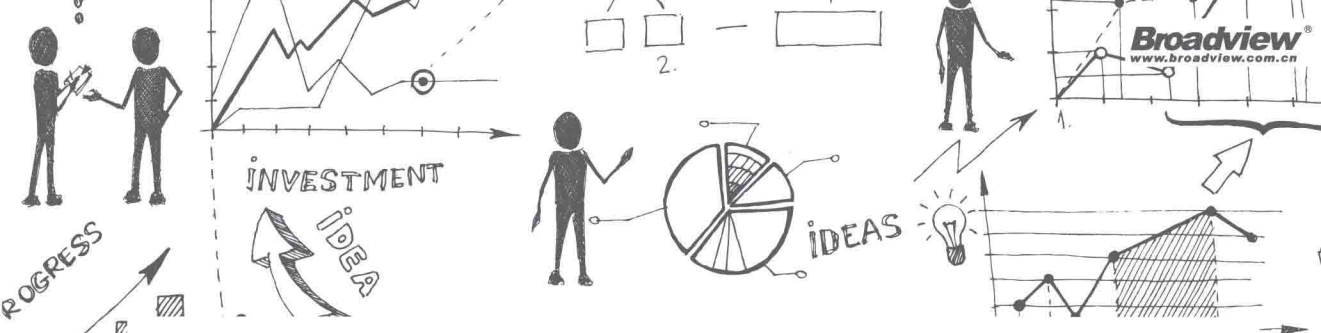


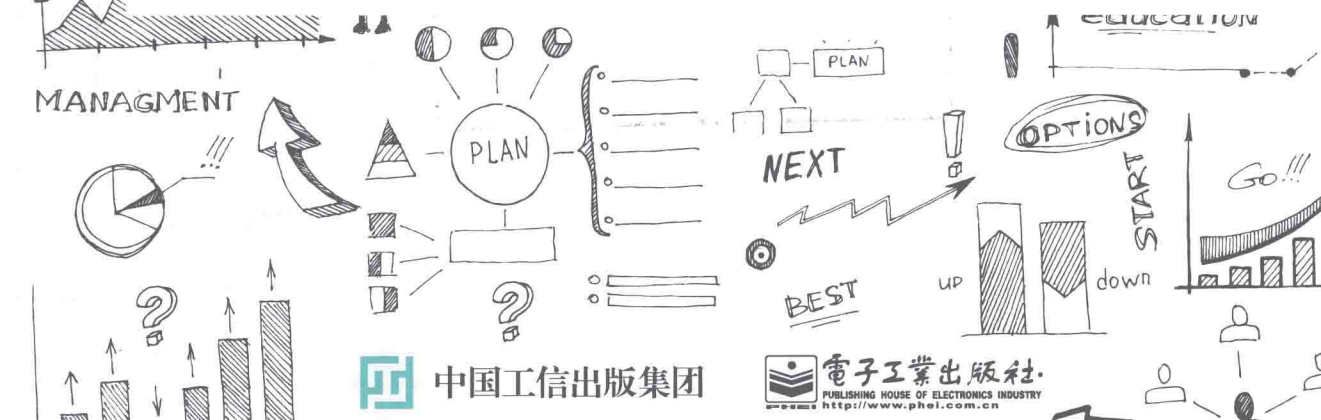
没有数据，经济舞台将苍白无力

没有数据分析，掘金又从何谈起

数据分析

从这里开始

大数据时代 小数据分析



中国工信出版集团

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

大数据时代
小数据分析

屈泽中 编著

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

内 容 简 介

本书是一本大数据时代下进行小数据分析的入门级教材，通过梳理数据分析的知识点，将各类分析工具进行串联和对比，例如：在进行线性规划的时候可以选择使用 Excel 或 LINGO 或 Crystal Ball。工具的应用难易结合，让读者循序渐进地学习相关工具。JMP 和 Mintab 用来分析数据，分析的结果使用 Excel、LINGO、Crystal Ball 来建立数据模型，最后使用 Xcelsius 来动态展示数据分析的结果。书中以两个人的对话为叙述方式，场景描写多，容易进入学习状态，完全是用生动的故事和实用的案例尽可能地贴近生活和工作，让数据分析生动有趣，基本上有高中数学知识就可以理解线性规划等数据分析内容。

本书不仅介绍 Excel 而且介绍使用其他工具软件进行数据分析，可用来拓展互联网公司、传统企业、电商企业、管理咨询公司等各行各业从事数据分析工作的分析师和管理者对数据分析的认知，也适合初中级数据分析师或者想进入数据分析行业的有志之士参考阅读。

图书在版编目（CIP）数据

大数据时代小数据分析 / 屈泽中编著. —北京：电子工业出版社，2015.7
ISBN 978-7-121-26469-6

Ⅰ. ①大… Ⅱ. ①屈… Ⅲ. ①数据处理 Ⅳ. ①TP274

中国版本图书馆 CIP 数据核字（2015）第 142313 号

责任编辑：孙学瑛
印　　刷：北京京师印务有限公司
装　　订：北京京师印务有限公司
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：22.75　字数：468 千字
版　　次：2015 年 7 月第 1 版
印　　次：2015 年 9 月第 2 次印刷
定　　价：69.00 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。

推　荐　序

很高兴泽中的首部专著《大数据时代的小数据分析》能在出版后的短短时间内销售一空，很高兴能受邀为本书的第二印作序。

泽中在公司负责运营数据的挖掘与分析，理论精、实践强，是公司运营参谋领域当之无愧的“大拿”。更难能可贵的是，他能以共性的数据工具为手段，打通工作生活界限，深入钻研，迅速成为了国内数据挖掘分析领域的实战性专家。

一个公司最重要的资产就是他的员工，特别是那些勤于思考，能以自己的思考能力和职业贡献不断定义、刷新岗位价值的关键员工。泽中正是如此。他心有所属，沉稳安静，不浮躁不喧哗，始终像一个沉稳的驭手驾驭着自己的工作和生活，在安静沉稳的表象下显示着举重若轻的自信和大气。

数据的挖掘和分析是科学时代的产物，它的真理性依赖于事实数据本身的质地和分析过程的科学严密，而不能诉诸感性论辩的激情和机锋。和泽中接触，我能感觉到他在数据挖掘分析工作上的先天优势，也能体会到长期游弋于理性数据世界对他人格风采养成的直接影响。

我相信，这样一本畅销著作的出版，对他的人生来说，肯定是一件具有里程碑意义的大事；但在领导和同事看起来，这本书完全像是某种完美计划下瓜熟蒂落的圆满产物，固然令人惊喜，但更让人觉得顺理成章，与他的职业形象十分吻合。

泽中的书是一本通俗的、工具色彩浓厚的统计学专著，与我看过的一本美国原版通俗统计学专著有异曲同工之妙。

这就是查尔斯·韦兰教授写的《赤裸裸的统计学（naked statistics）》，其副标题是“stripping the dread from the data（让数据不再恐怖）”。这是一本深受欢迎的通俗性统计学读物，也是我自己离开大学课堂后，涉猎得不多的统计学书籍之一。《经济学人（the economist）》对该书的评价比作者在副标题中自谦的表达更加热情洋溢：他的书让统计学生动有趣，在对事物化繁为简后，大千世界潜在的迷人魅力展现。（He makes statistics interesting and fun. His book strips the subject of its complexity to expose the sexy stuff underneath.）

我不知道泽中是否看过这本书，但很显然，两本书在写作方法上不谋而合，这不止是一种技巧，而是显示了作者极大的诚意和融会贯通之功。和韦兰教授一样，为了激活读者的兴趣点，泽中努力提炼和展现生活本身的精彩，在他们的笔下，让包括我在内的许多人望而生畏的、以数学作为基础工具的统计科学得到了彻底的“善意”包装，显得精彩纷呈又平易近人。这里有你的事业炼金术，如怎样成为一个优秀的面包店长；这里有你的爱情阅人术，如最佳男友模型的建立；这里还能教你怎样做一个精明的消费者，如怎样识破手机绑定消费的秘密。

面对这些最接地气、接人气的事例，即使如我这样典型的文科男的确也会对迷人的通俗统计学，对不管是叫数据挖掘分析还是叫数据魔术的这样一门知识和技能变得跃跃欲试，不管是“涨姿势”还是长能力，这样的书都对我们有着开卷有益的强烈暗示。

当然，我们每个人被这样的数据能力所吸引和折服其实并不奇怪（这应该也能解释为什么在我这个文科男的书架上能轻易找到这本通俗统计书），我在近期阅读的另一本畅销书，《人类简史：从动物到上帝》中读到的一些论断不失为很好的解释。

作者写道：

我们的祖先很希望了解这个世界，投入大量时间和精力希望能找出支配自然界的法则。人类建立的现代科学和先前的知识体系有很大不同：第一：愿意承认自己的无知；第二：以观察和数学为中心；第三、取得新能力。科学革命不是知识的革命，而是无知的革命，真正让科学革命起步的重大发现，就是发现“人类对最重要的问题其实毫无所知”。而对于像是宗教这些前现代知识来说，他们假设世界上所有重要的事情，都已经为人或神所知。这些全知者可能是某些过去的智者，某个全能的神，或者是某些伟大的神，通过经典或者口传，将这些知识传给后人。而对于普通人而言，需要做的就是钻研这些古籍和传统却加以理解。在当时，如果说，《圣经》《古兰经》或者《吠陀经》居然遗漏了某些宇宙的重大秘密，而这个秘密居然能被一般的肉眼凡胎发现，这简直是不可思议的事情。

现代科学没有需要严格遵守的教条，但研究方法有一个共同的核心：收集各种实证观察，并以数学工具整理。早期的知识体系，用故事构造理论，而现代科学则用数学。传统的神话和经典，讲到的法则都用语言叙述，而不用数学公式。

各种宗教当然仍然是现代社会不可小觑的精神力量，但作为一种认识工具，我们显然不管怀有什么样的宗教情怀，都还是要把自己归属于现代科学的大旗之下。除非彻底厌弃，否则对于这个世界我们一定还是保留着旺盛的好奇，因而对数据能力的向往总是在某些时刻激励着我

们在某个时刻打开类似这样的一本书。

当然更不可否认的是，是计算机技术的发展使我们对技术能力怀有的好奇和向往找到了实现的可能（泽中本身数据挖掘的技术岗位也是这样的产物），强大的计算机统计分析软件解放了我们中的许多人对繁复的数学知识、术语、公式所怀的天然恐惧，大大降低了我们获得数据挖掘分析能力和从事数据分析活动的门槛。在公司组织一次培训课堂上，谈到统计分析软件的应用时，六西格玛专家夏老师的一席话让我印象很深，他说，在现在这样一个技术手段门槛如此之低的时代，脱离数据分析而谈论管理是一种悲哀。

泽中当时也在课堂之上，我相信，对夏老师的这句话，他一定比我感悟更深。在我看来，他的这本书也正是这样一种基于管理科学共识的宝贵实践。

在强大的统计分析软件帮助下，在模拟的对话式氛围之下，带着每个人在人生中最有同鸣感的那些有趣话题，按着书籍中的泽中老师的指导，你轻轻地敲击键盘按键，关于工作、生活的那些科学认识就从这神奇的屏幕上涌现。这些就是你在自己的工作生活世界中创造的新知识，至少其中的某些部分，将成为你对这个世界的独特贡献。

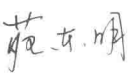
这真是一个神奇的过程，让你如此直观地感受到自己知识和能力的建立和深化。

以泽中自己为例，我也相信他必将以自己高超的数据能力和数据研究发现，为公司改善管理做出杰出贡献，从而不断证明自身对公司的卓越价值。

当笔者写这篇序言的时候，大到国家小到企业和我个人，正在因为发生在天津港的安全事故，而对化工企业的安全生产忧心满怀。安全、环保是化工企业的最大风险和发展约束，也是对国家人民所付的最大责任。我希望泽中也能更加注重对安全环保数据的深入挖掘和分析，使公司安全和环保风险的管理控制更加准确、更加理性。

我相信打开这本书的读者，一定是对工作、生活有追求、有想法的人，按照这本书的帮助掌握数据统计分析工具，提高自己的认识水平和分析研究能力，不管是对在生活中实现个人情怀、还是在岗位上实现职业梦想都会大有裨益。

再次祝贺泽中，能为读者奉献出这样一部值得一读的诚意之作，希望他能继续在数据挖掘分析领域阔步前行！



2015-8-25 于宁波

序　言

笔者自 2008 年的一个偶然机会第 1 次接触“数据挖掘”（Data Mining）这个新名词以来，在数据挖掘应用相关领域度过了 6 年。笔者的专业是化工，整天应该与塔、釜、换热器、化学反应和物料守恒等打交道。开始接触这个专业的目的是为了利用数据分析的一些功能来优化生产运营，让企业以更高的效率、更低的成本和更好的质量运营，为此需要数据积累、数据分析和数据模型。

2008 年，国内企业在数据挖掘应用中摸索起步，远不如现在大数据火热。如今大数据最火的商业应用主要集中在互联网、银行和电信等领域。基于行业应用限制，笔者无法接触到真正的大数据挖掘，但是幸运的是还是碰到了职业和兴趣的重合点。

这几年的摸索是笔者职业生涯中很重要的一段时光，因此有必要将自己一路走来的心得与体会、感悟和挫折整理出来，一则是对自己的这段职业生涯做一个交代，特别是对一路引导、鼓励和支持笔者的师友和家人；二则是合理地引导类似笔者半道出家的学习者，对数据分析有兴趣却没有深厚的统计学知识和 IT 功底人士，笔者相信本书的内容对于广大对数据分析应用感兴趣的初学者来说都是一种宝贵经验。在学习数据分析的道路上笔者深刻认识到一个道理，即一个成功的数据分析实践的核心因素不是数据分析技术，而是对业务理解和分析思路。这也是当初学习数据分析的初衷，初学者切不可为数据分析而分析数据。

大数据挖掘需要精通数据库、计算机编程和深厚的统计学基础，有的甚至涉及运筹学范畴，是一门复合型的应用科学。大数据的案例现在是一抓一大把，如国外典型的“啤酒与尿布”的案例，在了解数据分析之前不妨来看看几个有趣的应用案例。

（1）数据新闻让英国撤军

2010 年 10 月 23 日《卫报》利用维基解密的数据做了一篇“数据新闻”，即将伊拉克战争中所有的人员伤亡情况均标注于地图之上，地图上一个红点代表一次死伤事件。用鼠标单击红点后弹出的窗口则有详细的说明，包括伤亡人数、时间和造成伤亡的具体原因。密布的红点多

达 39 万个，显得格外触目惊心，如图 0-1 所示。此新闻一经刊出立即引起朝野震动，推动英国最终做出撤出驻伊拉克军队的决定。

图 0-1　伊拉克战争中所有的人员伤亡情况

（2）大数据与乔布斯癌症治疗

乔布斯是世界上第 1 个对自身所有 DNA 和肿瘤 DNA 进行排序的人，为此他支付了高达几十万美元的费用。他得到的不是样本，而是包括整个基因的数据文档。医生按照所有基因按需下药，最终这种方式帮助乔布斯延长了几年的生命。

（3）Google 成功预测冬季流感

2009 年，Google 通过分析 5000 万条美国人最频繁检索的词汇将其和美国疾病中心在 2003—2008 年间季节性流感传播时期的数据进行比较。并建立了一个特定的数学模型，最终成功预测了 2009 冬季流感的传播，甚至可以具体到特定的地区和州。

（4）奢侈品销售

PRADA 在纽约的旗舰店中每件衣服上都有 RFID 码，每当一个顾客拿起一件 PRADA 进入

序　言

笔者自 2008 年的一个偶然机会第 1 次接触“数据挖掘”（Data Mining）这个新名词以来，在数据挖掘应用相关领域度过了 6 年。笔者的专业是化工，整天应该与塔、釜、换热器、化学反应和物料守恒等打交道。开始接触这个专业的目的是为了利用数据分析的一些功能来优化生产运营，让企业以更高的效率、更低的成本和更好的质量运营，为此需要数据积累、数据分析和数据模型。

2008 年，国内企业在数据挖掘应用中摸索起步，远不如现在大数据火热。如今大数据最火的商业应用主要集中在互联网、银行和电信等领域。基于行业应用限制，笔者无法接触到真正的大数据挖掘，但是幸运的是还是碰到了职业和兴趣的重合点。

这几年的摸索是笔者职业生涯中很重要的一段时光，因此有必要将自己一路走来的心得与体会、感悟和挫折整理出来，一则是对自己的这段职业生涯做一个交代，特别是对一路引导、鼓励和支持笔者的师友和家人；二则是合理地引导类似笔者半道出家的学习者，对数据分析有兴趣却没有深厚的统计学知识和 IT 功底人士，笔者相信本书的内容对于广大对数据分析应用感兴趣的初学者来说都是一种宝贵经验。在学习数据分析的道路上笔者深刻认识到一个道理，即一个成功的数据分析实践的核心因素不是数据分析技术，而是对业务理解和分析思路。这也是当初学习数据分析的初衷，初学者切不可为数据分析而分析数据。

大数据挖掘需要精通数据库、计算机编程和深厚的统计学基础，有的甚至涉及运筹学范畴，是一门复合型的应用科学。大数据的案例现在是一抓一大把，如国外典型的“啤酒与尿布”的案例，在了解数据分析之前不妨来看看几个有趣的应用案例。

（1）数据新闻让英国撤军

2010 年 10 月 23 日《卫报》利用维基解密的数据做了一篇“数据新闻”，即将伊拉克战争中所有的人员伤亡情况均标注于地图之上，地图上一个红点代表一次死伤事件。用鼠标单击红点后弹出的窗口则有详细的说明，包括伤亡人数、时间和造成伤亡的具体原因。密布的红点多

试衣间，RFID 会被自动识别；同时数据会传至 PRADA 总部。每一件衣服在哪个城市、哪个旗舰店、什么时间被拿进试衣间和停留多长时间，数据都被存储起来加以分析。如果一件衣服销量很低，以往的做法是直接收回；如果 RFID 传回的数据显示这件衣服虽然销量低，但进试衣间的次数多，则说明这件衣服的下场会截然不同，或者在某个细节的微小改变就会重新制造出一件非常流行的产品。

除了国外这些经常用于商业培训课程的案例外，数据分析其实并不遥远，在国内也不乏应用。例如，共和国的开国元帅林彪就曾经依靠敏锐的数据嗅觉和军事天赋成功捣毁敌营总部。

目前国内的大部分高校还没有开设数据挖掘这门专业课程，大数据分析需要依靠庞大的数据库，即需要各专业的人士通力合作，是一个团队作业。类似笔者这种半道出家的个人学习者在不具备团队协作的条件下，可以在样本数据的分析下工夫，样本数据也可以称为“小数据”，因此本书的名称定为《大数据时代的小数据分析》。

本书主要介绍应用数据分析的一系列工具，如：Excel、LINGO、Crystal Ball、JMP、Minitab 和 Xcelsius 等，涉及的分析有预测、风险分析、优化求解、假设检验、相关分析、回归分析和聚类分析等。但所有这些软件都不是最新版本，如 Excel 使用 2010 版；Minitab 使用的 V15 版。在使用软件时最重要的不是版本的最新，而是理解其功能和特点，灵活地运用。即使是 Excel 2003 版本，只要运用得当，同样能发挥强大的功能。很多不同功能的软件都可以完成，本书主要结合不同软件的不同特点介绍其应用。

书中涉及一些专业名词和原理，如标准差和假设检验等，本书没有给出生涩难懂的定义，而只是通俗地解释这些名词。这样做原因有二：一则作为半道出家的笔者不愿，也不会定义这些理论；二则定义这些名词或原理只会让本来就让人头疼的数据分析显得更加枯燥。如果读者需要准确理解这些专业名词，可以参考其他资料。

本书中列举的一些应用都是尽可能地贴近生活和工作，让数据分析看起来尽可能有趣一些，在排列各章节的顺序时也尽量遵循软件的功能之间的逻辑关系。

本书在每一章均会应用一些有趣的案例引出讨论的重点，其中两人按照师徒问答的形式模拟实际工作中的场景循序渐进地学习分析工具，让枯燥的数据分析显得生动一些。

本书适合的读者如下。

（1）对数据分析应用有兴趣的人士。

（2）对统计、数学和码农等深奥理论不感兴趣者。

（3）想尝试自身专业的数据分析，提高技能者。

（4）想尝试数据分析工作并寻找切入点者。

本书不适合的读者如下。

（1）喜欢拍脑袋和胸脯者。

（2）见了数据就想呕吐者。

（3）爱好 SAS/R/Python 等豪门软件的狂热者。

（4）统计、数学和 IT 专业的大牛。

（5）对数据有深刻理解的科学家。

笔者是从化工这个与数据分析无关的专业开始学习数据分析的，相信只要读者能静心地读完本书也会有所收获。但是不能指望数据分析能解决所有的问题，它不是万能的。一个成功的数据分析实践的核心因素不是数据分析技术，而是对业务的理解和分析思路。

全书的原理讲解和工具操作同步，即在操作软件的同时理解其原理；列举的案例涵盖多个行业，根据案例引出所需要讨论的知识点；然后根据知识点举一反三，串联尽可能多的数据分析入门知识；同时将介绍其适合的分析工具。部分案例来自软件自带案例。

在编写本书之前笔者与人大经济论坛（http://bbs.pinggu.org/）合作开发过相关的视频培训课件，其中部分工具与本书中介绍的工具相同，有需要视频课件的读者可以试听（前 3 节免费）。

（1）Crystal Ball 初中级课程：http://www.peixun.net/view/208.html。

（2）Crystal Ball 高级课程：http://www.peixun.net/view/216.html。

（3）LINGO 初级课程：http://www.peixun.net/view/251.html。

（4）Minitab 初级课程：http://www.peixun.net/view/281.html。

由于笔者的水平有限，对数据分析的理解不够透彻，加之编写时间仓促，因此书中难免会出现一些错误或不准确之处，恳请读者批评指正。

本书配套资源下载链接：www.broadview. com.cn/26469。

目　录

第 1 章　知己知彼，百战不殆——风险与预测分析 / 1

1.1　预测从世界杯开始 / 2

1.2　手机绑定消费的秘密 / 5

1.3　笔记本电脑出国冒险记 / 25

1.4　慧眼识分布 / 36

1.5　分布 72 变 / 47

1.6　做最优秀的面包店长 / 74

第 2 章　运筹帷幄，决胜千里——效益最大化 / 101

2.1　换个思路来数鸡 / 102

2.2　做一个精明的农场主 / 128

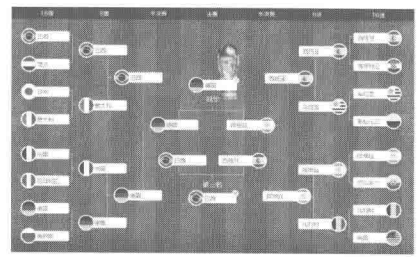
2.3　见识 LINGO 与 Crystal Ball 的威力 / 146

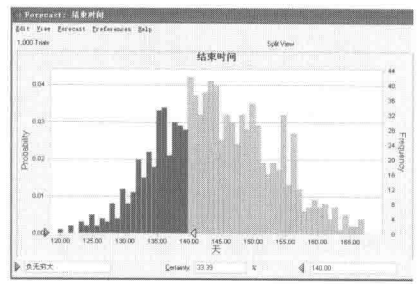
第 3 章　图个明白，精彩展现——JMP 精彩图表 / 192

3.1　图个明白——常用图形 / 194

3.2　图个明白——树图 / 208

3.3　图个明白——SPC 图 / 214





第4章 抽丝剥茧，明察秋毫——相关分析 / 227

4.1 假设检验——大胆假设，小心求证 / 228

4.1.1 小心求证——均值检验 / 235

4.1.2 小心求证——比例检验 / 252

4.1.3 小心求证——非参数检验 / 261

4.2 相关与回归分析 / 272

4.2.1 相关性与第三方变量 / 272

4.2.2 收入与支出关系——简单线性回归 / 280

4.2.3 最佳口感食品配方——多元线性回归 / 283

4.2.4 咖啡好喝，不能多喝——非线性回归 / 290

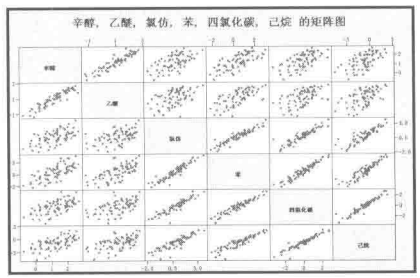
4.2.5 预防心血管疾病从减肥开始——二值 Logistic 回归分析 / 295

4.3 人以类聚，物以群分——聚类分析 / 300

4.3.1 美好一天从早餐开始——观测值聚类分析 / 302

4.3.2 海拔是否影响血压——变量聚类分析 / 305

4.3.3 为熊猫分类——K 均值聚类分析 / 307

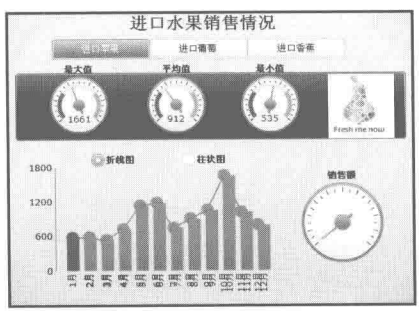


第5章 要里子，也要面子——数据展现的艺术 / 311

5.1 哪种水果更好卖 / 314

5.2 书店利润最大化 / 327

5.3 非诚勿扰——最佳男友模型 / 337



第1章

知己知彼，百战不殆
风险与预测分析

1.1 预测从世界杯开始

1.2 手机绑定消费的秘密

1.3 笔记本电脑出国冒险记

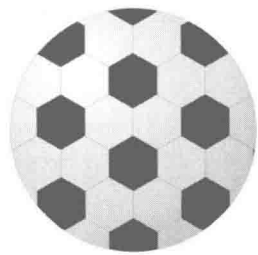
1.4 慧眼识分布

1.5 分布72变

1.6 做最优秀的面包店长

1.1 预测从世界杯开始

预测一般根据以往发生的事情来推断即将发生事情的风险概率，它与风险如影随形，即根据已知的风险来推测未知的风险。



【场景再现】

Mr Shu 和 Miss Ju 是一家公司运营分析部的职业数据分析师，前者是一名入职多年精通业务的资深数据分析师；后者则是职场新人，对数据分析有强烈的爱好。但由于在学校期间未接受过专业数据分析培训，因此 Mr Shu 负责对 Miss Ju 进行入职培训。

Mr Shu 和 Miss Ju 也都是资深球迷，自然不会错过 2014 年世界杯夺冠预测这样的练习机会。

Mr Shu：“2014 年世界杯即将开战，你觉得谁的夺冠概率比较大？”

Miss Ju：“我个人超级喜欢德国队的硬朗风格，但 2014 年的举办地在五星上将巴西的主场，可能南美国家的机会会更大一些吧。我们球迷完全是在这里拍脑袋，没有数据支撑，完全是靠猜测啦，你觉得呢？”

Mr Shu：“直观感觉也很重要吧？不仅是我们球迷在这里拍脑袋凑热闹，专门的投资公司也是不会放过世界杯这样的捞金机会，著名的投资公司高盛公司就对 2014 年的世界杯进行其模型分析。”

Miss Ju：“真的吗？这种巨无霸分析出来的应该算有理有据吧？分析的结果怎么样？”

Mr Shu：“高盛在推出世界杯报告预测之余，围绕历届赛事对经济和股市的影响大做文章，发表了一份长达 60 页的分析报告。根据该行的统计模型，4 强将为巴西、德国、阿根廷和西班牙，决赛则是巴西和阿根廷之争，主办国胜。”

Miss Ju：“哇，如此 TIPS，实在大路货！这个预测就好比一年一度的香港国际赛马日，评分最高的 4 匹马顺序归来。既无惊喜，更乏惊吓。从投机角度出发毫无刺激性可言，有什么数据支撑吗？”

Mr Shu：“高盛的经济学家通过建立数据模型分析了自 1960 年以来超过 14 000 场国际比赛，最终得出了本届世界杯的预测结果。世界杯五冠王巴西在家门口捧得第 6 座金杯的可能性高达 48.5%，而名列第 2 的则是桑巴军团的宿敌阿根廷。不过，潘帕斯雄鹰夺冠的几率为 14.1%，几乎只是巴西夺冠几率的 1/3 而已。”

Miss Ju：“那按照高盛的分析，巴西应该胜券在握了吧？不过好像也没什么惊喜，巴西本来就是夺冠大热门。”

Mr Shu：“是的，这份报告的撰写人，即高盛首席经济学家也表示：‘当然，这个结论一点也不令人惊讶。作为两支足球史上最成功的球队，巴西和阿根廷杀入世界杯决赛实至名归。但巴西在模型预测中具有如此巨大的优势还是多少让我们感到惊讶。’巴西几乎是阿根廷的 3 倍呢。”

Miss Ju：“其他强队的夺冠概率怎么样？”

Mr Shu：“南美双雄之外，德国队成为最大热门。其捧杯几率为 11.4%，成为欧洲球队中最具冠军相的球队；西班牙（9.8%）和荷兰（5.6%）位居这份榜单的第 4 名与第 5 名；喀麦隆、阿尔及利亚和洪都拉斯则被高盛认为夺冠完全不可能。”

Miss Ju：“作为德国的铁杆球迷，我认为德国自然是大热门。但同为欧洲强队的英国和意大利呢？要知道意大利也是四星上将啊。”

Mr Shu：“贵为欧洲传统豪门的英格兰队被高盛认为只有 1.4%的几率最终捧杯，高盛分析英格兰最终将会成为小组第 3 无法出线。并将一场不胜灰溜溜地离开巴西，而同组中晋级下一轮的将是意大利和乌拉圭。不过英格兰小组赛同组对手意大利的夺冠几率也只不过是 1.5%，与三狮军团难分伯仲。”

Miss Ju：“说的好像有点道理，这样说我也觉得巴西夺冠的可能性极大。”

Mr Shu：“当然高盛这种基于过往比赛的数据模型预测并非万无一失，4 年前的南非世界杯前该公司同样预测巴西是最大热门，夺冠几率高达 26.6%。但球队在 1/4 决赛就被淘汰，最终捧杯的是高盛当时预测的第 2 热门西班牙。”

Miss Ju：“还有持不同意见的分析结果吗？”

Mr Shu：“当然也有持不同意见的预测模型，如著名的风险分析公司@Risk 也做出了同样的预测模型，但模拟出来的结果与高盛公司分析的概率不太一样。”

Miss Ju：“@Risk 分析出来的结果怎么样？”

Mr Shu：“@Risk 利用数学模型分析的结果同样是巴西夺冠，但是结果比高盛的要谨慎很多。巴西的夺冠概率第 1，为 17%；概率第 2 则为西班牙的 12%。而接下来的 6 强分别是瑞士（8%）、希腊（8%）、德国（7%）、哥伦比亚（7%）、阿根廷（6%）和乌拉圭（5%）。如果不考虑主场因素，德国队的夺冠概率最大为 19.9%。”

Miss Ju：“Mr Shu 你有自己的预测吗？”

Miss Ju：“当然，不然怎么能自称资深球迷，看看图 1-1 就知道了。”

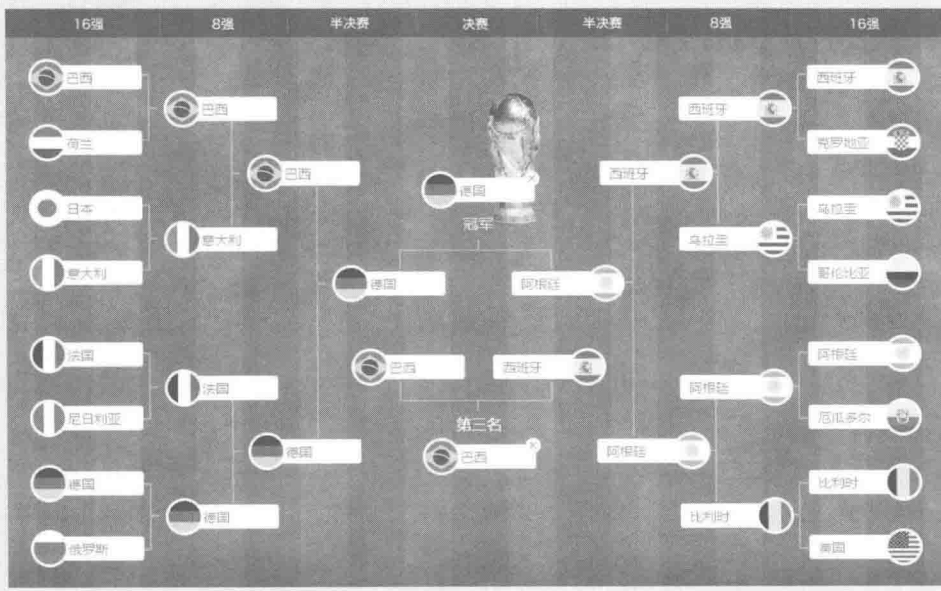


图 1-1　世界杯夺冠预测

Miss Ju：“看来每个人的分析都不太一样嘛，这主要是看中的影响因素不一样。还是章鱼哥保罗最厉害，准确率达到 92%。”

Mr Shu：“当然每场球任何一种结果的准确的概率都有 1/3，这种预测个人认为都是纸老虎，不靠谱。”

Miss Ju：“好像预测玩的就是概率嘛。”