

本章目的

- 广泛地介绍统计学科及其应用
- 介绍每位经理人应该怎样掌握电子工作表
- 介绍你应该怎样掌握本教材所使用的软件
- 区分描述统计学和推断统计学
- 讨论数据的来源
- 讨论数据的类型
- 介绍抽样方法的类型
- 研究怎样进行调查价值的评估

统计应用

雅韵公司的顾客满意度调查

雅韵是一家通过电子商务网站销售高级立体声播放设备的公司,追求给顾客提供优质的服务和高质量的商品。为了加深对顾客的了解,雅韵要求其顾客填写一份满意度调查表,并马上交回公司。调查表中有以下的一些问题:

- 从你订购商品起,到收到这些商品的天数:
- 在未来的 12 个月内,你准备花多少钱(单位:美元)去购买立体声播放设备?
- 根据你在雅韵最近的购买情况,你对雅韵提供的服务的总体评价如何?

☐ 比预想的要好得多	☐ 比预想的要好
☐ 和预想的差不多	☐ 比预想的要差
☐ 比预想的要差得多	
- 你如何评价最近在雅韵购买的立体声播放设备的质量?

☐ 比预想的要好得多	☐ 比预想的要好
☐ 和预想的差不多	☐ 比预想的要差
☐ 比预想的要差得多	
- 在未来的 12 个月内,你是否打算在雅韵购买其他商品?
是_____ 否_____

请你在此次调查进行审核。此次调查的结果能提供何种类型的信息?雅韵将如何使用这种信息去提高现有的服务和商品的质量?你认为这个调查中还应包括哪些问题?

在本教材中,每章开始部分的“统计应用”是实际情况的介绍,能使读者作为一名经理人如临实际的工商环境之中,根据具体情况的来龙去脉,运用统计方法帮助作出决策。在“统计应用”的情况介绍中所提出的问题,将在本教材接下去的章节中,运用统计技术予以解答。通过阅读本教材并完成练习后,你将会了解到 21 世纪企业经理人应该怎样运用统计方法作出决策,解决问题,提高其产品和服务的质量。

1.2 现代统计的产生与发展

从历史上看,现代统计学的产生和发展可找到三个不同的原因:政府需要搜集公民的数据(见参考文献 5,6,12,13 和 16)、概率论理论的发展以及电子计算机不断地升级换代。

在有记载的历史中,始终贯穿了数据的收集。在埃及、希腊和罗马的文化中,最初是为了税收和军事征募的目的收集数据。在中世纪,教会组织经常保留着有关出生、死亡和婚姻情况的记录。在美国,殖民时期就保留着各种记录(见参考文献 17)。从 1790 年起,美国宪法就规定每十年进行一次人口普查。对全国范围内的数据需求与**描述统计学**(descriptive statistics)的发展密不可分。描述统计学是关于收集、展示一批数据,并反映其特征的各种方法,其目的是为了正确地反映这些数据的特点。

与这种发展相应的是概率论数理统计公式的推出。概率论基础部分可以追溯至 17 世纪中叶数学家帕斯卡与赌徒切维拉·德·梅瑞之间的通信(见参考文献 5,6)。这些通信,对文艺复兴时期的博弈机会进行了探讨。经过贝努里、德莫夫和高斯等数学家的不断改进,产生了**推断统计学**(inferential statistics)。为了理解推断统计学的含义,需要有如下的定义:

总体 (或称为域)是所研究的全部单位或事件。

参数 是综合测量的整个总体的某个特征。

样本 是从总体中抽取的部分单位,并用于分析总体。

统计量 是根据样本数据计算的综合测量值,可用以反映或估计整个总体的某个特征。

推断统计学使人得以根据样本的结果对总体的某个特征作出估计或对总体作出决策。在 20 世纪,普阿松、费雪、戈赛特、纽曼、瓦尔德和图基等统计学家开拓并发展了推断统计学的很多方法,这些方法至今还被广泛地应用着。现在需要这些方法,是由于抽样的需求。因为当总体容量很大时,从整个总体中获取信息通常要耗费大量的费用和时间,并且比较麻烦。对总体指标的判断,需要根据从这个总体中获取的样本信息。概率论提供了这种从样本的结果反映总体特征的可能性。

电子计算机的不断升级换代,极大地促进了统计学的发展。现在,个人使用的计算机运算能力,已经达到了早期统计学家只能想像的程度。从 20 世纪 60 和 70 年代开始,称为**统计软件包**(statistical packages)的统计软件的开发,可以帮助使用者方便地进行统计分析方面的计算。在 20 世纪 80 年代,市场上销售的统计软件,如 SAS,SPSS

和 Minitab 等(见参考文献 9,11 和 15),可根据个人计算机市场的发展不断推出新的版本。当这些版本面世后,称为**电子工作表程序**(electronic worksheet programs)的另外一种类型的软件,如 Microsoft Excel,逐步完善并具有了统计运算能力。现在,软件工具的普遍使用已经使商务决策时统计方法的应用不断增加。当代的经理人面临着来自两方面的压力:必须更多地依据越来越多的信息作出商务决策;以及当全球经济变得越来越电子化和一体化时,只能增加商务中的统计应用。

1.3 每位经理人需要了解电子工作表应用的哪些内容

电子工作表程序能使使用者在已排列好行和列的工作表中,输入各种数据、公式和文字。该形式的表格,正是在管理和财务分析中长期使用的。工作表中既可输入简单的数值或公式,又可输入处理输入数据的指令,而一系列公式的有效组合还可建立综合处理数据的数学模型。

这些程序的实践的性质和改进的用户界面有导致被误用的倾向。许多使用者错误地把重点放在怎样进入程序或怎样填写工作表等操作技术上,而忽略了形成工作表所需的分析。这就造成了这类使用者可能会编制出与原来的信息前后矛盾的工作表。而且这种表是很难理解、很难验证的,对类似的问题也难以再次使用。

一个常用的分析方法是把一个问题分解成几个部分。就构成一份工作表的任务而言,该方法首先要求用户单独考虑求取的结论、输入所需的数据和用于导出结果的公式。在某些情况下,有的公式被看作是为其他可直接导出结论的公式提供数值的中间计算过程。一旦有了明确的定义,这些部分可在一份工作表的不同部位分别予以完成。如果情况许可,还可在相同工作手册或文档的不同的工作表中分别予以完成。这种方法,因涉及各个领域,促进了对设计方案的审核。而这正是非专业使用者在寻找结果前常要做的工作。使用电子工作表程序还提出了一个更为普遍的、任何类型的统计软件都会产生的问题:需要使用者理解程序计算过程中的假设和限制条件。只有这样,才能使使用者作出合适的选择,以避免被错误的结论引入歧途。

1.4 你需要了解本教材所使用软件的哪些内容

为了有效地使用本教材,读者应该使用 Microsoft Excel 来解决统计问题。市场上的电子工作表有单独销售的,也有包含在微软的办公软件一起出售的。尽管存在一些瑕疵,Microsoft Excel 还是为初学商务的学生提供了展现和演示统计方法的最佳途径,并且还具有很多其他的优点,使之广为流行。

为了适合各位读者运用 Microsoft Excel 的不同经验和能力,本教材能帮助你在如下几个方面应用 Microsoft Excel:

- 可以用 PHStat CD 客户版程序产生工作表,而这个程序可以在安装了视窗系

统的 Microsoft Excel 97 或以上版本中运行。这只需要对 Excel 有所了解。

- 按照每一章的最后部分“Excel 手册”中的指令去做,你就可以自己动手使用工作表了。这需要对 Excel 非常了解。
- 可以检索、检验和研究所附 CD,CD 对章内问题提供了解决方案。这只需要对 Excel 略微有点了解便可。

因为“Excel 手册”部分的设计和技术处理与 PHStat 和所附 CD 是一致的,使用 Microsoft Excel 的不同方法所产生的工作表是完全相同的或是非常相似的。在本教材中,这些方法自始至终都是可以交替使用的。

以下将进一步介绍这 3 种方法。

使用 CD 客户程序(PHStat)

一般而言,使用 CD 客户程序(PHStat)是产生含有统计题解的新工作表的最为便捷的方法。可以把 PHStat 看成能自动地产生工作表的“傻瓜机”,这种表与所附 CD 手册以及在“Excel 手册”中讨论的工作表是相似的。为了进行分析,PHStat 可以对数据进行预处理,而若用手工来做这种工作,会非常地不方便,而且容易出错。PHStat 也综合了 Excel 的基本特征,同时能用新的方法进行深入的分析。使用 PHStat 对 Microsoft Excel 的要求很低。

附录 F 和附录 G 中讨论了安装和使用 PHStat 程序。这是 PHStat 的最新版本,在所附的 PHStat 阅读文件中,给出了怎样安装你自己的 PHStat。当正确安装完毕后,PHStat 会插入 Excel 的菜单中,如图 1.2 所示。

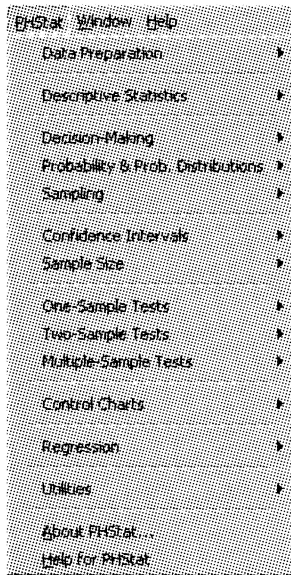


图 1.2 PHStat 菜单

对那些通过操作来学习的人,本教材在很多章的后面包括了“Excel 手册”部分,它可以用于检验同一章前面部分所使用的工作表。这种方法需要对 Microsoft Excel 非常熟悉,因为你需要掌握怎样绘制散点,公式输入错误时怎样改正以及其他一些输入技巧。

使用所附 CD 的工作簿

使用所附 CD 的工作簿,包括每一章之内的问题和章后部分习题的完整分析,这只需对 Excel 有所了解便可,你只要知道用 Microsoft Excel 正确使用工作簿的操作技

巧。所需的基本技巧,将在本章后面的“Excel 基础”中予以讨论。

所附 CD 中的所有工作簿的结构是相同的。为了加深理解工作簿是如何产生其结果的,所有工作簿都以一张汇总的工作表开始,此表列出了工作簿中所包含的内容提要,需要时数据表包含了列的数据以供分析之用。中间的计算和结果可在其他一张或几张工作表中找到,在这种工作表中,含有实际结果的单元格用淡青绿色表示,而含有数据的每个单元格则用淡黄色表示,这些数据的变动会使其结果也随之发生变化。

提示

在本教材中,Microsoft Excel 的运用并不能保证它能在实际的统计活动中的应用(见 1.2 节)。还要注意的,通常 Microsoft Excel 所做的统计运算含有可能导致错误结论的缺陷(见参考文献 7)。

在使用大量的数据或具有异常统计特性的数据时,特别容易发生这种情况。一些缺陷总会导致错误的结论,此时,这种缺陷会被注明,并会提供其他方法以产生结果。对所有其他的情况,选择所有要分析的数据以避免缺陷的影响,以使你从 Microsoft Excel 所输出的结果中能够得出正确的判断。

1.5 为什么需要数据

开展工商活动的基础是获得有用的信息。你可以把数据想像成所需的信息,能帮助你在某些情况下作出更为正确的决策。许多场合都需要数据,包括:

- 一名市场研究者需要对产品的特性进行评估,以区分不同的产品。
- 一家药品制造厂商需要判别一种新药是否比现在正使用着的药更有效。
- 一名生产经理按惯例要检查生产过程,以检验其生产的产品质量是否符合公司的标准。
- 一名审计人员想通过查看某家公司的财务报表,以确认这家公司是否是依据了通行的会计准则做报表。
- 一名财务金融分析人员想判断在未来的五年中,哪些行业中的哪些公司最具有成长性。
- 一名学生想获得同学们最喜欢的摇滚乐队的数据,以满足自己的好奇心。

收集数据有六个主要的理由,现把其列于展示栏 1.1 中。

展示栏 1.1

收集数据的主要理由

1. 为调查提供所需的数据。
2. 为研究提供所需的数据。
3. 需要用数据来反映所提供的产品或服务的质量的好坏。
4. 需要用数据来评价是否符合标准。
5. 在决策过程中,需要用数据来协助。
6. 数据能满足我们的好奇心。

在本章开始时所举的雅韵公司顾客满意度调查中,就说明了以上 1、3、4 和 5 点。例如,雅韵公司的调查结果就是搜集到了数据,然后才对数据进行分析,以根据需要、评价标准、测量服务和产品的质量好坏制定出可供选择的行动方案。

在开始进行统计分析时,最重要的是识别出数据的来源是否合适。若数据存在偏差、模糊不清或其他错误等缺陷,即使是最富有想像力和经验最为丰富的统计学家都不可能弥补这些缺陷。

1.6 数据的来源

以下是数据收集的主要来源,现列于展示栏 1.2 中。

展示栏 1.2

数据收集的主要来源

1. 从政府、行业 and 个人的出版物中获得数据。
2. 设计一次试验以获取必要的数据。
3. 进行一次调查。
4. 从观察研究中获取。

数据的收集被称为原始来源,数据的汇编被称为次级来源。就如在展示栏 1.2 中所示,获取数据第一种方法就是把政府、行业 and 个人的出版物作为数据来源。对公共机构或私人组织而言,美国政府就是数据的主要收集和汇编者。劳工统计局就有收集就业数据的职能,并且要公布每月的消费者价格指数。人口普查局除了宪法规定每十年进行一次人口普查外,还要进行有关人口、住房和制造业的一系列调查。此外,还要进行一些旨在了解犯罪、旅行和健康保健等的专项调查。

除了美国政府以外,各种行业的出版物提供了各行业的数据,如莫迪公司这样的投资服务商会提供某一公司财务数据。如 A. C. 尼尔森这种为大企业集团服务的公司,会为客户提供有关该客户的产品与其竞争对手的相关产品的对比信息。每天的报纸充斥了关于股价、气候和各种运动的统计数字信息。

第二种收集数据的方法就是试验。在一次试验中,对整个过程都要进行严格的控制。例如,在检验洗衣机洗净程度的研究中,研究人员不是去询问顾客他们认为哪种牌子的洗衣机洗衣效果最好。而是通过实际洗涤脏衣服,来研究哪种牌子的洗衣机洗涤效果最佳。正确的试验设计,通常在更为高级的课程中予以讨论,因为这要涉及复杂的统计运算。然而,为了对检验和试验有所了解,本教材的第 7 到第 10 章中给出了试验设计的基本思想。

第三种收集数据的方法就是进行调查。它对所调查人们的行为不进行任何控制,仅提出诸如信仰、态度、行为和其他特征方面的问题,然后对他们回答的结果进行整理、编码、列表和分析。

第四种收集数据的方法是通过观察研究。研究人员通常是在自然状态下,进行直接的观察。大部分关于动物行为的知识都来自于这种方法。其他一些领域的科学知

识,如天文学和地理学,因为很难进行实验和调查,只能通过观察进行研究。商务中的观察研究可搜集到大量有关的信息,用以帮助作出正确的决策。例如,专家小组作为流行的市场手段,常用于引出对开放式问题的自由讨论。会议的组织者引导讨论,所有的参加者都要回答所提的问题。其他安排更严密的讨论方法也可不断地引入到小组讨论中,以获取信息并使意见逐步趋于一致。这些方法包括组织行为心理工具,如头脑风暴法、德尔菲技术和名义小组法。近年来,因为这些工具已影响到商务的整体品质管理理念,因而变得越来越常用。全面质量管理(TQM)强调的是小组协作和授权雇员对每个产品和服务加以改进。

提示

数据来源和知识的更新

从来没有像现在这样,可获得大量及时的和精确的数据,这要感谢信息技术的广泛使用。当产品在超市、百货商店和其他渠道被销售出去时,条形码自动地记录下来存货的数量。自动取款机(ATM)和其他网上银行使得交易能被及时记录下来。旅行机构有精确到最近一分钟的关于航班和旅馆的空位数据。十年前要花数小时、甚至数天的交易,现在只要在瞬间便可完成。

图书馆的使用也有了新的含义。人们不再限于诸如书籍、定期刊物、杂志、手册和报纸等印刷出来的媒体。基于计算机的信息系统,如使用 CD 盘上的数据库、在万维网上冲浪,或与其他网络用户用电子邮件交换信息等,使你能通过电子技术方便地搜寻和找到数据。本章中运用微软办公软件这部分内容,介绍了从万维网中获取数据的方法。

为了能设计一项试验、开展一次调查或进行一次观察研究,你必须搞清楚数据的不同类型。将用一次调查来反映获取数据中涉及的问题,因为在其他研究中也存在大部分相同的问题。

1.7 数据的类型

统计学家用不断改进的调查来研究各种随机变量。对这些随机变量的观察所获取的数据,单位与单位(或人与人)之间实际上总是不同的,因为世界上没有两片完全相同的树叶。

图 1.3 列举了根据两种不同类型的随机变量所获取的观察结果或数据:分类型或数值型。

分类型随机变量(categorical random variables)会导出分类型的回答,如回答是或否。例如,对问题:“你最近持有股票或债券吗?”的回答,就限于简单的是或否。另一个例子是回答雅韵公司调查中的问题:“在未来的 12 个月内,你是否打算在雅韵购买其他商品?”

数值型随机变量(numerical random variables)会导出如你的身高是多少英寸的回答。其他的例子如对在雅韵公司的顾客满意度调查中所问的问题:“在未来的 12 个月

内,你准备花多少钱(单位:美元)去购买立体声播放设备?”的回答,或者对“你现在订阅了几份杂志?”的回答。数值型变量有两种类型:离散型和连续型。

数据类型	问题类型	回答
分类型	你最近持有股票或债券吗?	是 ☐ 否 ☐
数值型	离散型 → 你现在订阅了几份杂志?	_____ 本
	连接型 → 你的身高是多少?	_____ 英寸

图 1.3 数据的类型

离散型随机变量(discrete random variables)产生于对计数问题的回答,例如:“你订阅了几份杂志?”就是一个离散型的随机变量,因为其回答一定是有限个的整数,如你订阅了 0 份、1 份、2 份杂志等。

连续型随机变量(continuous random variables)产生于对量度问题的回答,例如人们的身高就是一个连续型随机变量,因为任何的回答都要根据测量工具的测量结果,要么是连续的数值,要么是落在某一区间的数值。如由于测量工具的精确程度不同,你的身高可能是 67 英寸、67¼英寸、67½英寸、或者是67⁵⁸/₂₅₀英寸。

从理论上讲,不可能出现两个人的身高是完全相同的这种情况。因为测量的工具越精确,就越有可能区分出他们身高的不同。但是,大部分测量工具都不是十分精确的,无法测出细小的差别。因而,即使随机变量确实是连续的,还经常会在试验或调查的数据中发现取值相同的观测值。

本节习题

基本练习

- 1.1 若在某快餐店内有三种不同的饮料出售:果汁、茶和咖啡,试解释不同的饮料为什么是分类型变量。
- 1.2 若某快餐店出售的果汁有三种规格:小杯、中杯和大杯,试解释不同规格的果汁为什么是分类型变量。
- 1.3 若我们要测量从互联网上下载一份 MP3 文件所需的时间,试解释为什么下载时间是数值型变量。

概念应用

- 1.4 试判断以下是分类型的还是数值型的随机变量。若是数值型的随机变量,指出是离散型的还是连续型的。
 - a. 每户家庭拥有的电话台数。
 - b. 最早使用的电话类型。
 - c. 每月拨打的长话次数。
 - d. 每月拨打的长话中,最长一次的通话时间(单位:分钟)。
 - e. 最早使用的电话机颜色。

- f. 每月长话的费用(单位: 美元、美分)。
- g. 你是否拥有一部手机。
- h. 每月拨打本地电话的次数。
- i. 每月拨打的本地电话中,最长一次的通话时间(单位: 分钟)。
- j. 在住所内,是否有一条电话线与计算机的拨号器相连。
- k. 在住所内,是否有一台传真机。

1.5 假设以下信息是在开学第一周内,从曾到过学校书店的学生中收集的:

- a. 购书的费用。
- b. 购买了几本教材。
- c. 在书店中用了多少时间。
- d. 所学的专业。
- e. 性别。
- f. 是否有个人电脑。
- g. 是否有立体声播放设备。
- h. 在本学期注册了多少个学分。
- i. 是否在书店中购买过衣物。
- j. 在书店购买后的支付方式。

试把以上的每个变量划分为分类型的变量和数值型变量。若是数值型的随机变量,指出是离散型还是连续型的。

1.6 对以下每个随机变量,判断其是分类型的变量还是数值型的变量。若是数值型的随机变量,指出是离散型还是连续型的。

- a. 互联网络服务提供商的个数。
- b. 每月为互联网服务所支出的费用。
- c. 每个月的上网时间。
- d. 上网的最初目的是什么。
- e. 每周收到的电子邮件的数量。
- f. 一个月内,在网上购物的次数。
- g. 一个月内,在网上购物的费用总额。
- h. 个人计算机上的硬盘是否具有反复刻录的功能。

1.7 对以下每个随机变量,判断其是分类型的变量还是数值型的变量。若是数值型的随机变量,指出是离散型还是连续型的。

- a. 在过去的一个月中,购买服装的费用总额。
- b. 拥有冬季大衣的件数。
- c. 你最喜欢的百货商场。
- d. 在过去的一个月中,购买服装所用的时间。
- e. 最喜欢在哪个时段去购买服装(工作日的白天、工作日的晚上或周末。)
- f. 拥有几双冬季的手套。
- g. 外出购物时,最初所使用的交通工具。

1.8 假设以下信息是罗伯特·凯勒在麦德隆县存贷款协会处申请家庭抵押贷款时留下的:

- a. 居住地: 纽约的斯托尼·布洛克。
- b. 住房类型: 独院住宅。

- c. 出生年月：1962 年 4 月 9 日。
- d. 每月的支出：1 427 美元。
- e. 职业：报纸记者和编辑。
- f. 雇主：日报社。
- g. 工作年限：14 年。
- h. 在过去的十年中，有过几份工作：1 份。
- i. 家庭年收入：66 000 美元。
- j. 其他收入：1 600 美元。
- k. 婚姻状况：已婚。
- l. 子女人数：2 名。
- m. 抵押贷款的金额：120 000 美元。
- n. 抵押贷款的年限：30 年。
- o. 其他贷款：汽车。
- p. 其他贷款的金额：8 000 美元。

试划分以上回答的数据类型。

- 1.9 在调查中，最常包含的变量是收入。有时，这个问题可以这么问“你的收入是多少？（单位：千美元）”，或者要求被访问者在给出的一系列数据区间中进行选择：“请在与你的收入水平相应的位置打×”。
 - a. 在第一种方式下，试解释为什么收入即可视为离散型的变量，又可视为连续型的变量。
 - b. 若你要进行一次调查，在以上两种提问的方式中，你比较喜欢哪一种？试说明原因。
 - c. 在以上两种提问的方式中，你认为哪一种的回答率会更高？为什么？
- 1.10 若两名学生在同一考试中都取得了 90 分，通常会对考试成绩这个随机变量是否是连续的产生什么争论？
- 1.11 假设一家大型连锁百货商店的市场部经理想在一个大都市地区进行一次调查，以了解职业妇女每月在购买服装时所耗费的时间。
 - a. 试对要调查的总体和样本进行描述，并指出该经理想收集到什么样的数据？
 - b. 请列出(a)中所需的问题大纲，包括你认为与这次调查相关的三个分类型问题和三个数值型问题。

1.8 抽样方法的类型

根据 1.2 节中的定义，样本是为了分析总体、从总体中选取的一个部分。与对总体进行全面普查不同，统计抽样主要是选取一个对总体具有代表性的样本，抽取的样本可以提供用于估计整个总体特征的信息。

抽样过程中首先要定义一个抽样框。抽样框中含有总体中全部或部分单位，它可以是总体各单位的编号、名录或一份地图这样的数据源。样本是从抽样框中抽取的，