同时,所记录的估计应收该客户的账款金额会被更新。一段时间后,顾客将会支付发票上所列的货款。公司记录收到的付款或现金,并更新应收账款中的该客户账户金额以反映付款。

(二) 支出循环

支出循环(Expenditures)致力于为公司的经营购买商品(主要是存货),不管是简单地以批发

价格购买再以零售价格转售,还是用于产品制造,它是与商品和服务的购买和付款相关的一组周期性的业务活动和相关的信息处理操作,见图 6-1-2。其他支出包括日常商业经营活动,比如临时工、咨询费、云服务、差旅费等。

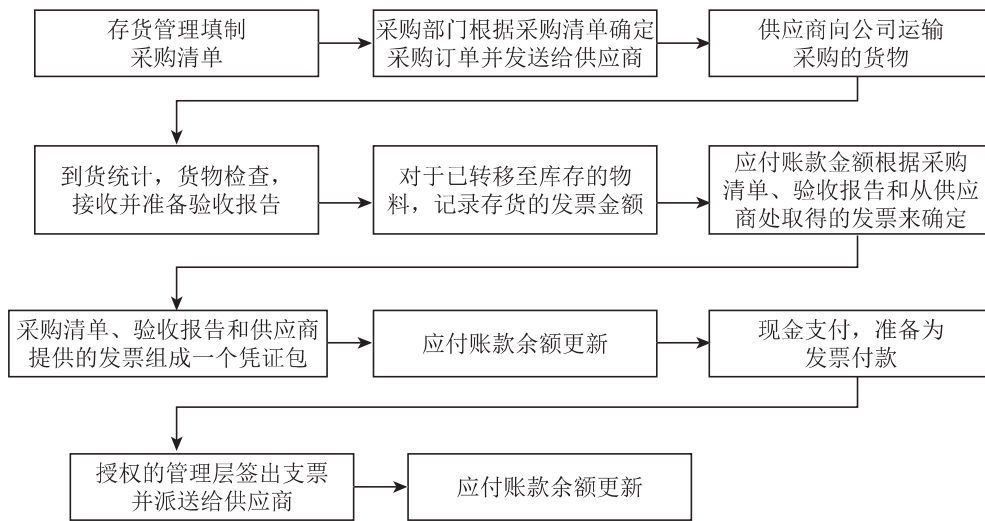


图 6-1-2 支出循环

在支出循环中,主要的外部信息交换是与供应商进行的。在组织内部,有关购买商品和材料的需求信息从收入和生循环、库存控制以及各个部门流入支出循环。货物和物料到达后,其收货通知会从支出循环流回这些来源。支出数据也从支出循环流向总账和报告程序,以包含在财务报表和各种管理报告中。

支出循环的主要目标是最大限度地减少获取和维护库存、供应品以及组织运行所需要的各种服务的总成本。为了实现此目标,管理层必须做出以下关键决策:

- ◆最佳库存和供应量是多少?
- ◆哪些供应商以最优惠的价格提供最好的质量和服务?
- ◆组织如何合并各个部门的采购以获得最佳价格?
- ◆如何使用信息技术来提高入库物流功能的效率和准确性?
- ◆组织如何保持足够的现金来利用供应商提供的折扣?
- ◆如何管理对供应商的付款以使现金流最大化?

这些问题的答案指导组织如何执行图 6-1-2 中所示的支出循环活动。

典型的支出过程是从存货控制确定需求(库存不足)并准备采购申请以满足需求开始的。请购单是基于采购需求所准备的,请购单发送到采购部门。采购订单发送给接受订单并把货物运到公司的供应商。供应商通常会在发货后不久开具发票。供应商所发的货物在接收部门接收,接收部门对收到的货物进行清点、检查、接收。

接收部门会准备验收报告以记录收到的物品。然后将商品运到仓库进行存储。库存控制随着物料的到达更新库存记录。在这一过程中,应付账款汇集了这三个文件:

- ◆因采购编制的采购单
- ◆因接收编制的验收报告
- ◆从供应商处取得的发票

这三个文件的组合构成了一个凭证包,这个凭证包提供了在付款之前必须给供应商的三个必不可少的文件:订单证明(采购单)、收据证明(接收报告)和账单证明(发票)。应付账款随着

采购更新,并且准备现金支出,将款项支付给供应商。通常是用经授权的签字者签发的支票(纸质的或电子的)进行支付。在支票发出之后,应付账款随着付款进行更新。

(三)经营(生产)循环

经营循环(Production)是与产品制造相关的一组周期性的业务活动和相关信息的处理操作,经营(或生产)循环因公司和产品的不同而不同,然而,就产品循环而言,存在一种可以使用的会计数据输入、处理、输出的模式。通常,在经营周期有四个主要的活动:

- ◆ 产品设计
- ◆ 计划和安排
- ◆ 生产经营
- ◆ 成本管理

如图 6-1-3 所示,每一活动都有输入、处理和输出。

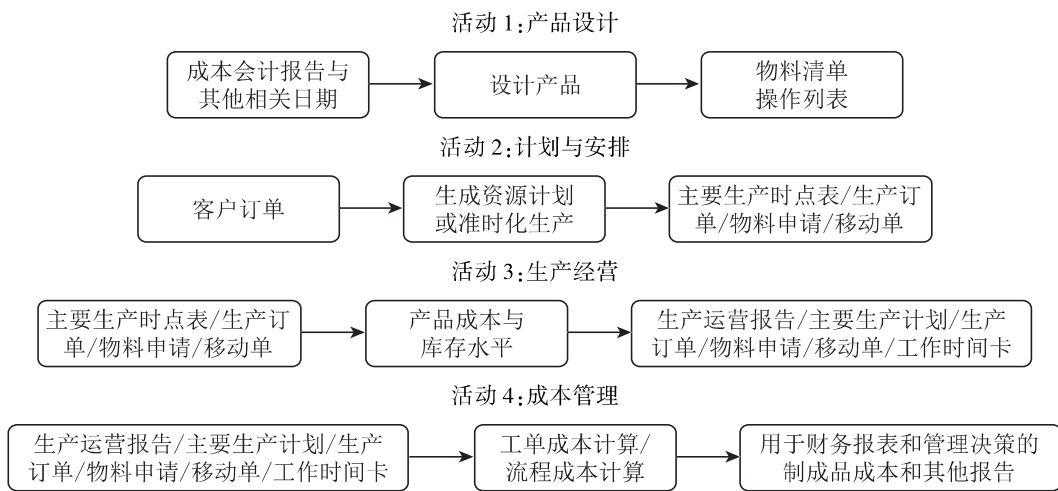


图 6-1-3 经营循环

活动 1:产品设计将成本管理报告作为输入,进入到除其他相关资源之外的设计过程。成本信息有助于指导产品的设计,以有效地反映不同设计的成本。输出包括物料清单(产品原料)和操作列表(产品配方)。

活动 2:计划和安排把客户订单作为输入,以决定是否使用物料需求计划(MRP)或精益生产。具体而言,MRP 用客户订单预测来决定何时生产、生产多少。精益生产(Lean Manufacturing, 或译为“及时生产”“准时化生产”)计划和计划产量是建立在实际的客户订单的基础上的,这就意味着库存是已被订购的,且进行及时生产以供应客户订单。该活动的输出包括主生产计划、生产订单(授权实际开始制造产品)、材料申请从库存中提取必要的物料,并转移票据(授权将产品从一个阶段转移到下一个阶段)。

活动 3:生产经营使用来自活动 2 制造产品过程的输出文件。此项活动最主要的会计职能是查明生产成本、追踪相关的存货水平。活动 3 的输出是经过修订和更新的输入文件,以及用于计算生产过程中所用的劳动力成本的工作—时间卡。

活动 4:成本管理包括活动 3 的输出文档。流程中所使用的这些文档中的数据,是用工作订单成本核算还是用流程成本核算方法来确定产品成本,取决于所生产的物品的性质。最终输出的是商品制造结算单的成本以及其他在财务报告和管理决策制定中所使用的相关的必要报告。

尽管会计人员主要参与第四步成本管理,但他们必须了解其他三个过程,才能设计出能够向管理层提供管理现代化制造公司生产周期活动所需信息的报告。

(四) 人力资源和工资循环

人力资源和工资循环 (Human Resources and Payroll) 是与有效管理员工劳动力相关的一系列业务活动和相关数据处理操作。

人力资源循环的主要作用是确保企业拥有相应技能的人才来履行其使命。该流程的关键步骤是聘用、培训、转岗、解雇员工。此外,人力资源循环是提供职工福利、保持良好的职业道德、维持安全、干净、多产的工作环境的重点。

工资循环的主要功能是为员工所完成的工作提供相应的报酬并注意与工资相关的税费。工资循环的五个主要步骤是:

- 第一,更新重要的工资数据;
- 第二,记录时间数据;
- 第三,准备工资单;
- 第四,发放工资;
- 第五,支付税款和其他扣除额。

图 6-1-4 展示了每小时员工的工资核算流程,该流程开始于最新的、准确无误的主文件。主文件随着来自人力资源部门、政府机构、保险公司和员工的与工资相关的数据的更新而更新。人力资源部门更新有关新员工、升职、转岗、解雇的主文件,政府机构提供税率数据,保险公司提供该公司的评估保险费率的的数据,雇员提供最新的扣除额,例如由于其家庭变更而要求扣除的数额,如婴儿的出生或员工希望增加退休计划账户的金额。

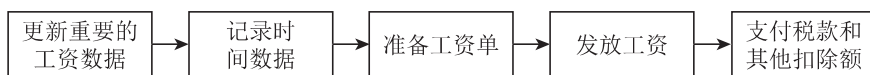


图 6-1-4 工资单流程

(五) 融资循环

融资循环 (Financing) 是与通过向投资者出售公司的股份、借钱以及支付股息和利息筹集资金有关的活动。

(六) 物业、厂房和设备循环

物业、厂房和设备循环 (Property, Plant, and Equipment) 包括三个主要的步骤,如图 6-1-5 所示:

- 第一,资产的取得;
- 第二,资产的折旧与维护;
- 第三,资产处置。如图 6-1-5 所示。

物业、厂房和设备通常意味着公司的大量投资,且将使用多年。因此,收购或购买这些资产需要进行仔细规划和分析。公司取得资产后,对其进行维护并计提折旧。为满足财务和管理会计报告需求,企业要在几种适当的折旧方法中选择其一对资产的使用情况进行估计(土地除外,因为它不是可折旧资产),比如直线法、双倍余额递减法、工作量法。在固定资产完成其服务目的之后,它将被以出售或者报废的形式进行处置。其账面净值与市场价值的差额将根据具体情况记入收益或损失。有时,资产可能会遭受意外损失,这也应该记录下来。

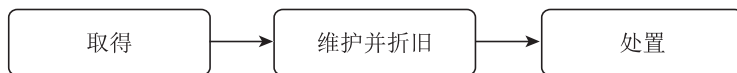


图 6-1-5 物业、厂房和设备循环

(七)总账和记账程序

总账和记账程序(General Ledger)包含四个主要步骤,如图 6-1-6 所示:

- 第一,更新总账;
- 第二,传递调整分录;
- 第三,编制财务报表;
- 第四,编制管理报告。

依据会计分类账和日记账以及公司财务部门的日记账分录更新总账。调整分录由公司财务主管传递到总账。公司的财务报表根据总账中的数据进行编制。管理报告如预算和财务执行情况报告,也是根据总账中的数据编制的。

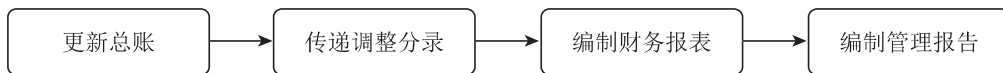


图 6-1-6 总账和记账程序

三、独立的财务和非财务系统

实施**独立的财务和非财务系统**(Separate Financial and Nonfinancial Systems)的主要挑战是数据维护。财务和非财务系统记录、追踪和报告业务的健康经营情况,两者的区别在于一个使用财务指标,另一个使用非财务指标。换言之,财务和非财务指标是用不同的手段衡量相同的事物,并把用非财务指标衡量的经营活动反映出来。独立的财务和非财务系统的问题之一在于确保数据在两个系统中准确连接。当两个系统分离时,数据应该是可比的,以确保两个系统用不同的指标所衡量的是相同的事物。如果系统所利用的数据不是在同一个地方或数据库中,必须建立并保持广泛的控制,使得使用者相信数据的质量和可靠性。例如,记录一项销售时,财务系统只是简单地记录销售额、现金收据或应收账款,而并不记录其他关于该销售的任何信息,比如所涉及的销售人员、具体的时间点、天气、顾客是一个人还是一群人、顾客询问产品相关信息的次数等。如图 6-1-7。

因此,许多组织开发了其他信息系统来收集、处理、存储和报告会计信息系统中未包含的信息。不幸的是,多个系统的存在产生了许多问题并导致效率低下。通常,同一数据必须由多个系统捕获和存储,这不仅会导致跨系统冗余,而且如果只在一个系统中更改数据而在其他系统中未更改数据,则可能导致不一致。此外,很难集成来自各个系统的数据。

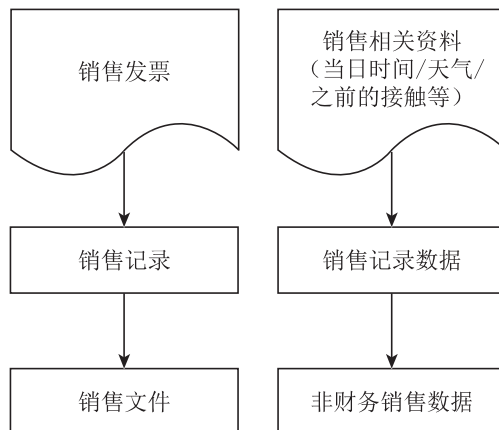


图 6-1-7 ERP 被应用之前的流程

四、企业资源计划(ERP)

在关系数据库和廉价的计算机内存问世之前,除了创建单独的程序和文件来控制和维护数据之外,企业别无选择。例如,公司会创建一个程序和数据文件用于日常销售,创建另外一个程序和数据文件用于记录应收账款。这两个文件必须定期核对,以确保数据的准确性。最好的方法难道不是把公司所有的数据存储在一个数据库里面吗?关系数据库的发明和电脑内存成本的降低促进了**企业资源计划**(Enterprise Resource Planning, ERP)系统的引进。ERP系统是一个将组织活动的各个方面(例如会计,财务,营销,人力资源,制造,库存管理)集成到一个系统中的系统,它将公司所有的数据存储到一个中央数据库中。此外,ERP系统存储了从员工健康计划、设备维护记录、客户电话号码、资金支出计划到营销活动,以及介于这两者之间的与公司各个方面相关的非财务信息。然而,这是企业资源计划系统的目标,它并不是总能实现的。企业资源计划系统被分为几个典型的商业循环,如图6-1-8所示。

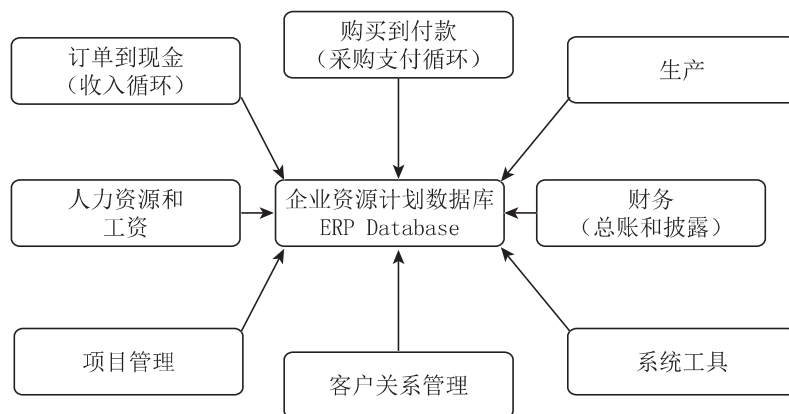


图 6-1-8 企业资源计划(ERP)

企业资源计划主要包括:

- ◆ 订单到现金(收入);
- ◆ 购买到付款(支出);
- ◆ 生产;
- ◆ 人力资源和工资;
- ◆ 财务;
- ◆ 项目管理;
- ◆ 客户关系管理;
- ◆ 系统工具。

订单到现金(收入)循环处理订单录入、运输、存货、现金收据和佣金。

购买到付款(支出)循环处理购买、已收货物、库存和仓库管理、非库存和现金支付。

生产循环处理生产调度、物料清单、物料需求计划、工程、在制品、质量控制、成本管理和服务公司的运营。

人力资源和工资循环维护和处理人力资源、工资、福利、时间记录、培训和教育以及政府报告。

财务循环(总账和报告)维护和处理总账、应收账款、固定资产、现金管理、应付账款、预算、财务报表以及其他报告。

项目管理循环记录并维护账单、成本、时间、完成情况、费用以及活动管理。

客户关系管理循环维护和支持销售、营销、服务、联系客户、佣金以及呼叫中心支持。此外,公司尽可能地通过分析大数据和物联网来预测客户需求。

为了建立和维护主文件数据,还有系统工具、数据流、评估控制等。

五、企业资源计划如何整合资源

企业资源计划如何帮助组织克服分离财务和非财务系统面临的挑战,并将其各个方面的作业进行整合?

企业资源计划(ERP)系统克服了财务和非财务系统分离所带来的问题,因为它们将公司运营的各个方面与传统的会计信息系统集成在一起。大多数大型和中型组织都使用 ERP 系统来协调和管理其数据、业务流程和资源。ERP 系统收集、处理和存储数据,并向信息经理和外部人员提供评估公司所需的信息。

正确配置的 ERP 系统使用集中式数据库在整个业务流程中共享信息并协调活动。这很重要,因为作为一个业务流程一部分的活动通常会触发组织中许多不同部分的一系列复杂活动。例如,客户订单可能需要安排其他生产以满足不断增长的需求。这可能会引发订购更多原材料的订单,也可能有必要安排加班时间或雇用临时工。精心设计的 ERP 系统使管理层可以轻松访问所有有关这些活动的最新信息,以便更有效地计划、控制和评估组织的业务流程。

六、数据库管理系统(DBMS)

如图 6-1-9 所示,数据库管理系统(Database Management System, DBMS)是介于数据库和访问它的应用程序之间的接口或程序。数据库、DBMS 和通过 DBMS 访问数据库的应用程序称为数据库系统。数据库管理员(DBA)负责协调、控制和管理数据库。数据库管理系统有助于创建、检索、更新、管理、保护数据,它控制两个主要的组件:

- ◆数据;
- ◆允许访问、检索、修改以及必要时锁定的数据库程序。

数据库规划或蓝图定义了数据的逻辑结构,或者是人类查看数据的方式。数据库管理系统提供了数据集中化的视角,使得各个地方不同的使用者都可以访问,该系统还可以限制哪些数据可以被特定的某一个或某一些用户访问、检索、修改。数据库管理系统规划使得使用者无须知道数据实际位于何处也可以访问数据库。

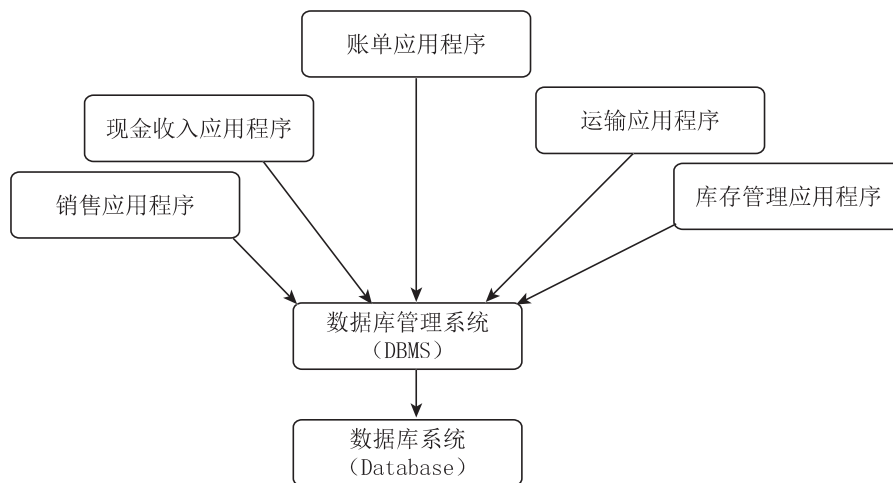


图 6-1-9 数据库管理系统 (DBMS)

七、数据仓库

在当今快速发展的全球经济中,管理层必须根据战略目标不断重新评估财务和运营绩效,并根据需要快速更改计划。由于战略决策需要访问大量历史数据,因此组织正在建立单独的数据仓库,称为**数据仓库(Data Warehouse)**。数据仓库是一组由详细汇总的数据组成的大型数据仓库,主要用于分析而不是处理交易。它本质上是从不同的程序、资源和数据库检索到的公司所有数据的存储库或存储位置。这些数据通常是干净且有组织的,使得它能够被搜索到。

八、企业绩效管理(EPM)

企业绩效管理(Enterprise Performance Management, EPM)也称为公司绩效管理或业务绩效管理,包括监控和评估经营业绩。企业绩效管理(EPM)与以下四个方面相联系:

- ◆ 商业智能和 ERP;
- ◆ 预算和预测;
- ◆ 财务报告;
- ◆ 分析绩效结果并确定改进方法。

九、EPM 如何促进业务规划与绩效管理

ERP 系统是帮助公司进行日常经营管理的,而 EPM 是通过分析、理解、报告对业务进行管理。尽管 EPM 已经存在了几十年,但它的方法论作为工具变得越来越复杂,EPM 的软件也不断改进和发展。在最早阶段,EPM 只是简单地包括面对面会议和电话。第一个 EPM 应用软件专注于收集和提供会计、预算和财务绩效信息。电脑自动生成的电子表格的出现使得繁琐的手动创建的电子表格逐渐消失,有助于更具战略性地制订计划、更好地编制预算和改进报告。后来,EPM 专用软件包被开发出来,使得财务和会计部门很多财务合并和报告的职责变得自动化。基于 Windows 的客户端/服务器系统已经让位于基于网络的程序。软件即服务(SaaS)应用程序已经被广泛应用,使员工能够更专注于更高层次的战略任务,而不是管理与 IT 相关的问题。

第二节 数据管控

本节从定义、框架、生命周期、保留政策和保护几方面来讨论数据管控。

一、数据管控

数据管控(Data Governance)由组织内全部的管理数据组成。主要涉及数据可用性、使用性、全面性和安全性的管理,数据管控非常重要,因为组织的数据具有内在价值。如果没有设计合理且运行良好的数据管控程序,数据可能会被破坏、贬值、变得无法使用、丢失甚至被盗。数据管控计划应该包括监督机构,一套程序和控制以及一套实施程序和控制的政策或指令。

在数据管控计划的实施阶段,应该选择和培训数据管理员,由他们担任指定数据看管人的角色。数据管理员应该对数据的可用性、使用性、全面性和安全性负责。同时应建立各种控制帮助管理员履行职责。输入、处理和输出控制辅助数据管理员维护数据质量。输入控制包括数据输入控制,例如适当的数据输入界面或表单设计、现场检查、限制检查、完整性检查、有效性检查,以及包括数据匹配、恰当的文件标签、交叉平衡测试的批处理总计处理控制、同步更新控制。输出

控制包括用户对输出、对账、数据传递控制的审核。数据可用性应包括适当的容错能力和内置于信息系统内的冗余、不间断电源供应和备份生成器、备份和测试备份过程以及实时镜像。通过适当的职责分离,数据变更管理和授权结构以及独立的检查和审计可以保持数据的完整性。

通过使用深度防御的方法有助于实现数据安全,包括在整个组织内实施不同层次的数据安全控制。换言之,就是不仅仅依靠锁住前门,而是锁住所有的办公室和衣柜门以防肇事者穿透前门。此外,数据安全取决于员工在适当的数据安全程序方面的培训、身份验证控制、授权控制,如访问控制矩阵、防火墙和其他网络安全工具、数据加密和补丁管理。如果员工意外(或故意)使数据暴露,世界上最好的控制也会被损坏。当然,不管在安全防护上花多少钱,都不能完全保护数据免受黑客入侵。这是数据风险管理发挥作用的地方。

二、数据管控框架

两个主要的数据管控框架通常被会计专业人员所使用。一个是由发起组织委员会(COSO)制定的,另一个是由国际信息系统审计协会(ISACA)制定的。COSO 框架处理一般的数据管控,而 ISACA 框架的重点是处理与信息技术(IT),特别是信息及相关技术控制目标(Control Objectives for Information and Related Technologies, COBIT)框架有关的数据管理。

COSO 发布了三个版本:COSO 内部控制—综合框架(COSO-ICIF),COSO 企业风险管理(ERM,2004),COSO 内部控制—综合框架(2013)。COSO-ICIF 于 1992 年出版,帮助专业人士设计和实施有效的内部控制。这些框架打破了设计有效的内部控制的任务,进入五个重点领域:控制环境、风险评估、控制活动、信息与沟通以及监控。此外,这些框架帮助用户解决在单位或活动中运营、财务报告、依赖程度方面的内部控制问题,如图 6-2-1 所示。

(一)COSO 内控综合要素框架

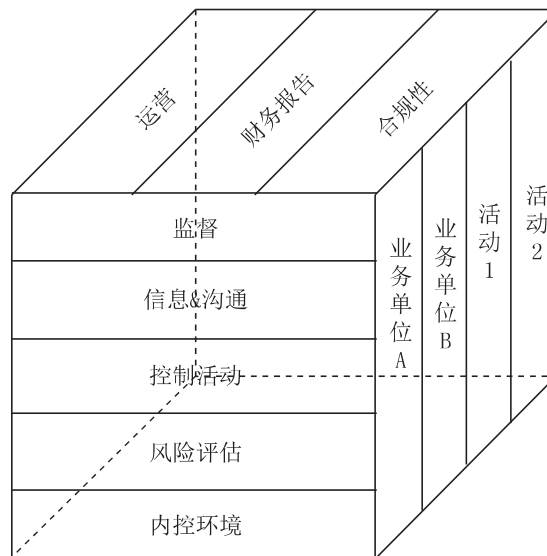


图 6-2-1 COSO 内控综合要素框架图

1.控制环境

内部控制的环境部分识别人们经营的公司以及为自己的公司创造的文化或环境。控制环境解决诸如顶层风格之类的因素(比如,高层管理者对有效的内部控制的态度、管理人员和雇员的忠诚度及能力、管理风格)。例如,通过组织对数据的管理来评估尽职程度。

2. 风险评估

风险评估部分解决每个实体都会面临来自外部和内部的风险这一事实。为了解决风险,必须识别和评估首要风险是如何影响公司目标的。然后决定如何控制风险以实现公司目标。例如,识别数据不同的组合方式,每种方式可能的结果以及预计发生的可能性是非常重要的。这些步骤将有助于优先调整数据管理改进计划。

3. 控制活动

控制活动部分包含用于实现目标和妥善管理风险的政策和程序。例如,员工需要多久更改一次密码?关于员工笔记本电脑中敏感信息的政策是什么?

4. 信息与沟通

信息与沟通部分涉及管理和控制公司运营所需信息的获取、处理和报告。例如,为员工提供有关数据安全实践的定期培训。

5. 监控

监控部分涉及检查以确保流程正常工作、识别不足之处、报告这些不足之处以使得它们能够被纠正,使整个系统得到改善。一种识别数据管理流程和安全缺陷的方法是聘请咨询公司执行渗透测试(后续会详细讨论)。

(二) COSO 内部控制—综合框架 2013

2013年,COSO发布了1992年框架的修订版。原来的立方体仍保持完整,但通过增加17条原则进行改进,以使其更好地应对当前的商业实践和技术进步。下面对17条原则进行讨论。

1. 控制环境

控制环境是内部控制的所有其他部分的基础。任何企业的核心都是员工——他们的个人属性,包括诚信、纪律、道德价值观、能力和经营环境。它是驱动组织和一切赖以生存的基础的引擎。

(1) 组织表现出对诚信和道德价值观的承诺。例如,它通过建立行为准则或根据被报告的违反行为准则的行为来创建流程。

(2) 董事会通过设立董事角色和职责、董事会会议的政策和实践、审查管理层的判断等方式,独立于管理层,并监督内部控制的发展和绩效。

(3) 在实现目标的过程中,管理层通过为不同的管理职能定义角色、汇报关系以及权限级别,建立了董事会监督结构、汇报关系以及适当的权力和责任。

(4) 组织承诺根据目标吸引、培养、留住有能力的个人。例如,它创建所需的技能和专业知识水平、评估能力,并选择合适的服务提供者。

(5) 在追求目标的过程中,组织要求个人对自己的内部控制责任负责,方法是明确界定责任和业绩衡量标准,并将薪酬与业绩挂钩。

2. 风险评估

组织必须识别、分析、管理其风险。风险管理是一个动态过程,管理层必须考虑可能会阻碍其目标实现的外部环境和业务内部变化。

(6) 组织明确规定目标,以使通过明确识别财务报表账户、披露、判断、估计能够识别和评估与目标相关的风险。

(7) 组织在整个实体中识别实现其目标的风险,并分析风险,将其作为应对风险的基础。这些活动包括建立一个正式的风险识别流程、与人员会面、评估风险的可能性,以及评估他们对风险的反应。

(8) 组织考虑在评估风险时可能存在的欺诈行为。例如,组织可以执行恶意欺诈风险评估

并考虑可以规避控制措施的方法。

(9) 组织确定和评估会对系统内部控制产生重大影响的变化。例如,组织可能会评估其外部环境的任何变化或高级管理团队的变动。

3. 控制活动

控制策略和程序有助于确保管理层确定的应对风险和实现组织目标的措施得到有效执行。控制活动在业务流程和技术的各个级别和各个阶段执行。

(10) 组织选择和开发有助于把实现目标的风险降低到可接受水平的控制活动。一个方法是将控制活动反映到已识别的风险或实施更仔细的外包功能的监控。

(11) 组织选择和开发有关科技的一般控制活动以支持目标的实现。例如,评估终端用户计算、管理安全性和访问控制,或设置 IT 基础架构根据功能角色支持各种级别的 IT 访问。

(12) 组织通过政策部署控制活动,这些政策通过制定与记录政策和程序,以及对控制活动进行定期评估,来确定预期的内容和将政策付诸实施的程序。

4. 信息与沟通

信息与沟通系统捕获并交换沟通管理和控制组织运营所需的信息。必须在内部和外部进行沟通,以提供执行日常内部控制活动所需的信息。所有人员必须了解他们的责任。

(13) 组织通过创建信息需求清单,从外部资源或与财务无关的经理处获取信息,或者通过数据管理程序提高信息质量,从而获得或生成并使用相关的质量信息,以支持内部控制的功能。

(14) 组织内部沟通信息,包括目标和内部控制的责任,支持内部控制的运作。例如,组织应该沟通内部控制责任、制定与董事会沟通的政策,或建立举报制度。

(15) 组织通过审查外部审计委员会或调查相关的外部各方,就影响内部控制运行的事项与外部各方进行沟通。

5. 监控活动

监控必须监视整个过程,并根据需要进行修改,以便系统可以根据条件进行更改。评估确定内部控制的每个组成部分是否存在并起作用,及时沟通缺陷,并将严重问题报告给高级管理层和董事会。

(16) 组织选择、开发和执行持续的和/或单独的评估用以确定是否存在内部控制的要素和功能。例如,组织可以建立适当的基线,识别和使用激励指标,或者派内部审计师监控组织的运作。

(17) 组织评估并及时与负责采取纠正措施的各方(包括高层管理者和董事会)沟通内部控制缺陷。例如,评估和报告缺陷并监控纠正措施。

(三) 信息及控制技术控制目标(COBIT)

COBIT 专注于与 IT 相关的有效的内部控制。COBIT 框架提供了与 IT 相关的有效管理对内部控制的最佳实践。COBIT 将来自许多不同来源的控制标准整合到一个单一的框架中,该框架允许:

1. 管理层对 IT 环境的安全性和控制实践进行基准测试;
2. 确保用户存在足够的 IT 安全性和控制权;
3. 审核员来证实他们的内部控制意见,并就 IT 安全和控制事宜提供建议。

这是一套庞大且详细的手册,用于创造、执行、维护与 IT 相关的控制。该框架将 IT 分为四个主要部分:计划和组织、获取与实施、交付与支持、监督和评估。然后,这四个主要部分又分解为 32 个管理控制流程。

三、数据生命周期

生命周期由多个阶段组成,每个阶段都有自己的特征。虽然对数据生命周期所包含的阶段的实际数量存在争议,但是这八个阶段代表了关于数据生命周期的一般观点:数据捕捉、数据维护、数据合成、数据使用、数据分析、数据发布、数据存档和数据清除,如图 6-2-2 所示。

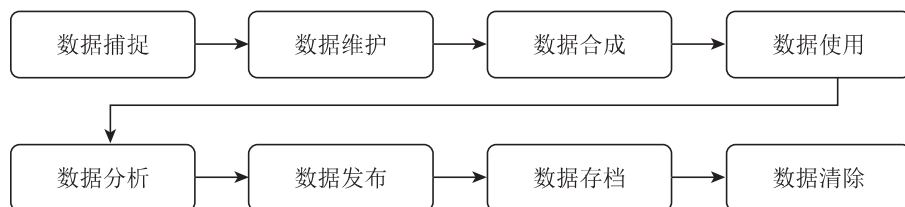


图 6-2-2 数据生命周期

(一)数据捕捉(Data Capture)

为了进行分析,首先必须记录或捕捉数据,这是创建组织内尚不存在的信息的行为。被捕捉的数据可能来自公司的外部或内部。今天的物联网产生了大量可能被捕捉到的数据。

捕捉数据的主要方式包括:

1. 利用数据

从组织外部获取已经存在的数据。例如,可以从外部实体购买或以其他方式获取数据。

2. 数据输入

可以在组织内创建新的数据值。

3. 信号接收

获取组织内的控制系统已经创建的数据。

(二)数据维护(Data Maintenance)

捕捉数据后,生命周期的第二阶段可以定义为向发生数据合成和数据使用的点提供数据。

数据维护涉及在不为企业获得任何价值的情况下处理数据,它通常涉及数据的清理和充实。为了有用,必须将数据转换为可用的形式。创造有用数据的过程可能包括清洗、擦洗以及通过“提取—转换—加载”(ETL)的方法进行处理。公司使用企业资源计划(ERP)系统和其他不太复杂的信息系统满足其信息需求。ERP 和其他信息系统使用数据库技术来组织和查询数据。为了使数据能够被加载到数据库中,数据必须被清洗和擦洗掉多余的字符和符号。此外,必须对其进行检查以确保日期数据填充日期字段,数字数据填充数字字段,字符数据填充字符字段。从本质上来讲,数据清洗和擦洗将非结构化的数据转换为可在组织的信息系统中使用的结构化数据。

(三)数据合成(Data Synthesis)

数据合成涉及统计方法的使用,这些方法结合了许多数据来源或测试,以获得更好的总体估计或被问到的有关数据问题的答案。一些人把这个称为数据建模或使用归纳推理以转换数据。另一些人将数据合成视为数据维护的子集。

(四)数据使用(Data Usage)

数据使用就是如何使用数据来支持业务的任务,此阶段可以定义为将数据应用到企业需要自行运行和管理的任务。如战略规划、客户关系管理(CRM)、处理发票、向供应商发送订购单等。

(五)数据分析(Data Analytics)

数据分析是以创建新的信息和产生业务洞察力为目的检查原始数据的科学性。它包含技能、科学、迭代探索和对过去业务绩效的调查,以获得洞察力并推动未来的业务规划。在它最基

本的层次,它意味着用数据分析的方法来回答问题。一些人将数据分析视为数据使用的子集。

(六) 数据发布(Data Publication)

数据发布是向组织外部发送数据的行为,可以发布数据以供广泛使用,通常将数据发送到业务部门。例如向客户发送结算单。

(七) 数据存档(Data Archival)

单个数据值可能会使用相当长的一段时间,但其生命周期最终都会到尽头。那时,组织应该将数据值存档。数据存档是从有效使用的数据中删除数据,并将其存储以备潜在的未来使用的过程,数据存档是存储数据的位置,但是它没有定期维护,且使用率可能很少。

(八) 数据清除(Data Purging)

数据清除涉及删除不再有用或不需要的数据,在数据生命周期结束时,将从企业中删除数据项的每个副本,这通常是从归档中完成的。管理者应创建数据保留政策以实施正确的数据清除实践。例如,美国国税局保留纳税申报表数据参数以备审计。对于内部数据,例如,清除 10 年里尚未被访问的数据。

四、确立记录留存政策

(一) 对于每个组织而言,有一个**正式的记录留存政策**(Documented Records Retention Policy)非常重要,它有时也被称为**记录管理政策**(Documented Records Management Policy),该政策规定了文件和其他记录的保留和定期销毁。

(二) 保持和维护太多的记录或者将它们存储得比需要的时间长,会为组织带来不必要的成本。

(三) 对于某些类型的文件,法律规定了最短的保留期限,例如与税收和雇员有关的文件,并且某些行业可能会受到特定法规的进一步管辖。甚至最低法定留存期有时也有不确定的描述。例如,税收记录必须至少保留 7 年,除非有可疑的欺诈行为,在这种情况下,没有最低记录期限。因此,有时最好保留纳税记录比最短期限更长,以防纳税人在未来某个不可预见的时刻需要证明没有欺诈行为。

(四) 如果某个特定法律案件所需的记录已按照常规留存政策销毁,则法院通常会假定该组织已履行其职责。但是,正如安然事件所表明的那样,一旦有可能提起诉讼,现在就制定记录留存政策或开始遵守过去未曾遵循的政策已为时已晚。

五、网络攻击检测和预防

所有与互联网有联系的组织都有遭受网络攻击的风险。恶意黑客使用各种方法和策略进行网络攻击,从钓鱼邮件、拒绝服务、SQL 注入到零日攻击。许多网络专家断言,这不是一个公司是否会受到攻击的问题,而是何时会受到攻击的问题。

公司如何保护自己免受这些攻击?一种方法是通过**渗透测试**(Penetration Testing)来确定公司何处容易受到攻击。渗透测试通常聘请专门识别和评估恶意攻击的易受攻击点的外部咨询公司。渗透测试有可能被视为对公司信息系统的经授权的攻击。咨询顾问试图采用一切可能的、能够想到的手段损害公司的系统,他们几乎总能成功地找到进入和可利用的弱点。不幸的是,没有人能够使用每种可能的方法来识别易受攻击点。黑客一直在寻找新的方法。

另一种检测和预防网络攻击的方法是通过**生物识别鉴定**(Biometrics)。每个信息系统用户都应该在系统中对自己进行身份验证。有三种主要的认证方法,即(1)你知道的东西,如密码;(2)你拥有的东西,如身份证;(3)你是谁,如你的指纹。前两种方法(密码和身份证)的问题在于

它们可能被遗忘、丢失或被盗。而生物识别很难丢失或被盗。因此,通过使用生物身份验证方法访问公司的信息系统可能会受到更严格的限制。

高级防火墙(Advanced Firewalls)提供了另外一种检测和预防网络攻击的方法。防火墙使用深度防御的概念,即在连续多层设置安全功能,因此,如果某个级别的安全性被攻破,它后面还有别的层来阻止攻击。这就好比一座大楼的前门是锁住的,大楼里的办公室也是锁住的,每个办公室里都有一个上锁的文件柜来存放敏感数据。除了深度防御之外,防火墙还使用数据包过滤和深度数据包过滤的方法来识别和阻止恶意网络攻击。数据包过滤检查发送到公司网络的数据包的源地址,并将其与恶意数据包来源清单进行比较。如果该地址在清单上,它不会被允许进入公司的信息系统。深度数据包过滤在允许数据包进入公司的信息系统之前,会扫描数据包的内容以查找是否含有恶意代码。

第三节 技术支持下的财务转型

本节讨论技术支持下的财务转型——换言之,就是技术如何使公司的财务运作地更快、更好、更有效。本节对系统开发生命周期(SDLC)、业务流程分析、机器人流程自动信息(RPA)、人工智能(AI)、云计算、软件即服务(SaaS)和区块链等主题进行讨论。

一、什么是系统开发生命周期(SDLC)

系统开发生命周期(The Systems Development Life Cycle,SDLC)是设计和实施新的信息系统的结构化路线图。SDLC方法是应用于大型、高度结构化的应用程序系统开发的传统方法,该方法的主要优点是增强了对开发过程的管理和控制。尽管SDLC有许多版本和变体,但在图6-3-1中给出了一个基本的五步法。

该方法可以扩展到包括几乎所有在文献中可以找到的变异。这五个步骤是:

- 第一,系统分析;
- 第二,概念设计;
- 第三,实体设计;
- 第四,实施和转换;
- 第五,运营和维护。

系统分析包括识别组织需求,收集关于修改当前系统、购买新系统以及开发新系统的信息。

概念设计包括为满足组织的需要创建计划。准备设计备选方案并创建详细的规范,从而为如何完成期望的系统提供指导。

实体设计包括概念设计以及为创建系统而建立的详细规范。该设计包括计算机代码规范、输入、输出、数据文件和数据库、过程和程序以及适当的控制。

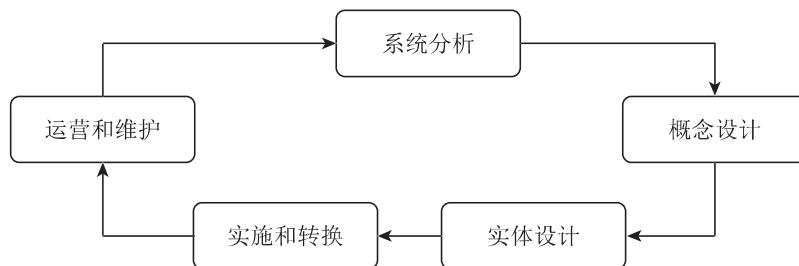


图 6-3-1 系统开发生命周期

实施和转换包括新系统的安装(硬件和软件)。新系统经过测试并对用户进行培训,且制定新的标准、程序和控制。

运营和维护包括运行系统、检查性能、根据需要进行调整以及维护系统。组织对系统进行改进和修复,直到确定维持旧系统不能使其获益,整个循环又重新开始。

二、业务流程分析在提高系统绩效方面的作用

由于更优质的信息有利于更好的决策,绩效系统的质量直接影响公司的整体表现。因此,公司总想找到改进的方法——即把事情做得更好、更快、成本更低,或者根本不做,以提高盈利能力。通过提高系统绩效的方法可以做到。公司的系统由一组业务流程组成。而业务流程只是公司运行业务的方式,包括公司每天为客户提供产品和服务,以及保持正常经营的任务、程序、团队和通信。

在许多新的、不断增长的流程中,流程的开发并不总是着眼于流程如何适应公司的总体结构和目标。相反,它们是基于最容易完成工作的方式创建的,而几乎不考虑长期后果。很多时候,最简单的或阻力最小的方法并不是完成任务的最佳方法。不幸的是,在业务流程就绪之后,惯性就开始起作用,并且很难进行更改,这可能会阻碍成长。一个随意设计的系统通常没有做好处理业务增长的准备,并且实际上可能反过来阻碍业务增长。

业务流程分析是研究公司所有业务流程以确定如何改进的系统方法。以下是四个基本步骤:

第一,明确流程、参与者以及当前正在做的事情,并有一个明确的起点和终点。最好从对业务最关键的流程开始。这些流程应该直接与公司的主要产品、收入或费用相联系。

第二,做一个流程的演练,以清楚和完整地记录它。作为这一步的一部分,采访关键员工,了解他们对该流程的看法。寻找流程中信息丢失或被误导的缺口。例如,遍历生产过程使分析人员能够直接观察该过程。与员工的直接沟通可以帮助分析人员完整地记录系统,分析人员可以在巡视过程中询问员工问题。

第三,检查当前的流程,找出强项和可以改进的地方,如瓶颈、摩擦点和弱点。寻找提高流程价值的方法。如果这个过程某一步骤没有为公司增加价值,那么就把它移除。如前所述,直接观察可以阐明需要改进的地方。此外,直接参与这一过程的人对其问题和瓶颈有深入的了解,并且他们最有可能为这些问题提供最佳的解决方案。

第四,在此基础上进行分析,提出改进方案。

三、机器人流程自动化

许多业务流程需要专业人员从各种数据源捕获和组合数据。例如,在对账过程中,专业人员需要来自会计总账包的数据,来自外部实体(如税务机关)的数据表,以及来自公司人力资源系统或其他数据系统的数据。然后,专业人员将完成导入或复制这些数据并将其粘贴到电子表格中的过程。这种类型的工作是细致的、耗时的,并且容易由于一时的注意力缺乏或精神疲劳而导致人为的错误。

机器人流程自动化(Robotic Process Automation, RPA)提供了一个有效的解决在执行这些任务时人为错误风险的方案。RPA使用的软件可以跨越各种数据平台,将数据捕获并记录到一个单独的处理设施中,如Excel之类的电子表格。RPA机器人能够模仿很多用户操作,如登录程序、复制和粘贴数据、移动文件和填写表单。“机器人”这个词让人联想到像科幻/奇幻电影中描述的那样会走路、会说话的机器。然而,在该种情况下,机器人实际上只是一组自动化指令,旨在

操纵机器或数据源来完成指定的任务。因为使用 RPA 只需要很少的计算机编程知识,所以大多数专业人员都可以通过培训来使用 RPA 软件自动完成所要做的任务。例如,假设一名员工每周需要将来自政府机构网站的 PDF 表单中的特定数据字段复制粘贴到 Excel 电子表格中,以便对数据进行分析。RPA 机器人无须手动复制和粘贴数据字段,而是可以通过编程进入政府机构的网站,检索并打开 PDF 表单、打开 Excel,将数据从该机构的 PDF 表单移动到公司的 Excel 电子表格中。目前的 RPA 软件在流程图类型的接口中使用了拖放式编程方法,由于用户不需要学习正式的编程语言,因此该任务相对容易学习并实现。

RPA 的优点包括但不限于以下几点:

1. 以比人类更低的错误率更快地完成单调的任务。
2. 留下所有操作和更改的审计跟踪。
3. 百分之百一致,因为它每次都以相同的方式执行编程功能(除非代码或数据损坏)。
4. 将人们从无聊和重复的工作中解放出来,使得他们可以专注于更有价值的工作,从而提高员工士气和生产力。
5. 能够一天 24 小时,一周 7 天不间断地工作。
6. 尽管最初的实施需要资本投资,但与 RPA 相关的长期成本大大低于人工成本。

四、人工智能与会计数据处理

人工智能(Artificial Intelligence, AI)是旨在感知、推理和理解的计算机软件。从历史上看,计算机软件通过一系列 if/then 条件工作,在这些条件下,每个操作都具有两个可能的结果(是/否,开/关,正确/错误,一个/零)。而基于推理、归纳、直觉、情感和生物化学的人类推理非常复杂,导致一系列可能的结果。此外,人工智能试图模仿人类的决策,这取决于知识和直觉的结合(例如根据经验记住变量之间的关系)。

(一) AI 在业务环境中的优势

1. 可以一天 24 小时工作;
2. 不会生病、死亡或被解雇;
3. 是非常快的数据处理器,尤其是在必须评估众多规则(程序)的情况下。

(二) AI 的类型

1. 神经网络是处理元素的集合,这些处理元素协同工作以像人类大脑一样处理信息,包括从以前的情况中学习并推广概念。
2. 基于案例的推理系统使用的过程类似于人类从以前的相似经历中学习的过程。
3. 基于规则的专家系统在设置的规则的基础上运行以得出答案。这些不能由系统本身更改,必须由外部来源(即计算机程序员)更改它们。
4. 智能代理是应用内置或学习的知识库来执行特定的、重复的和可预测的任务的程序,例如,向计算机用户展示如何执行任务或在网站上搜索特定的财务信息。
5. 专家系统是一个交互式系统,它试图模仿给定领域中人类专家的推理。当本地缺少人类专家时,它对于解决非结构化问题很有用。

(三) 利用 AI 来提高处理会计数据的效率和有效性的方式

1. 自动化数据输入和分析

从单据图像中获得的信息(例如费用报告、供应商发票、客户付款等)会自动分配给适用的会计总分类账。例如,收款人在 ATM 上存入支票,ATM 对支票数据(例如,付款人的银行账户数

据,美元金额等)进行拍照,然后从付款人的银行账户中提取资金并将其存入收款人的银行账户,对其进行分别记录。通过不断重复数据输入过程,计算机最终学会识别差异和异常。因此,可以更早地进行更正,并增强会计报告的完整性。

2. 减少欺诈

通常,审计人员只能手动审核付款文件中的一小部分(例如供应商发票、付款合同、员工费用报告等),而 AI 可以审核全部文件。由于 AI 具有通过分析大量数据进行学习的能力,因此它在识别欺诈方面更加有效。

3. 加强支出政策

扫描付款凭证的全部数量,例如,采购申请和采购订单、供应商发票、信用卡付款情况,可以分析员工的支出报告(包括支出文件),以识别不符合政策的支出(例如不允许的供应商和差旅费)。

人工智能系统的一个重要好处是,它们可以从错误中“学习”,并被编程不重蹈覆辙。智能系统可以被编程识别并与客户和供应商进行交流,捕获、编码和处理诸如发票和采购单之类的日常事务,跟踪付款期限,并确保批准被及时记录。目前市面上很少能买到这样的产品。公司通常会开发自己的软件,但是由于软件开发成本以及缺乏有经验的程序员,人工智能的实现成本可能会很昂贵。

五、云计算

计算机硬件和软件价格昂贵,需要维护,使用寿命相当短,而且需要相对频繁的资金投资。因此,许多公司选择将计算机业务外包。**云计算**(Cloud Computing)只是将计算机操作外包给外部供应商。“云计算”这个术语来自一种描述两个网络之间的云的网络映射技术。“云”由许多提供硬件和软件出租的提供者组成。公司付钱给供应商来管理他们的计算机进程并存储数据。因此,公司有效地将硬件资本支出和计算机过时风险外包给这些供应商。他们还可以通过直接花费这些成本来获得税收收益,而不是在资产的有效期限内对硬件和软件成本进行资本化、维护并计提折旧。

然而,由于云计算服务提供的敏感数据的方便和易用性,云计算也带来了一些风险。比如,云供应商无法为数据提供足够的安全性,数据可能会丢失、被盗或损坏;云供应商可能无法提供支持公司运营所需的服务水平。

六、SaaS 的定义及其优缺点

云服务主要分为三类:基础设施即服务(IaaS);平台即服务(PaaS);软件即服务(SaaS)。其中**软件即服务**(Software as a Service,SaaS)是云计算的一个常见例子。公司可以从供应商那里购买软件并通过互联网交付。软件过去以磁带、软盘或 CD 的形式分发。公司必须购买该软件的多个副本,然后在这些磁性介质上购买升级的新版本。SaaS 通过允许客户访问供应商主机的服务器软件,创建了一种新的软件分发方式。公司喜欢这种安排,因为 SaaS 有以下优点:

1. 它减少了公司在自己的计算机上安装和维护软件及数据的成本。
2. SaaS 提供了随用随付的服务,因此公司可以按月为软件付费。
3. SaaS 扩展方便,客户只购买他们需要的东西,然后可以根据需要添加功能和用户。
4. 该软件是自动更新和访问的,用户可以随时随地连接到互联网。

在风险方面,使用 SaaS 会使公司依赖 SaaS 提供商来使用软件。SaaS 提供商可能会遇到服务中断、强制对公司不想要的软件进行服务更改,或者遭遇数据泄露导致关键客户数据外泄的情况。

七、区块链

合同包括双方的口头或书面协议。合同提供两个或多个实体之间具有法律约束力的协议。然而,在数字世界中,合同可能难以执行,因为数据可以被快速、方便地更改。在加密货币的例子中,实际拥有该货币的人可能会有问题,因为该货币没有物质形式。一个人可以将加密货币发送给一方,然后立即将相同的加密货币发送给另一方。从本质上来讲,这个人可以用同样的货币购买商品和服务,就像伪造货币一样。因此,为了使合同对签约方具有效力,签约方必须能够锁定协议,而不必担心合同在事后被修改。这种保护措施被称为不可否认。不可否认本质上是防止一方通过否认签名的真实性来否定协议。然而,因为不存在合法的副本,不法的一方可能会试图更改或使合同无效。但是,如果合同的1,000个合法和真实的缔约方会议文件分别储存在1,000个不同的地点,不法的一方能够把1,000份副本全部更改吗?大概不会。如何将一个真实的副本存储在1,000个不同的位置呢?通过使用区块链技术。

区块链(Block Chain)技术,或一般的分布式账本(Distributed Ledger)技术,是比特币等加密货币的基础技术。分布式分类账是一个存储在几个地点或多个参与者之间的数据库,所有数据和事务都不是在一个中心位置进行处理和验证的。在所有相关方达成一致意见之后,才会存储数据。然后把文件盖上时间戳,并给出唯一的密码签名。因此,分布式分类账技术为数据库中存储的信息提供了可验证和可审计的历史记录。

区块链是一种特殊类型的数据库或分布式分类账,它具有单独的记录或块,这些记录或块在一个序列列表中链接在一起,称为块链。这些记录或块由网络中的多个对等的节点或各方进行验证。这些块被链接到其他块,使它们不变或不可变。由于契约记录在这些块中,然后链接到其他块,因此秘密修改契约的可能性实际上被消除了,因为这些块被合并到对等网络中的多个节点中。这些块本质上是不可变的,因为一个修改后的块不能被网络中所有其他节点上存储的其他块所识别。换言之,如果一个块被修改,网络上的每个人都会知道修改后的块与他们在节点上的块副本不匹配。

智能契约(Smart Contracts)是执行契约条款的计算机化交易协议,是使用密码签名的交易在区块链网络上部署的代码和数据(有时称为功能和状态)的集合,并嵌入到区块链协议中。智能契约设计的一般目标是满足通用合约条件(例如付款条款、留置权、保密性甚至执行),最小化异常(恶意和意外),以及最小化对受信任的中介机构(例如律师)的需求。因此,他们将书面合同提升到了另一个层次。智能契约是自动执行的软件程序,它通过软件代码而不是纸来代表双方之间的协议。软件代码运行并嵌入到区块链协议或平台中,不需要任何类型的中介(如代理或经销商)来控制执行的协议。软件代码表示协议的条款,并且是不可变的,因为它被合并到一个区块中,该区块是链的一部分。

区块链技术除了数字货币之外的应用可以降低或消除中介成本,提高交易透明度,并为可信任的分权提供框架。该框架还具有侦查逃税、减少腐败、追踪非法支付和洗钱以及侦查资产挪用的潜力。

举例来讲,区块链技术的应用之一是供应链。区块链技术能够通过精简的点对点网络无缝地交换数据,所有的交易都不可更改地存储起来,并可用于审计,这将使供应链系统更加透明和可靠。区块链技术和智能契约可以减少或消除伴随国际货物运输而来的大量文件以及处理国际金融交易所需的时间延迟。

第四节 数据分析

一、商业智能

(一) 大数据

大数据(Big Data)是每个人都声称能理解的术语之一,但没有人能给出一个明确的定义。一些人认为,大数据是一个不适合 Excel 电子表格的数据集。另一些人说,要想被认为是大的,数据集必须大于 1 艾字节(1 艾 = 10^{18})。大数据之所以成为可能,是因为硬件和软件处理常规数据处理系统所不能处理的超大数据集的能力取得了显著进步。大数据通常被用来分析非常大的数据集中的模式或趋势。无论你对规模的相对看法如何,大数据都带来了重大的机遇和挑战。例如,《财富》100 强公司的审计师可以下载客户的所有交易事项,并对所有人进行审计。

术语“量”“多样性”“速度”和“准确性”通常用于描述大数据。其中,体积指的是数据量。多样化是指处理数据的类型,如数字、文本、图像、音频和视频。速度是指数据生成和处理的速度,大数据通常是实时可用的。其准确性与数据的质量有关。使用大数据的机会是巨大的。大数据可以从许多来源收集数据,使得企业可以直接通过互联网搜索、社交媒体帖子和人口统计数据向对其产品感兴趣的潜在客户投放广告。医院可以获取患者的数据,以筛查被送进急诊室的无行为能力患者的有害药物的相互作用或药物过敏。同时,大数据也给个人隐私带来了挑战。所有这些数据都是实时提供给营销人员的,这令人担忧——更不用说数据泄露给客户和公司带来的成本,或者政府使用大数据时侵犯公民权利的风险。公司还面临着与潜在客户保持人性化互动的挑战,这些客户可能并不想成为公司的几个字节而已。

(二) 企业如何使用结构化、半结构化和非结构化数据

结构化和非结构化是数据频谱的对立面。**结构化数据**是计算机可以轻松使用的、有组织的数据。相反,**非结构化数据**不是以计算机可以轻松处理这些数据的形式组织的。**半结构化数据**介于其中。结构化数据的一个常见示例是已放入关系数据库的数据。关系数据库由行和列组成,数据被放在这些行和列相交的单元格中。在结构化数据库中,可以用结构化查询语言(SQL)查询其中可能包含的任何信息。例如,一个数据库可能包含两个表:第一个表用于记录与客户表相关联的销售记录(这意味着我们会有一个公司所有销售记录的列表,以及与这些销售记录相关的客户列表)。第二个表是现金收据,它也可以与那些已经付款的客户相关联。数据是结构化的,使得这些数据元素可以方便地组合在一起。如果销售表与客户表及现金收据表相结合,我们就可以很容易地看出哪些客户已经付款,哪些没有付款。没有付款的客户的查询结果包括公司的未收销售,换言之,就是公司的应收账款。

非结构化数据是未被组织起来以便查询的数据。非结构化数据可以是文本、图像、数字或音频。邮件是非结构化数据的一个很好的例子。半结构化数据由不同层次的结构化数据和非结构化数据组成。据估计,95%的大数据是由非结构化数据组成的。公司寻求使用结构化、半结构化以及非结构化数据提高运营绩效,以把更人性化的产品推荐给客户,找到问题的根源,找出产品和工艺中的瓶颈和缺陷,且能够更好地理解客户习惯。

(三) 数据转换

虽然数据→信息→知识→洞察→行动模型的来源并不明显,但是这个概念为专业人员理解数据的发展提供了一个有用的工具。数据本身并没有用,为了变得有用,它必须被组织成信息。

数据由数字、字符和符号组成。数据可以被看作是事实。有组织的数据是有意义的,因为它可以作为信息被使用。通过对数据进行组织,可以回答以下五个基本问题:何人、何物、何处、何时和如何组织数据。此外,考虑到其他特征,如强度、复杂性和相对等级,需要将数据转换为信息。知识可以被认为是一个人通过教育和经验获得的信息和能力。换言之,一个人获取信息并通过他或她自己的内部处理将其转化为知识。洞察力更为主观,在洞察力方面,人们钻研他们的知识库来建立联系或查看不太明显的模式。洞察力也可以被认为是判断力。数据、信息、知识和洞察力的真正价值在于做出明智的决定或采取行动。



图 6-4-1 数据转换

(四) 管理数据分析的机遇与挑战

商业智能和人工智能的首要目标是创建能够模仿人类洞察力的计算机程序——建立事实数据库,并在数据中找到联系和模式。其挑战在于捕捉像人类的洞察力和决策这样主观的东西,并把它转换成离散的、客观的计算机代码,可编程软件的详细指令。从本质上说,这个过程创建了成千上万条复制人类决策所必需的规则。这项工作的最大机遇是获取和保存专家知识、洞察力和决策技能,以便它们不会被未来的专业人员丢失。此外,在适当的信息技术控制下,计算机程序不会感到疲劳、打电话请病假或休假。计算机程序可以一天 24 小时、一周 7 天工作。计算机程序处理数据的速度更快,体积更大,考虑的细节也比人类决策者要多。

(五) 为什么数据和数据科学能力是战略资产

公司的数据被认为是公司最有价值的资产之一。利用该资产并做出更好决策的能力代表了公司数据的战略价值及其分析数据的能力。公司可以利用从分析多年来收集的数据中获得的独特见解。这些见解用于帮助获得和保留客户,以识别和纠正成本高于收益的流程,并通过识别公司生产或运营流程中的低效之处来降低成本。实际上,商业智能是使用应用程序、工具和最佳实践将数据转换为可操作的信息,从而利用公司数据的战略和运营价值。

(六) 商业智能

商业智能(Business Intelligence)与应用程序、工具以及最佳实践的集合相关联,这些应用程序、工具和最佳实践将数据转换为可操作的信息。商业智能提供有关组织当前状态并如何将其引导至预期方向的高级管理信息,以及有关组织关键成功因素的即时信息。商业智能工具依靠高级软件来搜索从组织内部和外部获得的大量数据,将有关组织的信息显示为条形图、饼图、柱状报告或任何其他适合于高层管理人员决策的形式,以发现模式、趋势和关系。

二、数据挖掘

(一) 数据挖掘的定义

数据挖掘(Data Mining)包括对大型数据集使用分析工具。从本质上讲,数据挖掘涉及大量数据的查询。其思想是在组织常规收集但在大多数操作设置中不使用的数据中查找模式、关系和见解。

(二) 数据挖掘的挑战

数据挖掘通常使用来自组织外部的数据。公司寻找有用的信息,并确定能够提高销售、降低成本和更有效的客户服务的趋势。使用数据挖掘的一个经典但可能是杜撰的例子是,发现购买啤酒和尿布之间存在显著的相关性。通过数据挖掘技术,一家大型零售商发现,在周五下午,购

买尿布的年轻父亲也会购买啤酒。因此,零售商开始将尿布和啤酒放在更接近的位置,以便利用这一新发现的相关性,增加尿布和啤酒的销量。另一个例子是信用卡诈骗。如果你曾经收到过信用卡公司的通知,提示你账户上有可疑活动,那么信用卡公司已经使用数据挖掘技术识别出你正常购买习惯之外的消费模式。

(三)数据挖掘

数据挖掘结合了信息技术和统计信息,其目标是从不同的角度分析数据并将其汇总为有用的信息。数据挖掘是通过不断进行回顾和重新思考(即增强先前分析)的迭代过程,努力获得所需的结果。一旦有所发现,进一步的数据挖掘可能会得出更多结论。数据挖掘并不是一门准确意义上的科学,它可以被看作是一门艺术,或是一门技能,因为虽然使用计算机系统来查找解决方案,但是从计算机输出中区分出有意义的模式高度依赖于分析人员的判断力、专业知识和直觉、统计能力或在数据之间找到关系的能力。样本容量是影响统计能力的主要因素之一。一般来说,样本量越大,样本中数据之间存在统计学上显著关系的可能性越大。然而,仅仅因为可以找到具有统计意义的关系,并不意味着这种关系具有任何实际意义。例如,一些减肥计划发现,与那些不尝试新饮食的人相比,那些尝试新饮食的人在尝试新饮食前后的情况在统计上有显著差异。然而,在三个月的时间里,这种差别可能只有两磅。大多数人并不认为在三个月内减掉两磅是重要的。因此,数据挖掘是一个具有挑战性的领域,需要运用知识、经验和智慧。

数据挖掘可以分为五个步骤:

1. 数据挖掘的第一步是发现异常,例如异常值、聚类、意外更改或与预期的偏差。这可能涉及识别有趣的异常数据记录,或者需要进一步调查的数据错误。
2. 第二步是找到变量之间的关系,然后以某种方式将这些关系聚类(分组)。聚类并不总是那么容易,因为事先不知道结构。
3. 一旦数据被聚类,第三步就是概括这些关系,以使所展示的结构适用于新数据以及原始数据库。
4. 第四步是回归分析(包括线性回归和多元回归)试图找到量化函数或方程,以最小的误差对数据进行建模,即估计数据或数据集之间的关系。
5. 数据挖掘的最后一步提供了数据集的表示形式,包括可视化和报告。

(四)使用查询工具来检索信息

结构化查询语言(Structured Query Language, SQL)是一种用于挖掘大型数据集的成熟工具,是用于在关系数据库管理系统(RDBMSs)中访问和操纵数据的语言,通过使用一组用于创建、更新和检索信息的命令来构造和执行查询。

SQL中有三个基本命令:

- ◆SELECT
- ◆FROM
- ◆WHERE

其中SELECT表示选择用户感兴趣的数据范围。换言之,用户希望查询的结果是什么, FROM标识数据所在的表。WHERE限制数据,以使其满足特定的标准,例如,WHERE Date < 12月31日。其他命令,如数学函数、分类数据、连接表等,也在SQL中使用。

在数据挖掘过程中,为了实现揭示模式和提供见解的目标而执行的常见任务包括:

- (1)检测异常以识别偏离预期的数据。异常数据可以帮助组织发现欺诈、软件问题或其他意外情况。

(2) 查找变量之间的关系以将数据分类到与已知结构不同的聚类中。聚类可能具有挑战性,因为发现新的组织结构可能需要多轮反复试验。

(3) 在已知结构的边界内对数据进行分类,这些结构已被普遍化以适应新数据集。

(4) 执行回归分析,包括线性回归和多元回归,以尽可能准确地对数据之间的关系建模。

(5) 汇总数据,以便可以将其呈现给使用该信息的人员。汇总是包含聚合信息的数据或文档的直观表示。

三、分析工具

(一) 将分析模型与数据契合的挑战

由于技术的进步,会计人员现在可以使用强大的数据分析或统计工具。这些工具使得会计人员能够分析以前需要统计学家提供分析服务的数据集。然而,在访问工具和知道如何有效地使用它们之间有一个至关重要的区别。这些工具非常善于将数据分析模型与数据集进行拟合。而且每个模型都是建立在一组准则的基础上的,这些准则包含了一定的潜在统计假设。如果任意地将数据分析模型应用于任何给定的数据集,可能会违反重要的基本假设,从而得出错误的分析、结果及结论。例如,许多数据分析模型依赖的数据集是随机的而且是正态分布的(想象一下标准的钟形曲线)。如果数据不是随机且正态分布的,但使用依赖于数据正态分布的模型进行评估,那么分析、结果和结论将是不可靠的。

(二) 描述型、诊断型、预测型和规范型数据分析

数据分析有四种基本类型:

- ◆ 描述型
- ◆ 诊断型
- ◆ 预测型
- ◆ 规范型

1. 描述型数据分析(Descriptive Data Analytics)

描述型数据分析描述发生了什么。顾名思义,描述型分析提供了描述实体的事件和实体运行的信息。描述型数据分析侧重于使用各种简明扼要的度量类型来描述数据,这种度量可以帮助我们理解数据是如何定位、如何变化的,以及数据分布可能是如何倾斜的。描述型统计基本上就是对数据进行汇总和报告。帮助我们查看数据位置的度量包括平均值(平均数)、中值(数据的中点)、众数(有多少数据点是相同的)以及百分位数(数据如何分组)。度量使我们看到数据是如何分散或变化的,包括范围(最高和最低的数据点)、四分位范围(数据在中间 50%的数据集)以及方差(平方的平均值偏离平均值的程度或与平均值有关的数据点是如何分散的)。最后,标准差是衡量数据点相对于均值如何分布的另一种方法。对于标准钟形曲线来说,68%的数据在均值 2 个标准差的范围内,95%在均值 4 个标准差的范围内,99.7%在均值 6 个标准差的范围内。标准差比方差更实用,因为方差是平方数,而标准差不是平方数。比如方差是用美元的平方衡量的,而标准差是以美元衡量的。它帮助我们了解数据分布的对称性,从而显示数据在多大程度上偏向一边或另一边。

2. 诊断型数据分析(Diagnostic Data Analytics)

诊断型数据分析关注的是事情发生的原因。公司使用历史数据进行更深入的研究(向下钻取),以发现模式和关系,从而对感兴趣的问题提供更深入的了解。例如,一家药店想知道为什么大量的顾客将他们的处方拿到竞争对手那里,药房对其长期收集的客户端数据进行诊断型数据分

析。它分析顾客处方的成本,从顾客家到药房的距离,药房非处方的宣传,顾客取药时的等待时间等信息。药房发现,顾客在取药时平均等待时间有所增加。通过分析客户数据,该问题被诊断出来。下一步是找出如何减少客户等待时间。

3. 预测型数据分析(Predictive Data Analytics)

预测型数据分析试图知道为什么会发生某些事情,然后创建一个模型来预测未来可能发生的事情。因此,预测型数据分析试图确定什么是可能发生的。它主要是试图预测基于描述型和诊断型数据分析的最可能的结果。预测分析旨在确定未来的趋势和倾向。例如,投资者使用描述型和定性分析来预测某些股票的未來表现。从本质上讲,他们使用预测分析来观察股价图表的极右边缘。预测型数据分析采用各种方法和技术为决策者提供见解和建议。它使用数据挖掘技术、大数据、统计模型和机器学习等工具来创建预测数据模型。这些模型用于识别数据中的模式和关系。

例如,许多杂货店都提供客户忠诚度计划,为登记的客户提供特价和打折优惠。这些程序还通过记录顾客在他们商店里的每一次购买记录来捕捉顾客的购买习惯数据。可以对这些忠诚度计划所获取的销售数据进行分析,以确定客户的购买模式,从而预测客户未来的购买行为。因此,杂货店可以将来自各种场景的数据输入到模型中,以预测客户的购买行为。

例如,根据预测模型,当冰淇淋价格打折10%时,顾客是否会购买更多的配料?公司使用预测分析来确定客户最可能接受的报价,以及发布这些目标报价的最佳时间。预测分析还可以很好地评估客户的终身价值,从而让公司提前知道在获取客户上要花多少钱,同时仍能实现每个客户可接受的利润率。

4. 规范型数据分析(Prescriptive Data Analytics)

规范型数据分析与预测型分析一起发挥作用,以确定对于所给定的未来的机会或问题要采取哪些行动。例如,投资者根据不同的风险容忍度和经济因素,使用规范型分析来确定投资的进入点和退出点。规范型分析通常使用人工智能技术来评估不同概率的许多不同场景的结果。然而,规范型分析只擅长于它们的基本模型和假设。换言之,由于没有人能预见未来,规范型分析仅提供了一种可能会发生什么观点,给出分析中使用的内部和外部数据。

(三) 聚类、分类和回归

三种流行的数据分析模型包括聚类、分类和回归。

1. 聚类(Clustering)

聚类专注于将观察分为几类(集群),以使同一聚类中的观察彼此更相似,寻求在数据集中找到相似的数据点。使用几种算法或工具来查找和分组相似的数据点。有些算法使用硬聚类,其中每个数据点都被分配到一个聚类或组。其他方法使用软聚类,其中每个数据点都属于一个基于概率模型或数据点应该包含在一个集群中的可能性的聚类。聚类是市场研究中最常用的方法,在市场研究中,公司希望更好地了解不同客户群体的偏好。

2. 分类(Classification)

分类试图将数据点分组到分类或类别中,这些数据是离散的。在回归问题中,当因变量(目标)是分类时,将结果与自变量(特征)相关联的模型称为“分类器”。许多分类模型都是二元分类器,例如在信用欺诈检测中的信用卡交易。多类别分类并不少见,例如将公司分类为多个信用评级类别。在分配评级时,结果变量为序数,这意味着类别具有不同的顺序或排名(例如,从低信用度到高信用度)。在度量规模上,顺序变量介于分类变量和连续变量之间。分类和聚类之间的主要区别是,分类已经设置和预定义了分类或类别,并试图将数据放入这些分类中。与此相反,聚类通过分析数据来寻找这些类别的优点。分类可以用来根据人口统计特征和他们对某些产品

的偏爱倾向对客户进行分类。然后可以构建分类数据模型,根据新客户或潜在客户购买这些产品的可能性对他们进行分类。

3. 回归(Regression)

回归专注于对连续目标变量进行预测,试图预测基于模型或方程上的数。数据本质上是连续的或数值的。用户试图找到最适合的数据,例如,回归方程可以用来建模或预测某一产品的成本。将一个参数或数量输入到方程中,成本是计算的输出。大多数读者已经熟悉多元线性回归(例如,普通最小二乘)模型,但是还存在其他学习技术,包括非线性模型。这些非线性模型对于涉及具有大量大型数据集特征的问题非常有用。属于回归类别的问题的一些示例是使用历史股市收益预测股票价格表现,或使用历史公司财务比率预测债券违约的可能性。

(四)线性回归

线性回归是一种帮助我们了解两个变量之间关系的工具。

除了包括一个以上的独立变量,多元回归与一元回归是完全相同的。回归分析提供的信息是在样本数据基础上的基于各种变量之间的关系,我们用决定系数(The Coefficient of Determination)来度量回归方程的拟合度。回归分析的优缺点等内容在本书第二章第三节已经做了介绍,此处不再赘述。

敏感度分析的作用是确定统计分析结果的可靠性。统计模型依赖于某些假设,并且可以根据基础数据的变化而变化。因此,这对确定结果对于假设和数据变化的敏感度是有用的。对假设和数据进行修改,以查看主要结果是否仍然成立。如果仍然成立,则认为结果对工具和数据的更改是不敏感的,由此所发现的结果应该是非常有力的。

例如,提出从另一家公司购买一条产品线的投资模型。该模型包括利率和劳动率的使用。一旦数据被分析,该模型建议采取该项目并购买产品线,敏感度分析将用于投入其他利率和劳动率,以确定该模型是否受标准利率和劳动率的影响很大,或者模型对于这些利率的变化并不敏感。在所有的数据分析模型中,敏感性分析都受到相同的统计假设的约束,因此绝不能盲目地把敏感性分析作为最终结论。

(五)时间序列分析

1. 时间序列分析的定义和应用

时间序列(Time Series)就是按照时间顺序排列的一组数据序列。时间序列分析就是发现这种数据的变动规律并用于预测的统计技术。

数据可以作为横截面数据集或时间序列数据集进行分析。横截面数据来自对多个实体在同一时间点的观察。例如,1,000家公司在其财年的最后一天的销售收入代表了所有1,000家公司销售收入总额的一个截面。使用这个数据集,可以计算出公司的平均值、中值、众数和方差。假设一个经理对一个产品线在一年内的成本行为感兴趣,经理可以收集生产线每天运行的成本数据。我们假设生产线每天都在工作。到年底,经理将拥有约260个数据点,每个工作日一个。这种类型的数据集称为时间序列数据集,因为数据是在一段时间内收集的,而不是在一个时间点上收集的。

分析的重点是变量如何随时间变化。例如,一年中生产线的成本是如何变化的?在时间序列分析中有四个主要的测量部分:趋势、季节、周期和随机。如果数据稳步增加或减少,结果图中将显示一条向上或向下倾斜的线,称为线性趋势。其他常见的趋势是指数型和S形趋势。时间序列数据也可能显示季节特征。例如,一家专门出租山地自行车的公司的大部分收入来自夏季,因此,该公司的时间序列图将显示从6月到8月的销售高峰或激增。与季节特征类似,时间序列数据也可能表现出周期性特征,因为商业环境往往会随着时间从增长到衰退而循环。与季节模

式不同,商业周期更难预测,因为它们的时间框架没有设定。商业周期的长度从一年到更长不等。最后,时间序列数据表现出一种随机或不规律的模式,一些人称之为噪音。随机模式表示影响实体、企业和整体经济的不可预测性事件。例如油价波动、地震和金融危机等。

时间序列分析的主要用途是确定趋势和模式以及预测未来。因为没有人知道未来,只有过去的历史能够帮助我们预测它。例如,时间序列分析使用历史数据来预测特定产品线的未来成本。

2. 时间序列分析的优点和局限性

时间序列分析的优点:

- (1) 预测预期的未来活动的的能力。例如,冬季的夹克销量要高于夏季。
- (2) 确定影响未来预期事件的因素。
- (3) 根据过去的活动来分析当前结果以确定趋势变化。

时间序列分析的局限性包括:

- (1) 随着时间的推移,起作用的因素的可靠性可能会降低。
- (2) 分析得出的结论可能会产生误导。例如,冬季夹克销量的增加可能是人口增加的结果,也可能是销量在减速增加。

(3) 假定过去适用的因素的可预测性在将来保持不变。

(六) 解释如何使用预测分析技术形成见解并提出建议

预测分析是一种使用数据(历史、当前和预测)、统计算法和机器学习技术来对诸如消费者或行业行为模式之类的趋势进行洞察的技术。此外,可以基于对数据的基本假设以及对未来结果的预测来提出建议。

预测分析可以通过计算概率来评估预测方案的可能性,以说明其可行性。

第一步,假设(例如产品类型、出售的数量、单位成本、季节性产品状态等)必须真实无误,以确保分析结果的准确性和质量。

第二步,对实现每种预测方案所需的条件提出建议。

第三步,当客户选择在线上购买商品并准备完成交易时,就会在零售部门中普遍使用预测分析,然后,网页将显示客户与初始商品一起购买的其他产品。例如,一位客户购买了台锯,结账后,网页上会显示客户与台锯一起购买的锯片、护目镜和手套。

(七) 探索性数据分析

探索性数据分析(Exploratory Data Analysis)更多的是一种方法,而不是一组技术和工具,不同于传统的建模和分析技术,它鼓励数据本身揭示其基本结构,而不是过早地采用假设或统计方法。它通常用作数据分析过程中的第一步或前一步。探索性数据分析使用视觉或图形工具(例如直方图和散点图)以及大量方法来寻找数据中的模式,识别并提取重要变量,发现异常值或异常现象中包含的数据集,测试假设和问题数据,并洞察数据集。

探索性数据分析的主要作用是开放地进行探索,以便获得对数据的一些新的、通常是不可怀疑的洞察力。探索性数据分析力求:

1. 最大限度地了解数据集。
2. 发现底层结构。
3. 确定重要变量。
4. 检查异常值和异常情况。
5. 测试基本假设。
6. 建立简约模型。

7. 确定最佳因素。

(八) 敏感性分析的定义及何时适用此分析

敏感性分析(Sensitivity Analysis)使用反复试验来确定变量或假设变化对最终结果的影响。在确定是否有理由花费更多资源以获得更准确的预测时,敏感性分析很有用。通过使用计算机软件,极大地方便了敏感性分析中固有的反复试验法。

敏感性分析的主要用途是资本预算,利率或收益的微小变化会大大改变项目的获利能力。

但由于以下原因,敏感性分析受到限制:

- (1) 使用假设代替事实;
- (2) 单独考虑变量而不是全部考虑。

(九) 模拟模型的运用(包括蒙特卡罗技术)

如果一种情况的参数是已知的(例如,生产产品的成本和约束条件),标准线性回归将预测正常产能下生产运行的总成本。然而,大多数实际情况需要使用许多不便实现的变量。因此,为了模拟实际情况,使用随机输入变量而不是对某些变量进行约束假设。虽然将模拟输入看作随机变量很方便,但它们实际上是基于一系列利率变化的可能值的范围和基于特定概率分布的取值。

最为人熟知的模拟应该就是蒙特卡罗模拟(或概率模拟),该模型模拟可能结果的概率,从一个可能值的分布给定一系列随机输入。设计随机输入的目的是模拟实际发生的概率。

例如,假设你的公司正考虑从一家大型跨国公司收购子公司,因为该子公司生产的产品与你的主要产品互补。然而,在向该子公司报价之前,公司应该考虑许多变数和未知因素。这些未知因素包括:子公司的产品根据成本效益原则可以用公司的设备生产,该交易中子公司的员工百分比,看不见的竞争组合产品,未来经济条件对新产品的影响是什么,利率对该交易的影响是什么,等等。

蒙特卡罗模拟可以从一个可能的输入范围内随机取一个值投入到所有的变量和未知参数,从而计算一个可能的输出。蒙特卡罗模拟对模拟结果进行记录,并重复模拟过程。该模拟使用从概率分布中随机选取的不同的值对模型运行数百次甚至数千次。这也被称为假设分析或目标寻求分析。

(十) 模拟模型的优点和局限性

模拟模型的优点:

可以在运行模型时包含所有可能的值,而不是只包含一个假定的最佳情况值。模拟要求计算机软件转换所有进出模型的可能值的分布。

模拟模型的局限性:

在所有的数据分析模型中,模拟的好坏仅取决于输入的质量。

正如古老的计算机科学谚语所说,无效信息输入=无效信息输出。因此,模拟测试的结果不能被盲目地认为是最好的结果。结果应该只是对整个决策过程的投入的一部分。

(十一) 假设(或目标寻求)分析

假设分析(What-If Analysis)是通过场景变化确定对模型结果的影响的过程。例如,分析师可以通过建立模型来预测潜在的利润。

1. 每售出 50 个、100 个或 150 个单位;
2. 每售出 50 个单位,分销成本降低 5%;
3. 每售出 50 个单位,销售工时增加 5%。

当决策者考虑到特定结果并需要确定如何实现目标时,就会发生目标寻求。

(十二) 大数据的局限性(包括其他数据分析)

数据分析并不适合所有应用程序和分析。数据分析可能不适合的情况包括但不限于：

1. 用户级别的数据结果不完整。通常,组织可用的数据仅限于从访问组织的网络资源(即网站)或在网络上查看该组织的广告所获得的数据。数据仅代表目标市场,因此,潜在的未开发市场可能存在,其数据未被捕获。

2. 很难解释为什么是这样的分析结果。数据通过以下方式进行处理:

(1)将数据分为几组并应用分析方法。

(2)使用算法直接分析数据。

这两个过程都可以预测结果并提供指导,但是算法分析方法往往使非技术人员更难证明选择某种行动方案而不是另一种行动方案。例如,将更多的资本支出花在产品 Z 而不是产品 P 上。

3. 数据受无用信息(通常称为噪声)的影响。单个不正确或无用的变量会破坏结果,并需要额外的工作时间来处理数据,以获得有意义的结果。

4. 用户级别的数据结果需要在使用前进行解释。通常,收集的数据从文本格式转换为数据可视化。数据可视化有助于识别趋势和相关性,这些趋势和相关性基于文本的数据中未被检测到的风险。

四、数据可视化

(一)利用表格和图形设计的最佳实践来避免复杂信息沟通中的偏差

数据可视化(Visualization)是向读者传递信息的一种方法。数据可视化的优势从谚语“一张图片胜过千言万语”就可以看出来。事实上,一个构思良好且构造合理的图像比阅读传达相同信息的文本更快、更完整地传达重要的信息。关键是正确的图像设计能够为使用者提供信息。在选择最有效的数据可视化技术来传达结果时,必须谨慎。有了今天的技术,任何人都可以创造漂亮的图片,并称之为可视化。然而,为了清楚地传达数据中包含的信息,可视化的创建者必须使用基本统计原理和假设的知识,并在选择适当的可视化时进行判断。一个糟糕的视觉设计会扭曲你想要讲述的真实故事。

(二)评估数据可视化选项并选择最佳表示方法

数据表是向读者提供信息的第一层次。图表有效地使用空间,可伸缩,并且可以很简单。由于电子表格的流行,大多数用户都熟悉表格和图表的制作,软件程序可以使数据表易于访问,并使数据的查找和控制相对容易。下面列出了可以而且应该应用于表和图的六个最佳实践:

◆规划。在第一个单元格接收数据之前,应该计划表和图的用途和内容。了解表格/图表的受众,并据此制订计划。

◆焦点。表格/图表的焦点应该是设计中最突出的部分,这样读者才会立即认出它。

◆校准。对于表格来说,文本必须排列整齐或对齐到单元格的左侧,数值数据必须对齐到单元格的右侧。列标题必须根据内容适当地放置。

◆大小。字符(文本和数字)的大小很重要。如果文本太小,就很难阅读,而且可能不会被阅读。此外,推荐使用常用字体,不常见的字体使用者可能会将注意力集中在字体上,而不是设计者想要传达的信息。

◆杂乱。杂乱是每个表格和图表的敌人。一定要留出足够的空白来帮助读者关注信息。

◆颜色。颜色可以是提供深度、焦点和对比度的强大工具。然而,过多或规划不当的颜色会影响表格/图表的信息。

数据可视化工具箱包含许多工具。就像扳手和螺丝刀一样,数据可视化工具或方法也有其

特定的用途。更常见的数据可视化工具和它们的用途将会在下面进行描述。有些更擅长于提供比较,有些擅长于提供分布,还有一些擅长于提供关系和趋势。它们也可以根据其预期用途进行分类。

1. 比较

柱状图用于比较数据类别或跨时间的数据。例如,我们可以比较公司不同年份的销售收入的金额。

饼图可以用来比较数据类别的比例。然而,大多数数据科学专家都不赞成使用饼图,除非各种情况有很大的不同。试图解释 10%的部分和 12%的部分之间的区别是有问题的。因此,最好只使用饼图上的几个部分。

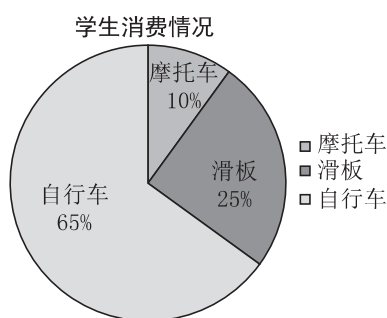


图 6-4-2 饼状图

2. 分布

从图形上看,直方图与柱状图相似。二者的区别在于其所描绘的信息不同。直方图用于显示单一或多个变量的频率分布情况。在直方图中,x 轴通常用于显示类别或范围,y 轴用于显示测量值、值、频率。直方图显示的是数据的分布形状。这样,人们能够直观地检查数据是正态分布还是指数分布。例如,可以使用直方图解释说明班级整体考试成绩,在图中可以显示成绩分布情况以及对个别学生成绩进行相对分析,除了这一点,直方图还可以用来显示客户群体的年龄分布情况。

我们以 2018 年我国(大陆)GDP 数据指标为例(数据来源于国家统计局网站)。2018 年我国(大陆)29 个省份的 GDP 合计值为 900309 亿元,首次突破 90 万亿元大关,创历史新高。按可比价格计算,GDP 比上年增长 6.6%。29 个省份(不包括港、澳、台地区)中,2 省 GDP 突破 80000 亿元,1 省 GDP 突破 60000 亿元,3 省 GDP 突破 40000 亿元,12 省 GDP 突破 20000 亿元,还有 11 省 GDP 在 20000 亿元以下。直方图如图 6-4-3 所示。

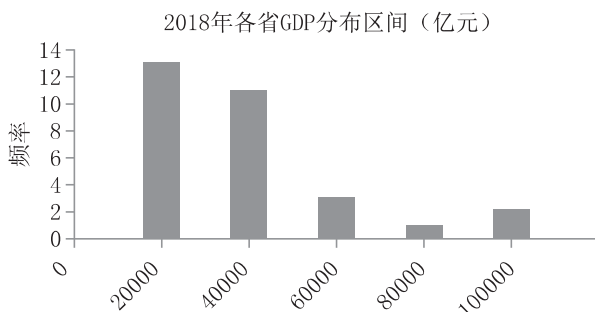


图 6-4-3 2018 年我国 29 个省份(不包括港、澳、台)GDP 分布区间

点图类似于直方图,只是它使用垂直的点而不是条来表示数据分布。

箱形图或箱线图使用五个标准度量显示分布的数据集:最小数据点,最低四分位数据点(第一四分位数),中位数或中间点的数据,数据点的第三个四分位数,以及最大的数据点。箱形图不

显示单独的值,而且可能会出现偏差,但它们也是显示异常值的少数技术之一。它们在显示分布之间的对比时也很有用。

3. 位置

地图和填充地图显示地理空间数据,因此可以根据地理位置查看数据点。当数据与地理位置(如国家、省、城市和邮政编码)相关时,地图和填充地图是可视化数据的强大工具。

我们根据国家统计局 2018 年第四次全国经济普查,调查的全国各省产业活动单位的数量(数据来源:国家统计局——第四次全国经济普查公告(第七号) http://www.stats.gov.cn/tjsj/zxfb/201911/t20191119_1710340.html),做一张地图作为可视化图表,见图 6-4-4 所示。



图 6-4-4 2018 年经济普查全国各省产业活动单位的数量地图

4. 关系

散点图显示根据相对于 x 轴和 y 轴的位置绘制的数据点。该技术提供了一个关于数据如何相关或定位以及其分布的视图,然而,散点图并没有显示两个以上变量之间的关系。

气泡图是散点图的一个版本。在这里,散点图上的点被合并成气泡,气泡大小不同,表示数据点的数量不同。与饼图类似,气泡图最好在气泡大小变化很大时使用。

热图通过颜色深浅的变化来显示变量之间的关系。热图提供了一种查看数值的直观的方式。事实上,热图可以显示大量的数据,而不需要过多的阅读器。热图还可以通过显示与周围显著不同的颜色强度来帮助识别异常值。然而,热图并不像其他技术那样精确,因为区分不同颜色的色调是很困难的。

5. 趋势

折线图可能是最常用的用于时间序列数据图形化的工具。折线图(或者线形图)展示了两个变量之间的关系;它们通常用于跟踪记录变量随时间推移发生的变化和趋势(在 x 轴上设置一个变量为时间)。折线图依次连接各个数据点以帮助推断一段时间内的变化趋势。折线图通常用来显示某些测量值随时间变化的情况,例如某一特定股票价格在五年内的变化情况或者一个月内每日客户服务电话的数量变化。折线图的一个限制是折线图只显示随时间变化的数据。

我们以 2020 年上半年全国房地产投资数据为例:2020 年,1—2 月份全国房地产开发投资同比下降 16.3%,1—3 月份同比下降 7.7%,1—4 月份同比下降 3.3%,1—5 月份同比下降 0.3%。

1—6月份,同比增长1.9%。可以看到,随着疫情得到有效控制,我国房地产投资金额也在稳步提升,由负转正。通过折线图(图6-4-5)所示。

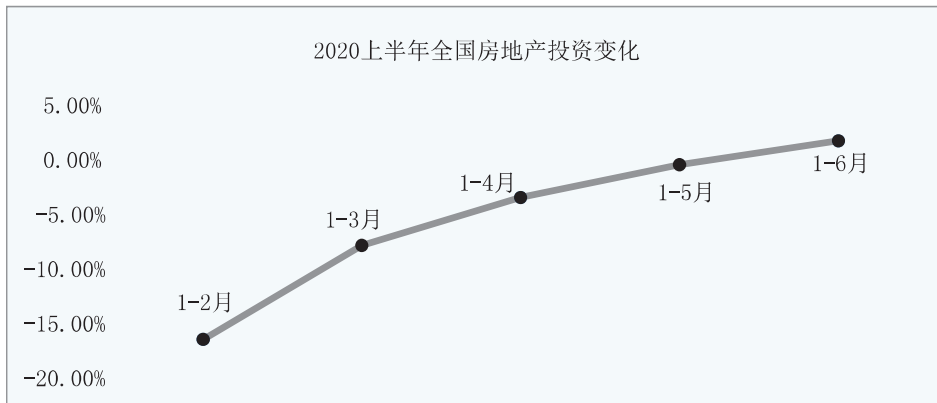


图 6-4-5 2020 年上半年我国房地产投资同比增幅

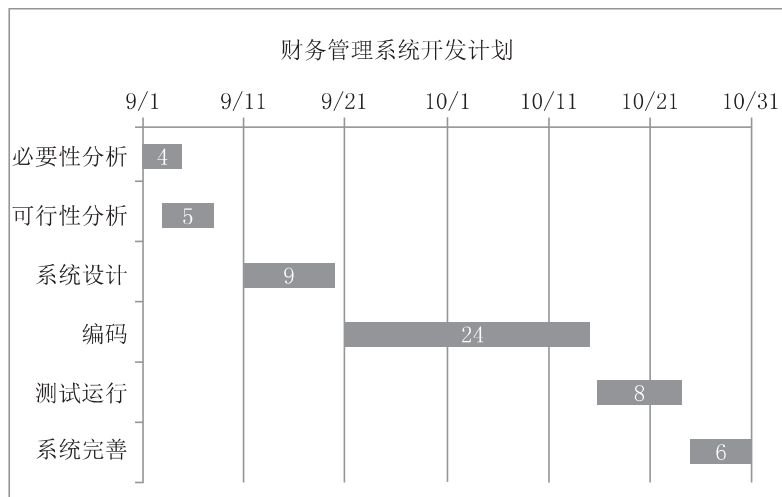
6. 进度

甘特图是一种特殊的水平条形图,可用于描述项目时间线,项目活动持续时间以及活动间的重叠。通过显示活动的起始日期和结束日期以及活动之间的重叠关系,甘特图为项目的管控提供了宝贵的帮助,帮助管理层掌握项目进度。例如,甘特图通常用于显示项目时间表,任务重叠,相对任务完成状况(即在显示实际任务执行期间的条形图中表明完成百分比的部分条形图),每项任务的资源分配情况,关键程序和可交付成果。

例:某公司计划开发新型财务管理系统

财务管理系统开发计划

任务名称	开始时间	时长(天)	完成时间
必要性分析	2019/9/1	4	2019/9/5
可行性分析	2019/9/3	5	2019/9/8
系统设计	2019/9/11	9	2019/9/20
编码	2019/9/21	24	2019/10/15
测试运行	2019/10/16	8	2019/10/24
系统完善	2019/10/25	6	2019/10/31



(三) 可视化技术的优点和局限性

1. 可视化技术的优点

(1) 数据可视化是指通过将数据放在可视化的背景(例如图形或图表)中来帮助读者或其他用户理解数据的意义。可以说明在基于文本的数据中可能经常被忽略的趋势、相关性或其他安排。计算机本身不需要使用可视化来进行决策。

(2) 数据可视化工具可以提供允许用户操纵数据或钻取数据以进行假设分析的交互式功能。可视化有助于在数据元素之间进行视觉比较,这有助于决策者在分析阶段识别模式、模式偏差和离群值。可视化通常被更广泛的受众理解,因为可视化用最少的术语将信息的核心表达出来。可视化选项包括直方图、箱型图、散点图、点图、表格、仪表盘、条形图、饼图、折线图和气泡图。添加颜色和形状有助于减轻决策者解释大量信息的负担而将精力集中在更重要的信息上,从而更有效地传达结果。

2. 可视化技术的局限性

(1) 可视化可能会产生误导。创建图像的编码人员可以影响观看者对该图像的解释。

(2) 可视化工具可能被操纵而呈现出有偏差的图片。

(四) 可视化工具最佳实践,确定沟通结果最有效的渠道

为避免复杂信息交流中的失真,图表设计的最佳实践准则包括但不限于以下内容:

1. 通过确保坐标轴从零开始并且不跳过数值来利用整个坐标轴。
2. 不要给设计增加过多的信息。
3. 限制用于确保突出显示的不同数据点间的颜色数量,避免使用相同颜色的其他阴影。例如,使用红色而不是浅红色、柔和红色、中等红色等。
4. 确保设计不要求用户解释数据之间的关系,显示计算以避免误解。
5. 在最终确定图表之前,请其他人提供有关设计的反馈。请记住,一个人清楚的事情实际上可能会使其他人感到困惑。