



海关高等教育教材
HAIGUAN GAODENG JIAOYU JIAOCAI

海关数据分析与SPSS应用

HAIGUAN SHUJU FENXI YU SPSS YINGYONG

段景辉◎主编 陈蓉◎副主编

 中国海关出版社



海关高等教育教材
HAIGUAN GAODENG JIAOYU JIAOCAI

海关数据分析与 SPSS应用

HAIGUAN SHUJU FENXI YU SPSS YINGYONG

段景辉 ◎ 主 编 陈 蓉 ◎ 副主编

图书在版编目 (CIP) 数据

海关数据分析与 SPSS 应用/段景辉主编. —北京: 中国海关出版社,
2019. 1

(海关高等教育教材)

ISBN 978-7-5175-0323-1

I. ①海… II. ①段… III. ①海关管理—统计分析—软件包

IV. ①F745. 2-39

中国版本图书馆 CIP 数据核字 (2018) 第 281628 号

海关数据分析与 SPSS 应用

HAIGUAN SHUJU FENXI YU SPSS YINGYONG

作 者: 段景辉 陈 蓉

责任编辑: 李 多

出版发行: 中国海关出版社

社 址: 北京市朝阳区东四环南路甲 1 号

邮政编码: 100023

网 址: www.hgcb.com.cn

编 辑 部: 01065194242-7529 (电话)

01065194231 (传真)

发 行 部: 01065194221/4238/4246 (电话)

01065194233 (传真)

社办书店: 01065195616 (电话)

01065195127 (传真)

www.customskb.com/book (网址)

印 刷: 北京中献拓方科技发展有限公司

经 销: 新华书店

开 本: 787mm×1092mm 1/16

印 张: 9.75

字 数: 200 千字

版 次: 2019 年 1 月第 1 版

印 次: 2019 年 1 月第 1 次印刷

书 号: ISBN 978-7-5175-0323-1

定 价: 30.00 元



海关版图书, 版权所有, 侵权必究

海关版图书, 印装错误可随时退换

“海关高等教育教材”丛书序

我国古代海关的起源，一般认为是在西周，至今已有三千多年的历史。其名称几经变迁，经历过关、塞、关楼、津、市、市舶司、月港督饷馆、钞关、户关、工关、榷关、常关等，直到清代康熙二十四年（1685年），中国历史上才第一次出现了正式的、以海关命名的进出境关口。1840年鸦片战争以后，中国沦为半殖民地半封建国家，丧失了关税自主权和海关行政管理权。为培养我国自己的税务、海关人才，清政府于1908年在北京创办了税务学堂。1913年，北洋政府教育部批准将税务学堂改名为税务专门学校。在四十余年的办学中，该校为中国海关培养了两千余名专业人才。

中华人民共和国的海关专业教育起步于1953年，以上海海关学校的设立为标志。1980年经国务院批准，上海海关学校升格为上海海关专科学校。1996年更名为上海海关高等专科学校。2007年3月，教育部批准在上海海关高等专科学校的基础上设立上海海关学院，设置了管理学、经济学、法学、文学等若干本科专业。上海海关学院作为全国唯一系列设置海关类课程和专业的本科院校，承担着传播海关专业知识、培养海关专门人才、进行海关学术研究、开展海关国际交流与合作的重任。

上海海关学院的发展，事关人才培养重任，事关国家的海关改革与发展，如何更好地服务国家战略，需要我们冷静思考、科学规划、抓住机遇、真抓实干，在新的起点上以新的办学思路、新的办学举措、新的办学业绩来适应新时代海关事业的改革与发展，以及国际经济贸易的发展。

在海关总署党组的领导和全国海关的支持下，升本后的上海海关学院严格按照教育教学规律组织教学工作，夯实教学管理基础，切实加强教学监控，狠抓人才培养质量，并积极探索构建以海关学为核心的学科群，探讨海关学基础理论，明确海关学的研究对象，建立包括海关管理学、海关法学、关税学等在内的二级学科理论体系，努力使中国海关拥有与自身地位相匹配的学科地位。

为满足迅速发展的海关高等教育的实际需要，2009年以来，学校连续两轮组织教师编写“海关高等教育教材”。该套教材涵盖了海关法律、关税、商品归类、海关估价、海关稽查、海关统计、风险管理、原产地规则、海关专业英语等诸多内容，具有涉及海关专业诸多领域、专业性强、偏重原理、强调理论



和实践相结合等特点。“海关高等教育教材”不仅能满足海关高等教育的需要，同时也是对海关实践的理论总结，对丰富和发展海关学科专业，构建以海关学为核心的学科群具有重要的意义。

自 2015 年起，在教材使用的基础上，我们认真总结教材编写方面的经验，贴合海关和外经贸事业发展对人才培养的实际需求，启动新一轮教材体系建设工作。我们将始终坚持以教材体系建设为抓手，推动学校走特色办学、内涵建设发展之路，着力在“三型一高”，即应用型、复合型、涉外型的高素质人才培养方面发挥应有的作用，早日实现建设海关特色鲜明、服务国家战略、具有国际影响力的世界一流高等学府的办学目标。我们相信，本着科学、务实、与时俱进的精神编写的教材，不仅能够为学校的教学提供科学适用的教材，也将为海关学科建设和人才培养做出积极的贡献。

上海海关学院院长

丛玉豪

2018 年 2 月

前 言

大数据时代代表着信息技术发展进入了一个新的时代，代表着爆炸性的数据信息给传统的计算技术和信息技术带来的技术挑战和困难，代表着大数据处理所需的新的技术和方法，也代表着大数据分析和应用所带来的新发明、新服务和新发展机遇。

由于大数据处理需求的迫切性和重要性，近年来大数据技术已经得到了全球的学术界、工业界，以及各国政府的高度关注和重视。早在 2010 年 12 月，美国总统办公室下属的科学技术顾问委员会（PCAST）和信息技术顾问委员会（PITAC）向奥巴马和国会提交了一份《规划数字化未来》的战略报告，把大数据收集和使用的工作提升到体现国家意志的战略高度。2012 年 7 月，联合国在纽约发布了一本关于大数据政务的白皮书《大数据促发展：挑战与机遇》，全球大数据的研究和发展进入了前所未有的高潮。为了紧跟全球大数据技术发展的浪潮，我国政府、学术界和工业界对大数据也予以了高度的关注。2012 年以来，国家自然科学基金、973 计划、核高基、863 等重大研究计划都已经把大数据列为重大的研究课题。为了推动我国大数据技术的研究发展，2012 年中国计算机学会（CCF）组织了 CCF 大数据专家委员会，还特别成立了一个“大数据技术发展战略报告”撰写组，并已撰写发布了《2013 年中国大数据技术与产业发展白皮书》。

目前，国内已有 50 多所高校开设了大数据相关专业。此外，中国人民大学、北京大学、中国科学院、中央财经大学、首都经济贸易大学还联合培养大数据分析硕士。从其课程设置的结构来看，除了统计学科的统计建模、数据挖掘等内容外，还需具备计算机学科的相关知识，以及应用领域的经济学和管理学背景。因此，大数据分析技术是集统计学科、计算机学科、经济与管理学科等一体的科学技术。

海关在履行进出口贸易统计职能中产生了各项进出口统计数据，政府和企业通过这些数据可以快速了解行业市场状况，及时调整策略，对于有效开拓国际市场，提高企业经营效率，应对市场竞争等有着重要作用。因此，海关数据的处理和应用逐步提上日程。本书从巩固学科基础，培养学生动手能力，提高



学生综合素质的教学目标出发,试图建立一个实践与理论相结合,着重培养学生实际动手能力为主的实验教学课程。SPSS 是世界上最早的统计分析软件,它的分析结果清晰、直观、易学易用,已经广泛地应用于经济学、数学、统计学、管理学、医学等各个领域。SPSS 是采用图形菜单驱动界面的统计软件,它突出的特点是操作界面极为友好,输出结果美观漂亮。学生只要掌握一定的计算机操作知识,精通统计分析原理,就可以利用该软件为数据分析服务。本书旨在使学生系统、正确地掌握数据统计与分析技术——SPSS 软件的理论基础,熟练掌握 SPSS 软件的应用,应用 SPSS 软件进行描述分析、频数分析、数据探索、交叉表分析、图形分析、回归分析、因子分析、聚类分析、时间序列分析、多重响应等。

通过本书的学习,学生不仅能巩固所学理论知识,更能切实掌握各种统计分析方法在统计软件 SPSS 中的实现,并能正确解释 SPSS 的运行结果。在实验教学中,培养学生的科学素质和良好的实验室工作习惯,强化学生的时间意识,为培养具有创新精神和实践能力的高素质人才奠定良好的基础。

编 者

2019 年 1 月

目 录

海关数据简介 1

 一、大数据定义 1

 二、大数据分析 1

 三、大数据技术 2

 四、海关数据 4

实验一 海关数据的建立和管理 7

 一、实验目的 9

 二、准备知识 9

 三、实验内容 18

 四、实验步骤与实验结果 18

 五、实验总结 19

 六、实验作业 19

实验二 海关数据的描述性统计分析 21

 一、实验目的 23

 二、准备知识 23

 三、实验内容 26

 四、实验步骤与实验结果 27

 五、实验总结 31

 六、实验作业 31

实验三 海关调研数据的方差分析 33

 一、实验目的 35

 二、准备知识 35

 三、实验内容 45

 四、实验步骤与实验结果 46

 五、实验总结 51



六、实验作业	51
实验四 进出口贸易额的相关与回归分析	53
一、实验目的	55
二、准备知识	55
三、实验内容	78
四、实验步骤与实验结果	78
五、实验总结	81
六、实验作业	81
实验五 构建海关统计指标中的因子分析	83
一、实验目的	85
二、准备知识	85
三、实验内容	93
四、实验步骤与实验结果	93
五、实验总结	95
六、实验作业	96
实验六 商品数据的聚类分析	99
一、实验目的	101
二、准备知识	101
三、实验内容	108
四、实验步骤与实验结果	109
五、实验总结	115
六、实验作业	115
实验七 海关税收数据的时间序列分析	117
一、实验目的	119
二、准备知识	119
三、实验内容	137
四、实验步骤与实验结果	137
五、实验总结	147
六、实验作业	147
参考文献	148

海关数据简介

一、大数据定义

“大数据”是一个体量特别大、数据类别多的数据集，这样的数据集无法用传统数据库工具对其内容进行抓取、管理。“大数据”的特点一是数据体量（volumes）大，一般在10TB左右，但在实际应用中，很多企业用户把多个数据集放在一起，已经形成了PB级的数据量；二是数据类别（variety）多，数据来自多种数据源，数据种类和格式日渐丰富，已冲破了以前所限定的结构化数据范畴，囊括了半结构化和非结构化数据；三是数据处理速度（Velocity）快，在数据量非常庞大的情况下，也能够做到数据的实时处理；四是数据真实性（Veracity）高，随着社交数据、企业内容、交易与应用数据等新数据源的兴起，传统数据源的局限被打破，企业愈发需要真实、安全的信息。

二、大数据分析

大数据不仅是数据多，更是对大数据进行分析，以获取很多智能的、深入的、有价值的信息。越来越多的应用涉及大数据，而这些大数据的属性，包括数量、速度、多样性等，都体现了大数据不断增长的复杂性，所以大数据的分析方法在大数据领域就显得尤为重要，可以说是决定最终信息是否有价值的决定性因素。大数据分析普遍存在的方法理论有以下四个方面：

一是可视化分析。大数据分析的使用者有大数据分析专家，同时还有普通用户，他们对于大数据分析最基本的要求就是可视化分析，因为可视化分析能够直观地呈现大数据特点，容易被读者所接受，就如同看图说话一样简单明了。

二是数据挖掘算法。大数据分析的理论核心就是数据挖掘算法，各种数据挖掘的算法基于不同的数据类型和格式才能更加科学地呈现出数据本身具备的特点，也正是因为这些被全世界统计学家所公认的各种统计方法，才能深入数据内部，挖掘出公认的价值。另外，也是因为有这些数据挖掘的算法才能更快速地处理大数据，如果某个算法需花上好几年才能得出结论，那大数据的价值也就无从说起了。

三是预测性分析能力。大数据分析最重要的应用领域之一就是预测性分析，从大数据中挖掘出特点，建立科学模型，之后便可以通过模型带入新的数据，从而预测未来的数据。

四是数据质量和数据管理。大数据分析离不开数据质量和数据管理，高质量的数据和有效的数据管理，无论是在学术研究还是在商业应用领域，都能够保证分析结果的真实性和价值。



三、大数据技术

大数据技术，就是从各种类型的数据中快速获得有价值信息的技术。大数据领域已经涌现出了大量新的技术，它们成为大数据采集、存储、处理和呈现的有力工具。大数据处理关键技术一般包括大数据采集、大数据预处理、大数据存储及管理、大数据分析及挖掘、大数据展现和应用（大数据检索、大数据可视化、大数据应用、大数据安全等）。

（一）大数据采集技术

通过 RFID 射频数据、传感器数据、社交网络交互数据及移动互联网数据等方式获得的各种类型的结构化、半结构化（或称之为弱结构化）及非结构化的海量数据，是大数据知识服务模型的根本。重点要突破分布式高速、高可靠数据查询或采集、高速数据全映像等大数据收集技术；突破高速数据解析、转换与装载等大数据整合技术；设计质量评估模型，开发数据质量技术。

大数据采集一般分为大数据智能感知层和基础支撑层。大数据智能感知层主要包括数据传感体系、网络通信体系、传感适配体系、智能识别体系及软硬件资源接入系统，实现对结构化、半结构化、非结构化的海量数据的智能化识别、定位、跟踪、接入、传输、信号转换、监控、初步处理和管理等。必须着重攻克针对大数据源的智能识别、感知、适配、传输、接入等技术。基础支撑层提供大数据服务平台所需的虚拟服务器，结构化、半结构化及非结构化数据的数据库及物联网资源等基础支撑环境。重点攻克分布式虚拟存储技术，大数据获取、存储、组织、分析和决策操作的可视化接口技术，大数据的网络传输与压缩技术，大数据隐私保护技术等。

（二）大数据预处理技术

大数据预处理技术主要指完成对已接收数据的抽取、清洗等操作。

抽取：因获取的数据可能具有多种结构和类型，数据抽取过程可以帮助我们将这些复杂的数据转化为单一的或者便于处理的构型，以达到快速分析处理的目的。

清洗：对于大数据，并不全是有价值的，有些数据并不是我们所关心的内容，而另一些数据则是完全错误的干扰项，因此要对数据“过滤”，从而提取出有效数据。

（三）大数据存储及管理技术

要用存储器把采集到的数据存储起来，建立相应的数据库，并进行管理和调用。重点开发复杂的结构化、半结构化和非结构化大数据管理与处理技术。主要解决大数据的可存储、可表示、可处理、可靠性及有效传输等几个关键问题。开发可靠的分布式文件系统（DFS）、能效优化存储、计算融入存储、大数据的去冗余及高效低成本的大数据存储技术，突破分布式非关系型大数据管理与处理技术、异构数据的数据融合技术、数据组织技术，研究大数据建模技术，突破大数据索引技术，突破大数据移动、备份、复制等技术，

开发大数据可视化技术。开发新型数据库技术，数据库分为关系型数据库、非关系型数据库及数据库缓存系统。其中，非关系型数据库主要指的是 NoSQL 数据库，分为键值数据库、列存数据库、图存数据库及文档数据库等类型。关系型数据库包含了传统关系数据库系统及 NewSQL 数据库。开发大数据安全技术，改进数据销毁、透明加解密、分布式访问控制、数据审计等技术，突破隐私保护和推理控制、数据真伪识别和取证、数据持有完整性验证等技术。

（四）大数据分析及挖掘技术

改进已有数据挖掘和机器学习技术，开发数据网络挖掘、特异群组挖掘、图挖掘等新型数据挖掘技术，突破基于对象的数据连接、相似性连接等大数据融合技术，突破用户兴趣分析、网络行为分析、情感语义分析等领域的大数据挖掘技术。

数据挖掘就是从大量的、不完全的、有噪声的、模糊的、随机的实际应用数据中，提取隐含在其中的，人们事先不知道但又是潜在有用的信息和知识的过程。数据挖掘涉及的技术方法很多，有多种分类法。根据挖掘任务可分为分类或预测模型发现，数据总结、聚类、关联规则发现，序列模式发现，依赖关系或依赖模型发现，异常和趋势发现等；根据挖掘对象可分为关系数据库、面向对象数据库、空间数据库、时态数据库、文本数据源、多媒体数据库、异质数据库、遗产数据库及环球网 Web；根据挖掘方法可分为机器学习方法、统计方法、神经网络方法和数据库方法。机器学习方法中，可细分为归纳学习方法（决策树、规则归纳等）、基于范例学习、遗传算法等。统计方法中，可细分为回归分析（多元回归、自回归等）、判别分析（贝叶斯判别、费歇尔判别、非参数判别等）、聚类分析（系统聚类、动态聚类等）、探索性分析（主元分析法、相关分析法等）等。神经网络方法中，可细分为前向神经网络（BP 算法等）、自组织神经网络（自组织特征映射、竞争学习等）等。数据库方法主要是多维数据分析或 OLAP 方法，另外还有面向属性的归纳方法。

从挖掘任务和挖掘方法的角度，着重突破：

1. 可视化分析。数据可视化无论对于普通用户或是数据分析专家，都是最基本的功能。数据图像化可以让数据自己说话，让用户直观地感受到结果。
2. 数据挖掘算法。图像化是将机器语言翻译给人看，而数据挖掘就是机器的母语。分割、集群、孤立点分析还有各种各样五花八门的算法让我们精练数据，挖掘价值。这些算法一定要能够应付大数据的量，同时还具有很高的处理速度。
3. 预测性分析。预测性分析可以根据图像化分析和数据挖掘的结果做出一些前瞻性判断。
4. 语义引擎。语义引擎需要设计人工智能，以从数据中主动地提取信息。语言处理技术包括机器翻译、情感分析、舆情分析、智能输入、问答系统等。
5. 数据质量和数据管理。数据质量与数据管理是管理的最佳实践，透过标准化流程和机器对数据进行处理，可以确保获得一个预设质量的分析结果。



（五）大数据展现与应用技术

大数据技术能够将隐藏于海量数据中的信息和知识挖掘出来，为人类的社会经济活动提供依据，从而提高各个领域的运行效率，大大提高整个社会经济的集约化程度。在我国，大数据将重点应用于以下三大领域：商业智能、政府决策、公共服务。例如，商业智能技术、政府决策技术、电信数据信息处理与挖掘技术、电网数据信息处理与挖掘技术、气象信息分析技术、环境监测技术、警务云应用系统（道路监控、视频监控、网络监控、智能交通、反电信诈骗、指挥调度等公安信息系统）、大规模基因序列分析比对技术、Web 信息挖掘技术、多媒体数据并行化处理技术、影视制作渲染技术，以及其他各种行业的云计算和海量数据处理应用技术等。

四、海关数据

（一）海关数据组成

海关数据构成主要分为三大类：一是通关备案数据，指在业务信息化管理系统中设置的通关管理参数和对应各类审批注册的备案数据，为通关作业预备基础业务标准数据和建立通关作业的底账；二是申报数据，指海关监管对象依法申报项目形式的数据，是海关对监管对象的直接记录，包括进出境运输工具货物和物品的申报数据，其中以对应进出境的报关单数据为主；三是业务流数据，指海关在实际监管业务工作过程中记录监管操作及其处置结果的数据，如审单记录与查验记录。

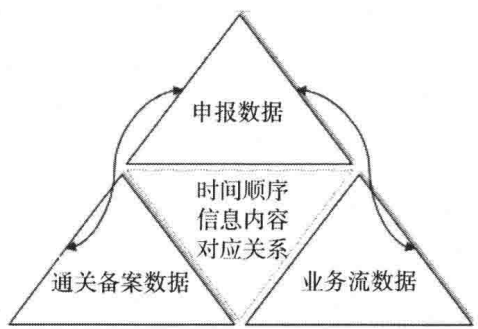


图1 通关备案数据、申报数据、业务流数据之间的关系

（二）海关数据特点

海关数据首先表现为巨大的体量，随着贸易量的增加，进出口总值与报关单量增长较快，因此，海关数据量是庞大的。其次表现为数据的复杂性，海关数据元结构包括商品编码、金额、数量、时间、运输工具、原产地等一系列信息，这些信息会随着相关贸易协议签订而随之发生改变。最后表现为数据的重要性，海关数据不仅是进出口监管的行政记录，更是国家外贸统计数据的唯一来源（全口径），也是海关自身的宝贵资源。

（三）大数据的海关应用

1. 拓展基础数据资源

一是丰富海关执法相关领域基础数据源，打破部门之间的数据壁垒，实现海关缉私行政执法数据、稽查数据、监管场所信息、口岸单位数据等的关联比对，尤其要打破目前各关区和海关各业务系统之间的数据壁垒，实现信息互联互通。二是海关预算和财务管理方面，引入海关通用财务管理、固定资产管理等现有信息系统的基础数据，打破现有系统之间数据分散、相互独立、无法互通的现状，实现“财”与“物”的联动比对，实现海关所有信息系统之间数据的对碰分析。三是推动海关有关部门进一步加大对物资采购、基建修缮等领域信息系统的开发建设，为实现财务管理、政府采购、基建修缮、资产管理的横向联动分析，以及海关总署—直属海关—隶属海关三级纵向联动分析奠定基础。

2. 完善现有海关系统

以海关执法廉政风险预警处置系统为依托，整合优化执法领域审计数据分析平台，在充分利用现有系统数据资源和监控功能的基础上，开发定制部分海关审计监控指标，完善构建按照不同系统用户形成的涵盖全国海关数据指标集中展示和预警模块。同时引入大数据思维，开发适应海关常规审计和经济责任审计实践需要的功能。一方面，突出对被审计单位贯彻落实海关总署党组重大决策、重要政策执行、业务发展动态和变化趋势的整体分析把握，更突出对执法统一性、自由裁量权规制、高风险节点指标等落实情况的监控。另一方面，要充分应用结构分析、多维分析、挖掘分析等分析方法，开展海关横向区域间对比分析、跨业务领域数据结合分析和纵向历史趋势分析，为有效发现系统性、区域性风险提供数据支撑。

3. 外购公认的优秀大数据服务

先进的分析技术，以及大数据工具的发展，正以前所未有的方式帮助用户获取新的统计角度和结果。Tableau、Qlikview 和 NoSQL 等工具和平台的迅速崛起带来了全新的分析视角和机会，基于成熟的分析、视觉化及数据管理的全新生态系统也以日新月异的速度提高信息使用者的分析能力。可提供这类工具的供应商不胜枚举，开放资源的开发商数量更是不计其数。作为海关审计的大数据建设，应该不仅仅局限于海关自身的技术开发力量，更为重要的是学会怎么利用现有的比较成熟的数据采集和挖掘手段，降低成本，提升海关审计监督效益。



实验一 海关数据的建立和管理



