Broadview
www.broadview.com.cn

CDA数据分析师系列丛书

从零进阶！
数据分析的统计基础

人大经济论坛 主编　曹正凤 编著

未来数据分析相关的就业岗位会有1000万人才缺口
CDA数据分析师系列丛书携你与时俱进！

電子工業出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
http://www.phei.com.cn

CDA数据分析师 系列丛书

从零进阶！
数据分析的统计基础

人大经济论坛 主编　曹正凤 编著

電子工業出版社
Publishing House of Electronics Industry
北京•BEIJING

内 容 简 介

全书共 6 章，分别讲解了数据分析的步骤和方法、描述性统计分析、抽样估计、假设检验、方差分析、相关与回归分析，使用简单的语言介绍了这些数据分析基本方法的核心思想和涉及的统计学、概率论等方面的理论内容，并使用图示的方法详细介绍了使用 Excel 2013 进行简单的描述性统计分析和使用 SPSS 进行相关的数据分析的过程与结果分析。

本书适合需要提升自身数据分析理论和实践能力的职场新人；在市场营销、金融、财务、人力资源管理中需要数据分析的人士；从事咨询、研究、分析等的专业人士。也可以作为数据分析师职业培训的教材，普通高等院校非统计专业数据分析的选修教材。

图书在版编目（CIP）数据

从零进阶！数据分析的统计基础 / 人大经济论坛主编；曹正凤编著. —北京：电子工业出版社，2015.2
（CDA 数据分析师系列丛书）

ISBN 978-7-121-25244-0

Ⅰ. ①从… Ⅱ. ①人… ②曹… Ⅲ. ①数据处理—教材②数据处理系统—教材 Ⅳ. ①TP274

中国版本图书馆 CIP 数据核字（2014）第 302724 号

策划编辑：张慧敏
责任编辑：徐津平
印　　刷：三河市双峰印刷装订有限公司
装　　订：三河市双峰印刷装订有限公司
出版发行：电子工业出版社
　　北京市海淀区万寿路 173 信箱　邮编　100036
开　　本：787×980　1/16　印张：12　字数：302.4 千字
版　　次：2015 年 2 月第 1 版
印　　次：2015 年 2 月第 1 次印刷
印　　数：4000 册　　定价：49.00 元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：（010）88254888。

质量投诉请发邮件至 zlts@phei.com.cn，盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线：（010）88258888。

序言：这是一个用数据说话的时代

在 CDA（注册数据分析师）Level I 级教材付诸印刷之际，关于数据分析这个职业及其价值的报道就有很多。比如，下面两条报道就充分体现了在大数据时代下，数据分析的价值。这在以前是从来没有过的。

LinkedIn 的最新投票结果显示，‘统计分析和数据挖掘’是 2014 年最大的求职法宝。LinkedIn 对全球超过 3.3 亿用户的工作经历和技能进行分析，公布 2014 年最受雇主喜欢、最炙手可热的 25 项技能，其中位列榜首的是统计分析和数据挖掘。

麦肯锡公司的一份研究预测称，到 2018 年，在“具有深入分析能力的人才”方面，美国可能面临着 14 万到 19 万人的缺口，而“可以利用大数据分析来做出有效决策的经理和分析师” 缺口则会达到 150 万人。数据科学家将成为 2015 年最热门的职业。

早在 2010 年 2 月，肯尼斯•库克尔在《经济学人》上发表了一份关于管理信息的特别报告——《数据，无所不在的数据》，文中写道：“世界上有着无法想象的巨量数字信息，并以极快的速度增长……从经济界到科学界，从政府部门到艺术领域，很多地方都已感受到了这种巨量信息的影响。”2011 年，麦肯锡发布了《大数据：下一个具有创新力、竞争力与生产力的前沿领域》，使人们在这篇文章里认识到了数据的力量。于是，一夜之间，面向数据分析市场的新产品、新技术、新服务、新业态正在不断涌现。从个人、企业到国家层面，都把数据作为一种重要的战略资产，逐渐认识到了数据的价值，不同程度地渗透到每个行业领域和部门，大大提升了企业的经营利润，推动了经济的发展。

这是一个用数据说话的时代，也是一个依靠数据竞争的时代。目前世界 500 强企业中，有 90%以上都建立了数据分析部门。IBM、微软、Google 等知名公司都积极投资数据业务，建立数据部门，培养数据分析团队。各国政府和越来越多的企业意识到数据和信息已经成为企业的智力资产和资源，数据的分析和处理能力正在成为日益倚重的技术手段。

作为一个数学和统计学的强国，数据分析、数据挖掘和大数据价值挖掘行业在我国仍属于朝阳行业，数据分析人才仍然比较稀缺。各行各业在平常工作中积累的各种各样的数据分析问题仍然没有得到及时有效地解决，有些问题，还是关乎本行业发展的至关重要的问题。数据积累越来越多，期待解决分析的数据问题也越来越多，人们逐渐习惯使用数据作为决策的重要参考依据。据艾瑞的研究报告，未来与数据分析相关的就业岗位会在 1000 万人左右，而目前来说国内合格的数据分析师不足 5 万人，建立一个科学有效的数据分析师培训体系迫在眉睫。

在这样一个用数据说话的时代，积累了丰富的数据分析培训经验的人大经济论坛承担起使命，几番调查研究，几番反复推演论证，在 2013 年，这个大数据的“元年”，CDA 注册数据分析师应运而生！

2003 年，人大经济论坛依托中国人民大学成立，在金融、管理、统计领域已积淀 11 个年头，在国内享有良好声誉。

2006 年，人大经济论坛数据分析培训中心设立，至今经历 8 个春秋，建立了大陆、台湾一线师资团队，培养人才已达 3 万余人。

2013 年，“中国数据挖掘与数据分析俱乐部 CDMC”在人大经济论坛旗下成立，2014 年改名为“CDA 数据分析师俱乐部”。来自政府、金融、电信、零售、电商、互联网、教育等行业人士加入会员，成功举办了数十场行业聚会。紧接着，积累了数据分析培训丰富经验的人大经济论坛在国内展开 CDA 数据分析师系统培训和认证考试，成功见证了 1000 余名数据分析师的成长。

2015 年，人大经济论坛将提供高水平、多层次的数据分析培训服务，以在行业积累 多年的影响力，吸引更好更多的优秀师资，瞄准行业内重要的数据分析问题和难点，攻坚突破，建立更加规范的行业培训体系，引领数据分析培训行业向规范化、有效化和前瞻化方向发展，为数据分析培训做出应有的贡献。

其实，数学（含统计）和英语一样重要，都是人们不可或缺的重要技能。既然英语全民这么重视，数学及其数据分析的技能更加需求于方方面面，更应被做大做强。让我们共同期待人大经济论坛办成另一个数据的“新东方”！

覃智勇

2015 年 1 月 1 日

前　言

感谢您选择“CDA 数据分析师”Level Ⅰ学习系列丛书”之《从零进阶！数据分析的统计基础》

众所周知，数据分析的基础是统计学，没有概率论和数理统计的知识，数据分析尤如无根之草，只能浮游在华丽的词藻上，数据分析师的职业发展之路也走不长远，因此掌握数据分析的统计学基础知识是非常重要的。统计学作为一门学科，其内容之多，不是一本书能概括得了的，而为了使零基础的读者，尽快进阶成一名具有统计基础的数据分析师，本书为此做出了积极的探索。取其精华，论其重点，使读者能尽快地掌握一些数据分析师必备的统计学知识，这是本书的目的，也是学习本书的读者所想要达到的最终目标。掌握本书中的内容，您将在数据分析师这个职业之路上走得更远，更踏实。

本书按照数据分析必备的统计学基础知识来讲解，以三国武将数据为案例背景，由浅入深、由易到难地向您展示统计学基础理论。由于三国时期的历史背景家喻户晓，读者可以不必费力地探究案例的背景知识，让您能够将更多的精力放在学习核心的理论知识上，从而为今后的数据分析工作打下坚实的理论基础。

读者对象

本书适合需要提升自身数据分析理论和实践能力的职场新人；在市场营销、金融、财务、人力资源管理中需要数据分析的人士；从事咨询、研究、分析等的专业人士，也可以将其作为数据分析师职业培训的教材，普通高等院校非统计专业数据分析的选修教材。

阅读指南

全书共 6 章，分别讲解数据分析的步骤和方法、描述性统计分析、抽样估计、假设检验、方差分析、相关与回归分析，使用简单的语言介绍了这些数据分析基本方法的核心思想和涉及的统计学、概率论等方面的理论内容，并使用图示的方法详细介绍了使用 Excel 2013 进行简单的描述性统计分析和使用 SPSS 进行相关的数据分析的过程与结果分析。其中，第 1 章为数据分析的步骤和方法介绍，第 2 章为描述性统计分析的内容，包括平均值、标准差及统计图的介绍，第 3 章至第 4 章为抽样推断的内容，包括抽样估计和假设检验，这是全书的重要内容，也是最重要的数据分析理论基础。第 5 章至第 6 章为统计分析的初步，介绍方差分析、相关分析和回归分析的原理及其软件操作实现。每章都根据所涉及的知识点的不同，选取了案例，并为读者准备了相应的思考题和练习题。

详细的章节内容如下：

第 1 章　数据分析概述

本章主要介绍数据分析的概念、分析步骤和分析方法，介绍如何在 Excel 2013 中安装数据分析工具，这是在后续课程中进行数据分析的基础。

第 2 章　描述性统计分析

本章主要介绍数据分析中最基本的分析方法——描述性统计分析，主要包括数据的集中趋势、离中趋势和数据分布的测度指标分析方法，直方图、茎叶图、箱线图等统计图的含义和画法，介绍如何在 Excel 2013 中实现数据的描述性统计分析。

第 3 章　抽样估计

本章主要介绍推断统计的基础——抽样估计，主要包括抽样估计的基础知识、三种在数据分析中经常用到的分布及中心极限定理等内容，详细介绍抽样误差、抽样估计方法和抽样组织形式等抽样估计的重要内容，本章介绍的内容为数据分析师在进行数据分析时所需要的最基础的知识。

第 4 章　假设检验

本章主要介绍数据分析中必须用到的一种统计分析方法——假设检验，它是抽样推断的主要内容之一，本章的主要内容包括假设检验的基本思想、步骤和假设检验中经常用到的检验统计量，并介绍 SPSS 中常用的几种 T 检验方法。

第 5 章　方差分析

本章主要介绍数据分析中比较多个总体的均值是否相等的检验方法——方差分析，主要包括方差分析的相关概念、单因素方差分析的原理、统计量构造过程等内容，并介绍如何在 SPSS 中实现单因素方差分析及对结果的分析。

第 6 章　相关与回归分析

本章主要介绍相关和回归分析，两者均是应用极其广泛的数据分析方法。主要内容包括变量间的关系、相关分析的概念和步骤、一元线性回归分析的相关概念和相关假定、一元线性回归方程及求法、一元线性回归分析和检验的内容，并介绍如何使用 SPSS 实现相关分析和回归分析。

本书特点

本书的主要特点有两个方面。

一是理论内容画龙点睛。

数据分析涉及统计学、概率论等众多内容，如何较快地实现从菜鸟到数据分析师的进阶，就必须有针对性地学习必要的知识，如何正确地引导读者抓住数据分析的精髓和要点，这是本书试图解决的问题。如果这个问题得到解决，那么将极大地节省读者进行数据分析的成本，从而较快地进阶为一名数据分析师。读者有时候也会有这样的一些感觉，当面对厚厚的一本统计学教材时，总有一种望而却步的感觉，失去了学下去的勇气。编写本书的本意，就在于让读者能在短期内，对数据分析师需要知道的基础知识做一个系统而完整的介绍，恰到好处地对精华内容进行展示，使读者能少而精地把握数据分析的基本要领，从而激起读者进一步学习的欲望。读完本书后，你会发现，其实要成为一名数据分析师，需要掌握的知识也不是太多，因为有些内容贵在精，而不在多。

二是一个案例贯穿始终。

本书在讲解统计基础理论时，均使用同一个案例，且该案例贯穿全书的始终。以三国时期的武将数据作为例子进行介绍数据分析的过程，由于三国的历史背景大家都较为熟悉，因此读者不必费力熟悉数据分析的业务背景，而是直接进入使用数据说话的奇妙世界。

学习方法

本书是数据分析师入门的基础理论部分，其宗旨在于将数据分析师需要具备的核心理论进行描述，而有些统计学理论没有全面的展开，因此读者需要根据自己的需要适当地查阅相关的理论知识，对所学的内容进一步巩固，达到由点到线，由线到面的学习效果。

书中前 4 章的结构是从易到难，由基础到提高，建议读者顺序阅读，以掌握数理统计的基本理论知识。从第 5 章开始，其结构虽然是从易到难，但章节之间相互独立，即可以从任意章节开始学习，不需要遵照从前到后的顺序阅读。

售后服务

为方便读者学习，本书提供了书中实例的源文件下载，请读者进入人大经济论坛（http://bbs.pinggu.org/），注册后搜索“CDA 教材源文件”关键词下载相应的源文件。

本书读者可以在人大经济论坛的“数据挖掘与商业智能”（http://bbs.pinggu.org/forum-133-1.html）就书中的问题进行提问，也欢迎大家就自己遇到的业务问题和大家讨论。同时，也可以向作者发邮件，作者邮箱为 czh_3104@tom.com。

致谢

本书由人大经济论坛策划，曹正凤负责编写和完成统稿。

丛书从策划到出版，倾注了电子工业出版社计算机图书分社张慧敏、石倩、官杨、张童等多位编辑的心血，特在此表示衷心地感谢！

为保证丛书的质量，使其更贴近读者，我们组织了人大经济论坛的多位版主和高级会员参与了本书的预读工作，他们是杨同梅、田佳、孙华枫、原瑜芬、叶阵雨、郑赟、李剑宇、江翊雪、陈鹏、刘莎莎、丁亚军。感谢各位预读员的辛勤、耐心与细致，使得本丛书能以更加完善的面目与各位读者见面，特别感谢覃智勇圆满地组织了本次预读工作和审校工作。

尽管作者们对书中的案例精益求精，但疏漏仍然在所难免，如果您发现书中的错误或某个案例有更好的解决方案，敬请登录社区网站向作者反馈，我们将尽快在社区中给出回复，且在本书再次印刷时修正。

再次感谢您的支持！

目　录

第 1 章

数据分析概述

本章主要介绍数据分析的概念、分析步骤和分析方法，介绍如何在 Excel 2013 中安装数据分析工具，以及如何安装 SPSS 数据分析软件，这是在后续课程中进行数据分析的基础。

1.1 什么是数据分析

从互联网上的词云分析中可以看到，“数据分析”这个词汇的热度很高，然而在这个喜欢炒作的年代，很多词汇和概念都是过眼云烟、昙花一现，究其原因为大多数人都是没有静下心来仔细思考和踏实工作。静下来想一想“数据分析是什么”是很重要的，当人们冷静下来，再让他们解释数据分析到底是什么时，要得到一个不错的答案恐怕是很难的。

不同的人对数据分析有不同的答案，比较常见的答案是，数据分析就是分析数据。从一大堆数据中提取到你想要的信息，就是数据分析。比较专业的答案是，有针对性的收集、加工、整理数据，并采用统计、挖掘技术分析和解释数据的科学与艺术。比较客观的答案是，从行业的角度看，数据分析是基于某种行业目的，有目的地进行收集、整理、加工和分析数据，提炼有价值信息的一个过程。

笔者认为，把数据分析看成是艺术有点过分夸张，而将其看成是过程又过于客观，但两者确实是数据分析从宏观到微观的一种很好的概括。但从本质上看，理解数据分析应从三个方面去把握：一是目标，数据分析的关键在于设立目标，专业上叫做“有针对性”；二是方法，数据分析的方法包括统计分析和数据挖掘两种；三是结果，数据分析最终要得出分析的结果，结果对目标解释的强弱，结果的应用效果如何。在某个实际分析的过程中，解决好了这三个方面的问题，就是一个好的数据分析师。

1.2 数据分析六步曲

概括起来讲，数据分析的过程主要包括：明确分析目的和内容、数据收集、数据处理、数据分析、数据展现和报告撰写等六个步骤。

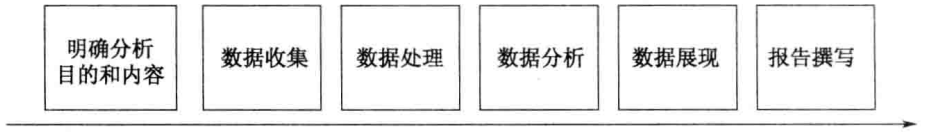

图 1.1　数据分析过程

1.2.1 明确分析目的和内容

在进行数据分析之前，数据分析师应对需要分析的项目进行一个详细的了解，或者自己本身就对此分析项目所涉及的行业有比较深刻的了解，对其内部的运行规律即使做不到了如指掌，至少也要有一个整体框架上的了解。数据分析的对象是谁？数据分析的商业目的是什么？最后的结果要解决什么样的业务问题？数据分析师对这些都要了然于心。对数据分析目的的把握，是数据分析项目成败的关键。只有对数据分析的目的有深刻的理解，才能整理出完整的分析框架和分析思路，因为根据不同的数据分析目的所选择的数据分析方法是不同的。

1.2.2　数据收集

当我们根据分析的目的，选定了相应的设计框架之后，一个重要的问题就出现了，如何能准确有效地收集数据，从而客观全面地反映所要研究的问题的真实状况。数据收集是一个按照确定的数据分析和框架内容，有目的地收集、整合相关数据的过程，它是数据分析的基础。通常数据收集的方法包括观察法、访谈法、问卷法、测验法等。

1.2.3　数据预处理

数据预处理是指对收集到的数据进行加工、整理，以便开展数据分析，它是数据分析前必不可少的阶段。概括起来，统计数据预处理的过程包括数据审查、数据清理、数据转换和数据验证四个步骤。

第一步：数据审查

该步骤检查数据的数量（记录数）是否满足分析的最低要求，字段值的内容是否与研究目的要求一致，是否全面，包括利用描述性统计分析，检查各个字段的字段类型，字段值的最大值、最小值、平均数、中位数等，记录个数、缺失值或空值个数等。

第二步：数据清理

该步骤针对数据审查过程中发现的明显错误值、缺失值、异常值、可疑数据，选用适当的方法进行“清理”，使“脏”数据变为“干净”数据，使得后续的数据分析得出可靠的结论。当然，数据清理还包括对重复记录进行删除。

第三步：数据转换

数据分析强调分析对象的可比性，但不同字段值由于计量单位等不同，往往造成数据不可比。对一些统计指标进行综合评价时，如果统计指标的性质、计量单位不同，那么容易引起评价结果出现较大误差，再加上分析过程中的其他一些要求，需要在分析前对数据进行变换，包括无量纲化处理、线性变换、汇总和聚集、适度概化、规范化，以及属性构造等。

第四步：数据验证

该步骤的目的是初步评估和判断数据是否满足统计分析的需要，从而决定是否需要增加或减少数据量。利用简单的线性模型及散点图、直方图、折线图等图形进行探索性分析，利用相关分析、一致性检验等方法对数据的准确性进行验证，确保不把错误和偏差的数据带入到数据分析中。

上述四个步骤是一个逐步深入、由表及里的过程。先是从表面上查找容易发现的问题（如数据记录个数、最大值、最小值、缺失值或空值个数等），接着对发现的问题进行处理，即数据清理；再就是提高数据的可比性，对数据进行一些变换，使数据形式上满足分析的需要；最后则是进一步检测数据内容是否满足分析需要，诊断数据的真实性及数据之间的协调性等，确保优质的数据进入分析阶段。

1.2.4　数据分析

数据分析是指通过分析手段、方法和技巧对准备好的数据进行探索、分析，从中发现因果关系、内部联系和业务规律，为商业目的提供决策参考。

到了这个阶段，要能驾驭数据、开展数据分析，就要涉及工具和方法的使用。其一要熟悉常规数

据分析方法，最基本的是要了解例如方差、回归、因子、聚类、分类、时间序列等数据分析方法的原理、使用范围、优缺点和结果的解释；其二要熟悉 1 + 1 种数据分析工具，Excel 是最常见的数据分析工具，一般的数据分析我们可以通过 Excel 完成，而后要熟悉一个专业的分析软件便于进行一些专业的统计分析、数据建模等。专业的数据分析工具主要包括：SPSS、R、MATLAB、SAS 等。

SPSS 是世界上最早采用图形菜单驱动界面的统计软件，它最突出的特点就是操作界面极为友好，输出结果美观漂亮。它几乎将所有的功能都以统一、规范的界面展现出来，使用 Windows 的窗口方式展示各种管理和分析数据方法的功能，对话框展示出各种功能选择项。用户只要掌握一定的 Windows 操作技能，粗通统计分析原理，就可以使用该软件为特定的科研工作服务。

R 软件是一套完整的数据处理、计算和制图软件系统。其功能包括：数据存储和处理系统；数组运算工具（其向量、矩阵运算方面功能尤其强大）；完整连贯的统计分析工具；优秀的统计制图功能；简便而强大的编程语言，可操控数据的输入和输出，实现分支、循环，并且用户可自定义功能。R 软件因其开源性、强大的统计计算等功能而受到统计人员的青睐。R 软件具备高效的数据处理和存储功能，擅长数据矩阵操作，提供了大量适用于数据分析的工具，支持各种数据可视化输出。R 软件的一大优势是分析人员可利用简单的 R 程序语言描述处理过程，以构建强大的分析功能。

MATLAB 是由美国 MathWorks 公司生产的商品化应用软件，该软件具有良好的用户界面和实时的人机交互环境，使用该软件可以进行程序设计、统计分析和数据挖掘。一般的统计分析功能都可以在 MATLAB 软件中实现，当然有的时候要适当进行代码的设计。在 MATLAB 软件中，有一大特色就是提供了众多的应用函数，这些函数丰富了软件的功能，也方便了用户。经过多年的改版和更新，MATLAB 的用户界面越来越接近 Windows 的标准界面，操作也越来越简单，编程环境也更加人性化，开发者编写的程序可以不用编译也能运行，同时具有良好的程序调试和纠错功能，这些都为软件的广泛使用提供了重要的支持。虽然 MATLAB 功能强大且界面友好，但由于其商业性质不同于 weka 和 R 的开源性质，用户使用正版 MATLAB 软件时需要支付一定的费用。在 MATLAB 软件中，提供了随机森林算法的接口，和 R 软件一样，用户需要对其参数进行设置，有些应用还需要进行适当的编程才能使用该算法。

SAS 是用于决策支持的大型集成信息系统，但该软件系统最早的功能仅限于统计分析，直到现在统计分析功能也仍是它的重要组成部分和核心功能。在数据处理和统计分析领域，SAS 系统被誉为国际上的标准软件系统，并在 96～97 年度被评选为建立数据库的首选产品，堪称统计软件界的“巨无霸”。SAS 是由大型机系统发展而来的，其核心操作方式就是程序驱动。经过多年的发展，SAS 现在已成为一套完整的计算机语言，其用户界面也充分体现了这一特点。它采用 MDI（多文档界面），用户在 PGM 视窗中输入程序，分析结果以文本的形式在 OUTPUT 视窗中输出。使用程序方式，用户可以完成所有需要做的工作，包括统计分析、预测、建模和模拟抽样等。

1.2.5 数据展现

一般情况下，数据分析的结果都是通过图、表的方式来呈现的，俗话说“字不如表，表不如图”。借助数据展现手段，能更直观地让数据分析师表述想要呈现的信息、观点和建议。常用的图表包括饼形图、折线图、柱形图/条形图、散点图、雷达图、金字塔图、矩阵图、漏斗图、帕雷托图等。

1.2.6　报告撰写

最后阶段，就是撰写数据分析报告，这是对整个数据分析成果的一个呈现。通过分析报告，把数据分析的目的、过程、结果及方案完整呈现出来，以为达成商业目的提供参考。

一份好的数据分析报告，首先需要有一个好的分析框架，并且图文并茂，层次明晰，能够让读者一目了然。结构清晰、主次分明可以使阅读者正确理解报告内容。图文并茂可以令数据更加生动活泼，提高视觉冲击力，有助于读者更形象、直观地看清楚问题和结论，从而产生思考。

另外，数据分析报告需要有明确的结论、建议和解决方案，不仅仅是找出问题，更重要的是解决问题，否则称不上是好的数据分析，同时也失去了报告的意义，数据分析的初衷就是为了满足商业目的而进行的。

1.3　数据分析方法简介

在第一节讲到数据分析概念时，数据分析的三个方面中，最重要的一个方面就是方法，即数据分析方法。在某些特殊的时刻，不同的数据分析方法的选择，会使一个数据分析得出截然不同的结果。数据分析方法分为两种，一个是统计分析方法，另一个是数据挖掘方法。

1.3.1　统计分析方法简介

1. 描述性统计分析

描述性统计分析（Description Statistics）是通过图表或数学方法，对数据资料进行整理、分析，并对数据的分布状态、数字特征和随机变量之间的关系进行估计和描述的方法。描述性统计分析分为集中趋势分析和离中趋势分析和相关分析三大部分。

集中趋势分析主要靠平均数、中数、众数等统计指标来表示数据的集中趋势。例如测试班级的平均成绩是多少？是正偏分布还是负偏分布？

离中趋势分析主要靠全距、四分差、平均差、方差、标准差等统计指标来研究数据的离中趋势。例如，当我们想知道两个教学班的语文成绩，哪个班级的成绩分布更分散时，就可以用两个班级的四分差或百分点来比较。

相关分析是研究现象之间是否存在某种依存关系，并对具体有依存关系的现象进行其相关方向及相关程度的研究。这种关系既包括两个数据之间的单一相关关系——如年龄与个人领域空间之间的关系，也包括多个数据之间的多重相关关系——如年龄、抑郁症发生率和个人领域空间之间的关系；既包括 A 大 B 就大（小），A 小 B 就小（大）的直线相关关系，也可以是复杂相关关系（$A=Y-B\cdot X$）；既可以是 A、B 变量同时增大的正相关关系，也可以是 A 变量增大时 B 变量减小的负相关关系，还包括两变量共同变化的紧密程度——相关系数。实际上，相关关系唯一不研究的数据关系，就是数据协同变化的内在根据——因果关系。获得相关系数有什么用呢？简而言之，有了相关系数，就可以根据回归方程，进行 A 变量到 B 变量的估算，这就是所谓的回归分析。因此相关分析是一种完整的统计研究方法，它贯穿于提出假设、数据分析、数据研究的始终。