

第1章

数据分析那些事儿

数据分析是“神马”

数据分析六步曲

几个常用指标或术语



>> 谁说菜鸟不会数据分析

出场人物：

牛董，关键词：私企董事长、要求严格、为人苛刻；

小白，关键词：应届毕业生，刚入职场的白骨精（白领+骨干+精英）、牛董助手、爱臆想；

Mr.林，关键词：小白同事、数据分析达人、成熟男士、乐于助人、做事严谨。

话说小白过五关斩六将，通过严格的面试，最终从众多优秀毕业生中脱颖而出，成为公司的一员。在报到的第一天，公司HR向小白介绍了她的职位——公司牛董的助理，负责文秘工作，可能需要做一些数据分析之类的活儿。小白一听到数据分析这个词，就感觉头皮发麻，这时，她的脑子里幻想出一些穿白大褂的科研人员在实验室的计算机前不断忙碌的场景。虽然在上大学时也使用过Excel，但是如果要做数据分析工作，她还真的有些不知如何下手。无数个问号涌到她的脑海中：数据分析到底要做什么呢？我要怎么做数据分析？老板想要看什么样的结果呢？……唉，只好边走边干了。

HR看出小白的心事，说道：关于数据分析你不用太担心，如果遇到难题，你可以请教我们公司的Mr.林，他在这方面可是专家喔！

小白一听有救星，立马兴奋起来，好像抓到救命稻草一样，想赶紧找到这位大师级人物，然后一股脑把疑问全倒出来。

HR：小白，你跟我来吧，我给你引荐下Mr.林。

小白：好的。

说着小白跟HR来到了Mr.林的办公桌旁，HR说道：Mr.林，这是我们公司新来的同事，叫小白，现担任牛董助理，她的部分工作涉及数据分析，到时候有问题还要麻烦您多多指点了。

小白紧跟着说道：Mr.林，您好，请多多指教。

Mr.林：呵呵，太客气了，有问题直接来找我就可以了。

小白趁机说道：我现在就有问题，您现在是否有时间帮我解答下？数据分析是干什么的，具体要怎么做？

Mr.林听完后，笑了起来：你还真是不客气呀，好吧，你刚进公司，我就先给你做个简单的培训，带你入个门吧，以后的修行可就靠你自己努力了。

小白用力点着头，HR见这“师徒”二人颇有一见如故之感，大致交代一番后，就离开了。

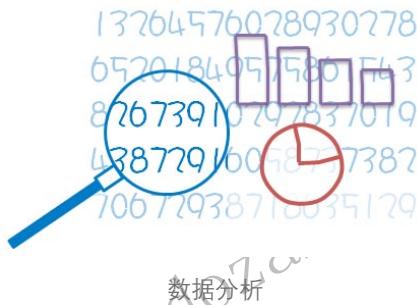
1.1 数据分析是“神马”

Mr.林接着说道：小白，既然要学习数据分析，借用现在的网络流行语，你能说说数据分析是“神马”么？

小白：呵呵，数据分析肯定不是“浮云”。从字面上理解，数据分析就是对数据进行分析。Mr.林，我只能理解到这一层面了，专业的解释就需要您来指教了。

1.1.1 何谓数据分析

Mr.林：简单来说，数据分析就是你刚才讲的——对数据进行分析。较为专业的说法，数据分析是指用适当的统计分析方法对收集来的大量数据进行分析，将它们加以汇总、理解并消化，以求最大化地开发数据的功能，发挥数据的作用。数据分析是为了提取有用信息和形成结论而对数据加以详细研究和概括总结的过程。这里数据也称观测值，是通过实验、测量、观察、调查等方式获取的结果，常常以数量的形式展现出来。



数据分析的目的是把隐藏在一大批看似杂乱无章的数据背后的信息集中和提炼出来，总结出研究对象的内在规律。在实际工作当中，数据分析能够帮助管理者进行判断和决策，以便采取适当策略与行动。例如，如果企业的高层希望通过市场分析和研究，把握当前产品的市场动向，制订合理的产品研发和销售计划，就必须依赖数据分析才能完成。

在统计学领域，有些学者将数据分析划分为描述性数据分析、探索性数据分析以及验证性数据分析（参见图1-1）。其中，探索性数据分析侧重于在数据之中发现新的特征，而验证性数据分析则侧重于验证已有假设的真伪性。

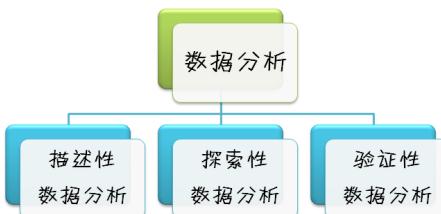


图1-1 数据分析类别

从另一个角度看，描述性数据分析属于初级数据分析，常见的分析方法有对比分析法、平均分析法、交叉分析法等；而探索性数据分析以及验证性数据分析属于高级数据分

>> 谁说菜鸟不会数据分析

析，常见的分析方法有相关分析、因子分析、回归分析等。我们日常学习和工作中涉及的数据分析方法主要是描述性数据分析，也就是大家常用的初级数据分析。

1.1.2 数据分析的广阔前景

Mr.林：数据分析作为一个新的行业领域正在全球迅速发展，它开辟了人类获取知识的新途径。

目前，数据库技术、软件工具、各硬件设备飞速发展，在这些软硬件技术与设备的支持下，信息技术应用已在各行各业全面展开，尤其是对通信、互联网、金融等行业的发展做出了巨大贡献，并且经过长期的应用积累大量丰富的数据。但大部分企业对其存储信息利用率极低。庞大的历史数据是否有价值？有何价值？是否可以综合利用分析？是否能够为领导决策提供参考依据？

回答是肯定的，数据分析这一项工作越来越受到领导层的重视，借助数据分析的各种工具从海量的历史数据中提取、挖掘对业务发展有价值的、潜在的知识，找出趋势，为决策层的决策提供有力的依据，对产品或服务的发展方向起到积极作用，有力推动企业内部的科学化、信息化管理。

从20世纪90年代起，欧美国家开始大量培养数据分析师，直到现在，对数据分析师的需求仍然长盛不衰，而且还有扩展之势。根据美国劳工部预测，到2018年，数据分析师的需求量将增长20%。就算你不是数据分析师，但数据分析技能也是未来必不可少的工作技能之一。

小白：数据分析师如此抢手的原因何在呢？

Mr.林：一个简单的原因就是社会越发达，人们对数据的依赖就越多。无论政府决策还是公司运营，科学研究还是媒体宣传，都需要数据支持。那么，对数据有如此大的依赖，就必然导致对数据分析的大量需求。因此，将数据转化为知识、结论和规律，就是数据分析的作用和价值。

小白：那数据究竟会庞大到什么地步呢？

Mr.林：据国际知名咨询公司估计，到2020年，全球每年产生的数据量将达到3500万亿GB，打个比方，就是用普通的DVD一张一张地摞起来，可以从地球摞两个堆一直到月球。



全球每年产生的数据量

面对这样庞大的数据，数据分析师的职责就不仅仅是单纯的分析了，更重要的是与相关部门进行合作，将数据真正应用到业务中，根据实际的业务发展情况识别哪些数据可用，哪些不适用，而不是孤立地在“真空环境”下进行分析。这就要求数据分析师不仅具备洞察数据的能力，还要对相关业务的背景有深入的了解，明白客户或业务部门的需求，从而将数据信息化、可视化，最后转化为生产力，帮助企业获得利润。这就是整个数据“供应链”。当然数据分析师也需要理解这个“供应链”。

小白：那怎样才能成为一名优秀的数据分析师呢？

Mr.林：学习数据分析需要时间和经验的积累，而不能一蹴而就。在工作中运用不同的分析方法对数据进行分析，并与业务部门同事积极沟通，加深自己对整个行业或研究内容的理解，相信在两到三年内，一个优秀的数据分析师就会诞生。

1.1.3 菜鸟与数据分析师

小白听完Mr.林对数据分析的介绍后，问道：数据分析这么有用，那成为一个优秀的数据分析师要具备哪些素质呢？以我现在的能力可以做数据分析师吗？

Mr.林看了看小白说：小白，先别着急。我们先来看看菜鸟与数据分析师之间的差别。刚好我这儿有张菜鸟与数据分析师想法对比图（参见图1-2），你可以对比一下，看看自己在哪些方面还是菜鸟，在哪些方面已经达到了数据分析师的水平。从他们思考问题的不同方式，就能发现为什么一个是菜鸟，而另外一个是数据分析师了。

菜鸟与数据分析师的区别

菜鸟会想	分析师会想
这曲线图真好看，怎么作的？	数据变化的背后真相是什么？
这数据可以做什么样的分析？	从哪些角度分析数据才系统？
高级的分析方法在这能用吗？	用什么分析方法最有效？
要作多少张图表？	图表是否表达出有效的观点？
数据文字说明外还需说什么？	数据分析的目的达到了吗？
数据分析报告要写多少页？	数据分析报告有说服力吗？
.....

图1-2 菜鸟与数据分析师的区别

Mr.林：小白，从菜鸟与数据分析师想法对比图中，你发现菜鸟与数据分析师有什么区别了吗？

>> 谁说菜鸟不会数据分析

小白：我发现，菜鸟很迷茫，目的不明确，而且一味追求高级的分析方法。而数据分析师的目的就很明确，一切都以解决问题为中心。

Mr.林：说得好！其实他们最主要的区别就在于目的是否明确，如果目的明确，所有问题就自然迎刃而解了。例如，分析师是不会考虑“需要多少张图表”这样的问题的，而是思考这个图表是否有效表达了观点？如果没有，需要怎样调整？如果目的不明确，菜鸟自然会在用什么样的分析方法、作多少张图表、需要多少文字说明、报告要写多少页等这些问题上纠结。

小白：我明白了，我不要做菜鸟，我要朝着数据分析师的方向努力。

1.1.4 数据分析师的基本素质

Mr.林看到小白斗志昂扬的样子非常高兴：别光说不做啊，要成为一名优秀的数据分析师，并非一件容易的事。虽然所学的专业与数据分析不相关，但你可以通过工作中的实践学习数据分析，需要付出大量的时间和精力，不经一番寒彻骨，怎得梅花扑鼻香？

下面，我给你介绍一名合格的数据分析师需要具备的五大基本能力和素质。

◎ 态度严谨负责

严谨负责是数据分析师的必备素质之一，只有本着严谨负责的态度，才能保证数据的客观、准确。在企业里，数据分析师可以说是企业的医生，他们通过对企业运营数据的分析，为企业寻找症结及问题。一名合格的数据分析师，应具有严谨、负责的态度，保持中立立场，客观评价企业发展过程中存在的问题，为决策层提供有效的参考依据；不应受其他因素影响而更改数据，隐瞒企业存在的问题，这样做对企业发展是非常不利的，甚至会造成严重的后果。而且，对数据分析师自身来说，也是前途尽毁，从此以后所做的数据分析结果都将受到质疑，因为你已经不再是可信赖的人，在同事、领导、客户面前已经失去了信任。所以，作为一名数据分析师就必须持有严谨负责的态度，这也是最基本的职业道德。



◎ 好奇心强烈

好奇心人皆有之，但是作为数据分析师，这份好奇心就应该更强烈，要积极主动地发现和挖掘隐藏在数据内部的真相。在数据分析师的脑子里，应该充满着无数个“为什么”，为什么是这样的结果，为什么不是那样的结果，导致这个结果的原因是什么，为什么结果不是预期的那样等等。这一系列问题都要在进行数据分析时提出来，并且通过数据分析，给自己一个满意的答案。越是优秀的数据分析师，好奇心也越不容易满足，回答了一个问题，又会抛出一个新的问题，继续研究下去。只有拥有了这样一种刨根问底的精神，才会对数据和结论保持敏感，继而顺藤摸瓜，找出数据背后的真相。



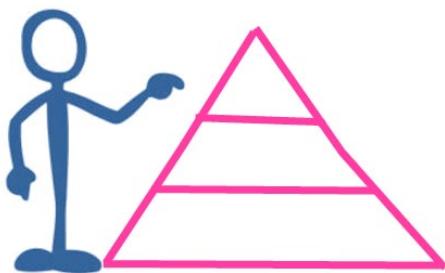
好奇心强烈

◎ 逻辑思维清晰

除了一颗探索真相的好奇心，数据分析师还需要具备缜密的思维和清晰的逻辑推理能力。我记得有位大师说过：结构为王。何谓结构，结构就是我们常说的逻辑，不论说话还是写文章，都要有条理，有目的，不可眉毛胡子一把抓，不分主次。

通常从事数据分析时所面对的商业问题都是较为复杂的，我们要考虑错综复杂的成因，分析所面对的各种复杂的环境因素，并在若干发展可能性中选择一个最优的方向。这就需要我们对事实有足够的了解，同时也需要我们能真正理清问题的整体以及局部的结构，在深度思考后，理清结构中相互的逻辑关系，只有这样才能真正客观地、科学地找到商业问题的答案。

>> 谁说菜鸟不会数据分析



逻辑思维清晰

◎ 擅长模仿

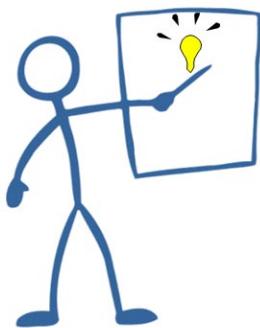
在做数据分析时，有自己的想法固然重要，但是“前车之鉴”也是非常有必要学习的，它能帮助数据分析师迅速地成长，因此，模仿是快速提高学习成果的有效方法。这里说的模仿主要是参考他人优秀的分析思路和方法，而并不是说直接“照搬”。成功的模仿需要领会他人方法精髓，理解其分析原理，透过表面达到实质。万变不离其宗，要善于将这些精华转化为自己的知识，否则，只能是“一直在模仿，从未超越过”。



擅长模仿

◎ 勇于创新

通过模仿可以借鉴他人的成功经验，但模仿的时间不宜太长，并且建议每次模仿后都要进行总结，提出可以改进的地方，甚至要有所创新。创新是一个优秀数据分析师应具备的精神，只有不断的创新，才能提高自己的分析水平，使自己站在更高的角度来分析问题，为整个研究领域乃至社会带来更多的价值。现在的分析方法和研究课题千变万化，墨守成规是无法很好地解决所面临的新问题的。



勇于创新

听到这里，小白就掰着手指头算自己符合几条优秀数据分析师的素质和能力。

Mr.林继续说道：这些素质能力不是说有就有的，需要慢慢培养形成，不能一蹴而就。

小白：……

1.2 数据分析六步曲

Mr.林：小白，刚才已经介绍了一名合格的数据分析师需要具备的五大基本能力和素质，现在我们就来看看如何进行数据分析吧！先问你一个问题，你知道数据分析是怎样一个过程吗？

小白一时语塞：这个……

Mr.林：不知道了吧！其实，数据分析过程主要包括6个既相对独立又互有联系的阶段。它们是：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、报告撰写等6步（参见图1-3）。

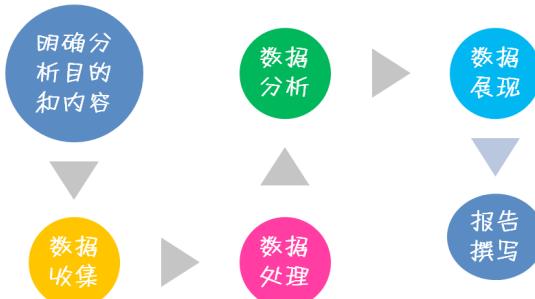


图1-3 数据分析六步曲

1.2.1 明确分析目的和内容

Mr.林：做任何事都要有目标的，数据分析也不例外。之前也说过，菜鸟与数据分析师的区别就在于菜鸟做分析时目的不明确，从而导致分析过程非常盲目。所以在开展数据分析之前，需要想想为什么要开展数据分析？通过这次数据分析我要解决什么问题？只有明确数据分析的目的，才不会偏离方向，否则得出的数据分析结果不仅没有指导意义，甚至可能将决策者引入歧途，后果严重。

当分析目的明确后，就需要把它分解成若干个不同的分析要点，也就是说要达到这个目的，需要从哪几方面、哪几个点进行分析，而这两点就是我们需要分析的内容。只有明确了分析目的，分析内容才能跟着确定下来。

明确数据分析的目的和内容是确保数据分析过程有效进行的先决条件，它可以为数据收集、处理以及分析提供清晰的指引方向。

小白：嗯，我记下了，在做数据分析前，先想想到底是为什么要做数据分析，以便明确数据分析方向，保证我不迷失方向。

1.2.2 数据收集

Mr.林：数据收集是按照确定的数据分析内容，收集相关数据的过程，它为数据分析提供了素材和依据。这里所说的数据包括第一手数据与第二手数据，第一手数据主要指可直接获取的数据，第二手数据主要指经过加工整理后得到的数据。一般数据来源主要有以下几种方式，如图1-4所示。



图1-4 数据的来源

◎ 数据库

每个公司都有自己的业务数据库，包含从公司成立以来产生的相关业务数据。这个业务数据库就是一个庞大的数据资源，需要有效地利用起来。

◎ 公开出版物

可用于收集数据的公开出版物包括《中国统计年鉴》《中国社会统计年鉴》《中国人口统计年鉴》《世界经济年鉴》《世界发展报告》等统计年鉴或报告。

◎ 互联网

随着互联网的发展，网络上发布的数据越来越多，特别是搜索引擎可以帮助我们快速找到所需要的数据，例如国家及地方统计局网站、行业组织网站、政府机构网站、传播媒体网站、大型综合门户网站等上面都可能有我们需要的数据。

◎ 市场调查

进行数据分析时，需要了解用户的想法与需求，但是通过以上三种方式获得此类数据会比较困难，因此可以尝试使用市场调查的方法收集用户的想法和需求数据。市场调查就是指运用科学的方法，有目的、有系统地收集、记录、整理有关市场营销的信息和资料，分析市场情况，了解市场现状及其发展趋势，为市场预测和营销决策提供客观、正确的数据资料。市场调查可以弥补其他数据收集方式的不足，但进行市场调查所需的费用较高，而且会存在一定的误差，故仅作参考之用。

小白：看来我以后做数据分析时，不用担心没有数据来源啦。

1.2.3 数据处理

Mr.林：数据处理是指对收集到的数据进行加工整理，形成适合数据分析的样式，它是数据分析前必不可少的阶段。数据处理的基本目的是从大量的、杂乱无章、难以理解的数据中抽取并推导出对解决问题有价值、有意义的数据。

数据处理主要包括数据清洗、数据转化、数据提取、数据计算等处理方法（参见图1-5）。一般拿到手的数据都需要进行一定的处理才能用于后续的数据分析工作，即使再“干净”的原始数据也需要先进行一定的处理才能使用。



图1-5 数据处理

小白：哦，数据处理是数据分析的前提，对有效数据进行分析才是有意义的，明白啦！

>> 谁说菜鸟不会数据分析

1.2.4 数据分析

Mr.林：数据分析主要是指通过统计分析或数据挖掘技术对处理过的数据进行分析和研究，从中发现数据的内部关系和规律，为解决问题提供参考。

在确定数据分析目的和内容阶段，数据分析师就应当为所分析的内容确定适合的数据分析方法。到了数据分析阶段，就能够驾驭数据，从容地进行分析和研究了。



数据分析

由于数据分析多是通过软件来完成的，这就要求数据分析师不仅要掌握各种数据分析方法，还要熟悉主流数据分析软件的操作。一般的数据分析我们可以通过Excel完成，后面也将重点为你介绍，而高级的数据分析就要采用专业的分析软件进行，如数据分析工具SPSS Statistics等。

小白：好的，期待您后续精彩的数据分析讲解。

◎ 数据分析与数据处理的区别

小白：Mr.林，我有个问题。我经常听到数据处理跟数据分析这两个词，它们之间有何关系呢？

Mr.林：从前面介绍数据处理和分析的过程，不难看出，数据处理是数据分析的基础。通过数据处理，将收集到的原始数据转换为可以分析的形式，并且保证数据的一致性和有效性。如果数据本身存在错误，那么即使采用最先进的数据分析方法，得到的结果也是错误的，不具备任何参考价值，甚至还会误导决策。

◎ 数据分析与数据挖掘的关系

小白：还有个问题，刚才提到数据挖掘，那么数据挖掘是什么呢？它与数据分析有什么关系呢？

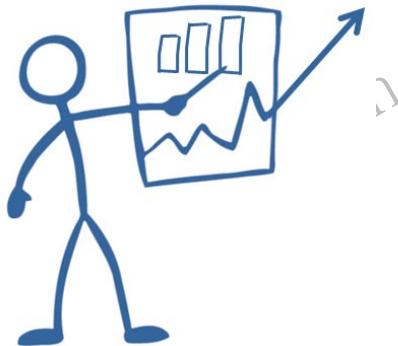
Mr.林：这个问题问得好！数据挖掘其实是一种高级的数据分析方法。数据挖掘就是从大量的数据中挖掘出有用的信息，它是根据用户的特定要求，从浩如烟海的数据中找出

所需的信息，以满足用户的特定需求。数据挖掘技术是人们长期对数据库技术进行研究和开发的结果。一般来说，数据挖掘主要侧重解决四类数据分析问题：分类、聚类、关联和预测。数据挖掘现在还不是你学习的重点，如果你有兴趣，可在掌握一定的数据分析知识后，再进一步查阅相关资料学习。

1.2.5 数据展现

小白：您刚才讲通过分析，隐藏在数据内部的关系和规律就会逐渐浮现出来，那么通过什么方式展现出这些关系和规律，才能让别人一目了然呢？

Mr.林：不错，孺子可教也！这就是我接下来要说的数据展现了。众所周知，每个人看待事物都有自己的理解方式，所以数据分析师在展现结果的时候一定要保证绝大部分人对结果的理解是一致的。



数据展现

一般情况下，数据是通过表格和图形的方式来呈现的，我们常说用图表说话就是这个意思。常用的数据图表包括饼图、柱形图、条形图、折线图、散点图、雷达图等，当然可以对这些图表进一步整理加工，使之变为我们所需要的图形，例如金字塔图、矩阵图、漏斗图、帕雷托图等。

大多数情况下，人们更愿意接受图形这种数据展现方式，因为它能更加有效、直观地传递出分析师所要表达的观点。记住，一般情况下，能用图说明问题的，就不用表格，能用表格说明问题的，就不用文字。

小白：原来图表的运用还这么有讲究呀，我一定好好学习如何用图表说话！

1.2.6 报告撰写

Mr.林：数据分析报告其实是对整个数据分析过程的一个总结与呈现，通过报告，把数据分析的起因、过程、结果及建议完整地呈现出来，以供决策者参考。所以数据分析报告

>> 谁说菜鸟不会数据分析

是通过对数据全方位的科学分析来评估企业运营质量，为决策者提供科学、严谨的决策依据，以降低企业运营风险，提高企业核心竞争力。

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，能够让阅读者一目了然。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，有助于阅读者更形象、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，没有明确结论的分析称不上分析，同时也失去了报告的意义，因为我们最初就是为寻找或者求证一个结论才进行分析的，所以千万不要舍本求末。

最后，好的分析报告一定要有建议或解决方案，作为决策者，需要的不仅仅是找出问题，更重要的是建议或解决方案，以便他们在决策时作参考。所以，数据分析师不光需要掌握数据分析方法，而且还要了解和熟悉业务，这样才能根据发现的业务问题，提出具有可行性的建议或解决方案。下面的图1-6示范了一个分析报告的基本框架。

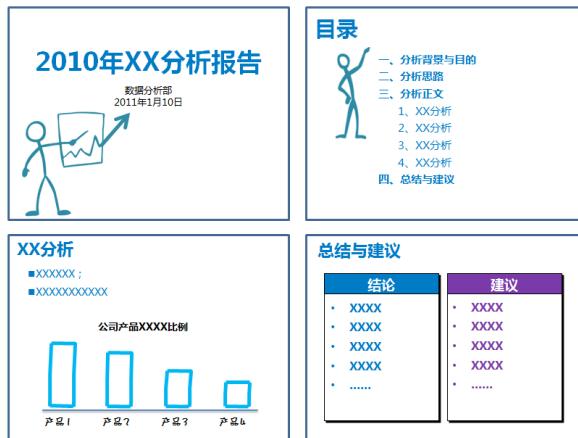


图1-6 数据分析报告示例

Mr.林：以上是数据分析的6个关键步骤，今天只是简要介绍，以后将逐一为你讲解。

听完Mr.林的长篇大论，小白说道：好的，我现在对数据分析已经有了一个基本的了解，之后还需要跟您进一步系统地学习数据分析相关知识。

1.3 几个常用指标或术语

Mr.林：一个优秀的数据分析师要有扎实的数据解读功底，因为在进行数据分析时，经常会遇到一些分析指标或术语，对这些指标或术语的理解不够充分就没法开展工作了。还

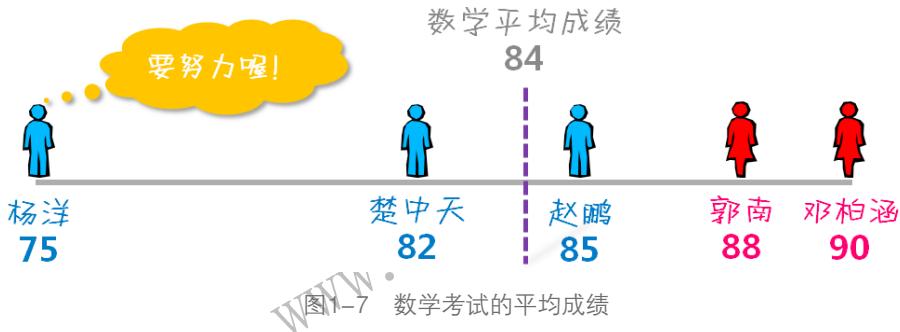
有点时间，我先给你简单介绍一些基础的分析指标或术语。

小白：太好了！拿只笔先~~

◎ 平均数

Mr.林：我们在日常生活中提到的平均数，一般是指算术平均数，就是一组数据的算术平均值，即全部数据累加后除以数据个数。算术平均数是非常重要的基础性指标。平均数是综合指标，它的特点是将总体内各单位的数量差异抽象化，代表总体的一般水平，掩盖了总体内各单位的差异。

例如现有某学期学生的数学考试成绩，通过计算成绩平均数，可得如图1-7所示的结果。将每位同学的数学成绩与平均数相比较，就能发现哪些同学的数学成绩高于平均数，需要保持；哪些同学的成绩低于平均数，需要继续努力。



当然，在平均数这个指标中，除算术平均数以外，还有其他平均数，如调和平均数和几何平均数等。我在后面还会教你平均数的相对应用。

小白：一个小小的平均数都有这么多学问呀。

◎ 绝对数与相对数

Mr.林：绝对数是反映客观现象总体在一定时间、地点条件下的总规模、总水平的综合性指标，也是数据分析中常用的指标，如GDP、总人口等。此外，绝对数也可以表现为在一定时间、地点条件下数量的增减变化，比如“*A*”国人口比“*B*”国人口多1000万人。

相对数是指由两个有联系的指标对比计算而得到的数值，它是用以反映客观现象之间数量联系程度的综合指标。计算相对数的基本公式是：

$$\text{相对数} = \frac{\text{比较数值(比数)}}{\text{基础数值(基数)}}$$

分母是用做对比标准的指标数值，简称基数；分子是用做与基数对比的指标数值，简称比数。相对数一般以倍数、成数、百分数等表示，它反映了客观现象之间数量联系的程度。

使用相对数时需要注意指标的可比性，同时要与总量指标（绝对数）结合使用。下面

>> 谁说菜鸟不会数据分析

我举几个例子，你一看就明白了，如图1-8所示。

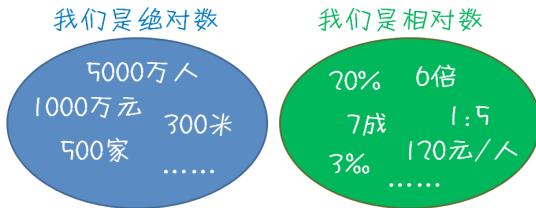


图1-8 绝对数与相对数

◎ 百分比与百分点

Mr.林：百分比是相对数中的一种，它表示一个数是另一个数的百分之几，也称百分率或百分数。百分比通常采用百分号（%）来表示，如8%，50%，168%等。由于百分比的分母都是100，也就是都以1%作为度量单位（如图1-9所示），因此便于比较，在数据分析中的应用非常广泛。

$$\frac{10}{100} = 10\%$$

图1-9 百分比的表示形式

百分点是指不同时期以百分数的形式表示的相对指标的变动幅度，1个百分点=1%。例如，图1-10中某公司发言人就混淆了百分比与百分点的概念。表示构成的变动幅度不宜用百分数，而应该用百分点。因此，这位公司发言人正确的说法应该是“公司今年的利润是45%，比去年的28%提高了17个百分点”。



图1-10 百分比与百分点

◎ 频数与频率

Mr.林：频数是指一组数据中个别数据重复出现的次数。如图1-11的左图所示，某校A

班学生共50名同学，按性别进行分组，分为男与女两个组别，男同学的频数为30，女同学的频数为20。

频率是每组类别次数与总次数的比值，它代表某类别在总体中出现的频繁程度，一般采用百分数表示，所有组的频率加总等于100%。还是用某校A班的例子，30个男同学在50个同学中出现的频率为60%，即 $(30 \div 50) \times 100\%$ ；而20个女同学在50个同学中出现的频率为40%，即 $(20 \div 50) \times 100\%$ ，如图1-11的右图所示。

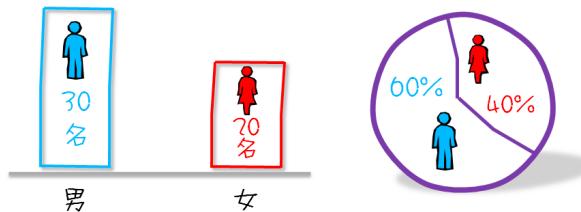


图1-11 频数与频率

小白：所以频数是绝对数，频率是相对数。

◎ 比例与比率

Mr.林：比例与比率都属于相对数。

比例是指在总体中各部分的数值占全部数值的比重，通常反映总体的构成和结构。比如，A班共有学生50人，男生30人，女生20人，则男生的比例是30/50，女生的比例是20/50。由此可以看出，比例的基数（也就是分母）都是全体学生人数，即为同一个基数。

比率是指不同类别数值的对比，它反映的不是部分与整体之间的关系，而是一个整体中各部分之间的关系。比如刚才的例子，男生30人，女生20人，则男生与女生的比率是30/20，如图1-12所示。这一指标经常会用在社会经济领域，比如我国的人口性别比就是用每100名女性数量相对的男性数量来表示的。

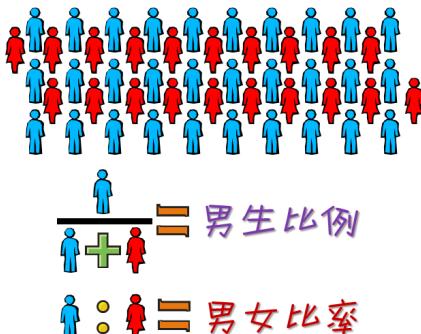


图1-12 比例与比率

>> 谁说菜鸟不会数据分析

◎ 倍数与番数

Mr.林：倍数与番数同样属于相对数，但使用时容易混淆。倍数是一个数除以另一个数所得的商。比如 $A \div B = C$ ，就是说A是B的C倍。需要注意的是，倍数一般是表示数量的增长或上升幅度，而不适用于表示数量的减少或下降。

番数是指原来数量的2的N次方倍。比如翻一番为原来数量的2倍（ 2^1 ），翻两番为4倍（ 2^2 ）。如图1-13所示，这位公司发言人在讲话中就混淆了倍数与番数的概念。因此，这位领导正确的说法应该是“公司产品销量翻一番（ $6.4=3.2 \times 2^1$ ），从去年的3.2万件提高到今年的6.4万件。此外，我们成本控制也很好，由20万元下降了一倍，今年成本为10万元”。

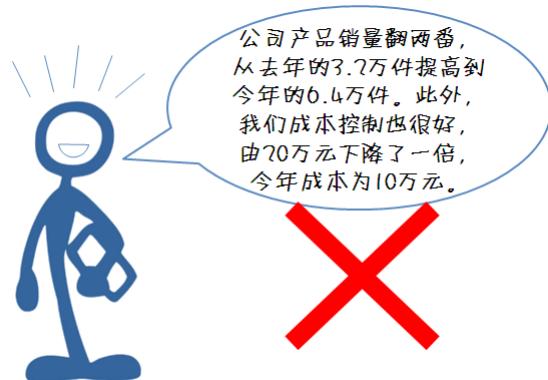


图1-13 倍数与番数

◎ 同比与环比

Mr.林：同比是指与历史同时期进行比较得到的数值，该指标主要反映的是事物发展的相对情况。例如2010年12月与2009年12月相比，如图1-14左图所示。

环比是指与前一个统计期进行比较得到的数值，该指标主要反映的是事物逐期发展的情况。例如2010年12月与2010年11月相比，如图1-14右图所示。



图1-14 同比与环比

1.4 本章小结

Mr.林：我讲的这些内容现在听起来虽然枯燥，但能够为你以后做数据分析打下坚实的基础，其中一些内容也只有通过实际的操作才能有更深刻的体会。

小白：我明白。

Mr.林：小白，你来总结一下今天的收获吧

小白：好的。

1、何谓数据分析，数据分析未来的前景，菜鸟与数据分析师的区别，以及数据分析师的基本素质。

2、数据分析的六部曲，首先是明确分析目的和内容，然后是数据采集，第三步是将采集回来的数据进行处理，第四步是根据研究目标和内容进行分析，第五步是将分析出的结果通过图表的方式展现出来，最后一步是撰写数据分析报告。

3、数据分析中经常用到的指标或术语，除掌握最基本的平均数、百分比等概念以外，还需要将容易混淆的指标进行明确区分，避免在以后阐述分析结果时出现错误。

Mr.林点点头：不错，不错，今天讲授的内容你都认真听了。在实际工作中仔细体会吧！