

第一章 导言

对于“统计”一词,人们并不陌生。在日常工作及生活中,人们经常看到、听到和用到这个词。例如:开会时主持人要统计出席会议的人数;球赛中教练员要统计各位队员的命中率和犯规次数;报刊上总是定期或不定期地公布诸如物价指数、人口增长率、国民生产总值等统计数据;每天的新闻广播里也不乏工农业生产速度、出口贸易额、居民生活费用指数等统计数字的报道……这里,人们所认识的“统计”主要是统计资料和统计数字,而统计学的含义远比这丰富得多,这是我们在学习统计学之前必须首先弄清楚的问题。

本章主要讨论以下问题:

1. 统计学的概念和特点; 2. 统计学的产生和发展; 3. 统计学的类别; 4. 统计学在经济学和经营管理中的应用; 5. Excel 与统计学; 6. 统计学中的重要概念及统计的研究程序。

第一节 统计学的概念

关于什么是统计学有许多种解释,这里略举几种:

- 统计学是关于数据资料的搜集、整理、显示、分析和推论的科学,目的在于帮助人们做出更有效的决定。

- 统计学是一门关于大量数据资料的搜集、整理、描述和分析的学科,目的在于探索数据内在的数量规律性。

- 统计学是在面对不确定的情况下,提供给人们做出聪明决策的科学方法,其过程包括资料的搜集、整理、陈列、解释和分析,并可进而加以推论,从而可获得合理的判断与有效的结论。

……

如果继续列举的话,还有许多,但基本上大同小异。综合各种解释,本书将统计学定义如下:

统计学是一门对群体现象数量特征进行计量描述和分析推论的科学。

这个定义包括统计学的三个核心要点:

第一,统计学的研究对象是群体现象。

统计学有两个基本概念:总体和样本。总体指调研者研究对象的集合;样本指来自总体的部分对象的集合。统计学的研究对象就是这些集合,而不是组成这些集合的某一个元素或个体。如果要知道一件物体的重量,只要把它放到秤上称一称;如果要知道一个人的身高,通过测量就可得知。而若要知道一批物件的重量、一群人的身高,就需要汇总和平均,采用一定的统计方法。若期望通过小小的样本信息了解总体,就需要抽样调查和估计检验,进行统计分析推论。虽然总体或样本的信息都表现在一个个元素或个

体上,研究总体不能脱离个体,但统计学研究的不是个体现象,而是通过个体所载有的信息来研究和说明群体现象。

第二,统计学所探索的是群体现象数量表现的内在规律性。

这里有两个要点:数量性和规律性。

其一,数量性。统计学的英文是“statistics”。这个英文名词有两个含义:作为复数,指的是通过观察或调查所搜集到的一定时间、空间下反映客观现象的特征的数字资料;作为单数,指的则是作为一门科学的统计学。统计的这一名词特性,很直观地说明了统计学与统计数字之间紧密相连的特点。数字是统计的语言,统计学是用数字来说明群体现象的特征。作为特征,可以是数量特征,如人的身高、年龄等;也可以是属性特征,如人的性别、民族等。统计学研究群体现象的特征,总是用数字来计量、说明的。例如,统计学所探索的某人口群体的性别特征表现为不同性别的人口数量和比例;民族特征为各民族人口数量、所占比例等。

其二,规律性。统计学研究的是个体之间存在数量差异的群体现象。如果一批物件重量都一样,一群人身高都相同,是用不着统计的。然而现实中群体现象总是由许多数量特征各异的个体组成的,而这些千差万别的个体数量特征下却掩盖着群体现象的某一数量规律性。例如,就每一个家庭的新生儿的性别而言,在出生前是很难确定的,可能是男性,也可能是女性。但如果对大量的新生儿的性别进行观察,就会发现男女比例约为107:100,这个比例就是新生儿这个群体性别特征的数量规律。又如,掷硬币和骰子,谁都知道随机地掷一次是无法确定结果的,即每一次抛掷结果各异,但如果我们反复不断地掷,当抛掷次数足够多时就会发现规律,即硬币出现正面或反面、骰子出现任一点数的机会都是均等的,这就是掷硬币和骰子的数量规律。再如,关于证券投资,谁都知道收益和风险宛如一对孪生兄弟,可是谁又都希望能够较低的风险下取得较高的收益,那么风险和收益之间究竟具有什么样的联系?通过大量的观察,根据大量的数据资料,就可以找出它们之间数量联系的规律性。这个可以通过建立一定的统计模型来获得,目前应用广泛的资本资产定价模型反映的正是这个规律。类似的例子还有很多。

第三,统计学研究的是对群体现象数量特征进行计量描述和分析推论的方法。

统计计量描述指收集、整理、计算并显示说明客观群体现象特征的统计数据资料;统计分析推论指通过随机抽样的样本特征对总体的特征进行推论。统计学研究的是科学地进行计量描述和分析推论的方法,即如何才能够准确、真实地收集到所需要的数据,如何科学地对这些数据进行分类和归总,进一步计算,如何在允许的误差范围内用样本数据推论总体数量特征,等等。

例如,指数是十分重要也是人们十分熟悉的统计数字。举个最简单的例子,假如去年北京市鲜牛奶的价格是每500克4元,鸡蛋的价格是每500克3元,今年牛奶下降为2元,而鸡蛋上升为6元,计算这两件消费品的物价指数。

方法一:以去年为基期。牛奶的价格降低了一半,为去年的50%,鸡蛋的价格上涨了一倍,为去年的200%,用50%加上200%再除以2,得出平均数125%,所以今年的平均物价比去年上涨了25%。

方法二:以今年为基期。去年的牛奶价格是今年的200%,鸡蛋的价格是今年的50%,用200%加上50%再除以2,得出平均数为125%,即去年的平均物价比今年高25%。

方法三:以去年为基期,将鸡蛋的价格指数乘以牛奶价格指数($50\% \times 200\%$)后,开平方根得出平均数100%,则物价既没有上升,也没有下降。

三种方法导致三种截然不同的结论,究竟哪一种方法才最真实地反映了物价的实际变动?这正是统计学研究的问题。

再如,有一家出版商根据过去的经验知道,出于财务上的考虑,要成功地发行某一种月刊,就要保证在它所针对的读者群中,至少有20%的读者会订阅此刊。现在一种月刊已设计好,出版商寄出了1000份给1000位可能的读者,其中有190位读者表示要订阅此刊。由于订阅的读者只占19%,出版商是否应该决定不出版此月刊了呢?或者这这是由于样本的原因,而对于全部可能的读者而言,订阅的比例不会小于20%?统计方法将帮助人们解决这个问题,而这也是统计学研究的众多方法之一。

统计学研究各种各样的统计方法,这些方法是分析认识客观现象的有力工具,帮助人们在各种不确定的或复杂的情况下做出明智的判断和决策。

第二节 统计学的产生和发展

一、统计学的产生

统计学形成于19世纪中叶,它的产生主要源于概率论、国势学和政治算术。

17世纪中叶起,在德国兴起的国势学派用记述的方法来研究一国的地理、人口、财政、军事、政治、法律制度等方面。18世纪中期,哥廷根大学教授戈特弗里德·阿亨瓦尔(Gottfried Achenwall)首次提出了“统计学”(statistik)这一名词,并定义为国家显著事项的学问,言下之意是通过这门科学,可了解国家理乱兴亡之迹。最初的“统计”只是文字记载,后来才发展出了数字和图表。

英国是资本主义发展较早的一个国家,也较早开始利用数字对人口、经济等方面进行记载和推断,其代表人物是约翰·格兰特(John Grant)。他于1661年出版了《对死亡表的自然观察和政治观察》一书,发表了对人口出生率研究的结果,并观察到一切疾病和事故在全部死亡原因中占有稳定的百分比。与格兰特同时代的英国经济学家威廉·配第(William Petty)的《政治算术》和对国民收入估算的方法,不仅对经济学,而且对统计学的发展也具有重大的意义。他们虽未创立“统计学”之名,但所用于探索社会和经济现象的数量规律性的方法,则具有“统计学”之实。

概率论的产生最早起源于对赌博中掷骰子的输赢问题的研究。在但丁的《神曲》中就记载了投掷三个骰子所可能出现的各种结果。伽利略也从事过这方面的研究。但直到1714年,瑞士数学家雅各布·伯努利(Jacob Bernoulli)的《推测法》问世,概率论才作为一种理论初具雏形。之后,概率的乘法运算法则、正态分布、先验概率、逆概率等理论相继问世,概率论逐渐发展起来。

将这三方面成果融为一体而创立了统计学的,是比利时统计学家阿道夫·凯特勒(Adolph Quetelet)。在统计研究方面,他先研究了天文、气象方面的统计资料,后又利用统计数字研究植物界和人类社会。他以概率论作为理论基础,用大量观察和综合平均的方法来进行研究,从而把概率论、国势学、政治算术观察群体现象进行数量分析的方法,融合为一门统计学。

二、统计学的发展

统计学的发展可大致划分为三个阶段:

第一阶段是统计学的初创阶段。从 17 世纪中叶开始到 19 世纪末,经过二百多年的漫长时期,描述统计和概率论的基本内容逐渐形成,研究的基本统计问题在于求得一套通用的数学公式来描述待研究现象的整体。

第二阶段是统计推断方法体系基本确定的阶段。从 20 世纪初“小样本 t 分布”的问世,到 40 年代末“统计决策函数”与“序贯抽样”的提出,在这个阶段,估计理论、样本分布理论、方差分析、非参统计、时间序列、随机过程等方面的研究都获得了重大进展,推断统计迅速发展,概率论体系也日臻完善。

第三阶段是统计方法与应用研究全面发展的阶段。从 20 世纪 50 年代起,统计学受计算机、信息论等现代科学技术的影响,新的研究领域层出不穷,如多元统计分析、探索性数据分析、现代时间序列方法、人工智能,等等。据美国学者统计,现代统计学是以指数式加速度发展,新的研究分支不断增加,统计应用的领域不断扩展。统计方法在各学科领域的应用则又进一步促进了统计方法研究的深入和发展。

这些阶段的划分只是大致的,其发展是渐进的、错综的,并没有明确的时间界限。

从现在统计理论研究的发展趋势看,统计理论研究的分支仍会不断增加,统计学将越来越具有交叉学科的性质,统计学应用的范围将更加广泛。一些过去与数字毫无联系的学科,如政治学、历史学、法学等,都已经在并将更多地应用统计方法进行研究和分析。目前阻碍统计方法推广应用的主要是定性资料的统计分析。这类问题包括国民经济的行业、部门等分类资料,不同的政治观点、思想准则,等等,如何对这些定性问题进行量化处理分析,是统计学家们正在研究的重大课题之一。探索性数据分析的应用前景也很广阔。这种方法重视数据直观的显示、原始数据中信息的提取及对特殊数据的识别和具体分析。贝叶斯统计理论研究在统计推断、预测、决策等领域将继续显示其强健的生命力,成为统计实践中十分有力的工具。

总而言之,统计学的发展在理论方面对现代数学理论的依赖越来越深,在数据处理和计算手段上对计算机硬件及软件技术的发展依赖越来越多,在应用方面其范围则不断增加,研究的对象从无穷大到无穷小,实非言过其实。

第三节 统计学的分类

与任何其他科学一样,统计学随着对它研究的逐步深入而发展进步,必然会出现各个相互联系而又区别的分支。或出于不同的观察角度,或出于不同的研究重点,故常常有不同的分类,分出的各类别也难免有些重叠。一般而言,统计学大致有两种基本的分类:描述统计学和推论统计学;理论统计学和应用统计学。

一、描述统计学和推论统计学

从统计学的概念,我们已经知道统计学研究的一个方面是搜集、整理、描述数字资料的方法,这部分内容就形成了描述统计学。例如,要计量居民生活费用的变动,首先要通

过一定的行之有效的调查方法,搜集反映居民生活费用的数字,然后对这些资料进行汇总、归纳和计算,将原始资料整理成为有条理的能够说明被研究现象特征的科学指标,最后再以相应的统计图和统计表将这些结果表现出来。

统计学研究的另一个方面是利用样本数据推断总体特征,这部分内容则形成了推论统计学。例如,要研究我国 13 亿人口的年龄构成、出生率、死亡率等人口特征,仅靠人口普查不够,因为人口普查需要大批训练有素的调查人员和大量的经费,不宜经常进行。在两次普查的中间年份所需的人口数据资料,通常是采用抽样调查的方法取得样本资料,依据样本资料所给予的信息来推算人口总体的上述特征。由于样本包含的总体信息必然不完备,用其推算总体难免有误差,为了使这种误差尽可能小或将其控制在可以接受的范围内,统计学就需要研究抽取样本和推算的方法,以使推算的结果尽可能准确,且失误的概率尽可能小。

二、理论统计学和应用统计学

理论统计学也即数理统计学,主要阐述统计学的数学原理,其理论基础是概率论。它所包括的主要内容有:概率理论、抽样理论、实验设计、估计理论、假设检验理论、决策理论、非参数统计、序列分析、随机过程等,着重于阐明统计方法的数学原理。

将理论统计学的基本原理应用于各个学科领域,就形成了各种各样的应用统计学。如社会统计学、经济统计学、教育统计学、生物统计学等。应用统计学着重研究统计方法如何应用于某个具体的领域内,如何解决具体领域内的具体问题。它除了包括各领域通用的方法,如参数估计、假设检验、方差分析等,还包括某领域特有的方法,如经济统计学中的指数法、现代管理决策法等,以及社会统计学中的反事实方法、针对社会网络的统计模型等。应用统计学着重阐明这些方法的统计思想和具体应用,而不是统计方法数学原理的推导和证明。

第四节 统计学在经济学和经济管理中的应用

自古以来,统计就服务于经济学或经济管理,在经济学和现代经济管理研究中,统计学发挥的作用更不容忽视。

一、统计学与经济学

经济学家的抱负是要告诉人们怎样利用有限的资源最大限度地满足人类的需要,由此引起对资源配置、市场供求、经济增长等许多问题的研究。对这些问题,不能只作定性的理论分析,还需要经验性的定量方法。例如,经济学研究有效需求不足的问题,显然研究消费与需求的数量关系将有助于对问题的深入了解和认识。而这样做既要有经济统计数据,还要有适当的统计方法。一般而言,构造一个复杂的模型并不太难,但要想出一个良好的方法获取数据以适当地估计模型,则是比较困难的事。数据获取的困难在于:

- ① 某些变量无法直接测得,如消费者偏好、企业信心、技术进步等,只好用“替代变量”,如何替代,难有完美的方法。还有些变量的数值属于保密范围,如有关个人财富的数字,财富持有人往往不愿透露。如何测算准确,需要完善的方法。
- ② 数据缺落或不足,例如我国目前大部分宏观统计数据只提供到省市级,而县乡级的许多统计指标往往存在数据

严重缺失的问题。③ 数据不准确。④ 数据的统计口径不同,尤其在涉及不同年份的纵向比较或不同地区的横向比较时,许多时候会发现不同的数据样本采用了不同的调查或处理方式,从而难以直接进行对比分析。要克服数据获取的困难,需要统计方法的突破。假如数据的问题已经解决,还有如何估计模型的参数问题。如凯恩斯在 20 世纪 30 年代曾提出总量消费函数的概念。为了测定这一消费函数,人们花了大量的精力估计一个看起来十分简单的边际消费倾向。虽说根据凯恩斯的理论,边际消费倾向平均而言稳定在一个大致不变的水平上,但并不排除每次估计都会有波动,如何考虑这个问题?并且,既然是估计,就有误差,如何考虑这个误差?又如何使误差保持在可以接受的范围内?这都需要统计学提供科学的统计方法。

二、统计学与经济管理

统计学无论在政府宏观经济管理还是在工商企业微观管理中的应用都是极其普遍的。

在宏观经济管理中,政府总是关注着当前的经济形势和未来的经济趋势。为此,政府进行大量的统计调查和统计分析预测;目前工农业的生产和销售数量、未来 12 个月的情况;每月的消费者价格指数及由此可反映出的通货膨胀情况;房屋动工数量、货币回笼速度等,这些只是许许多多统计指标中的几个而已。政府依据统计数据决定货币的投放,中央银行的利率等宏观经济政策,并利用统计数据评价、检查施政的绩效。

在工商企业微观管理中,管理人员经常需要在未来条件不确定的情况下做出牵涉金数额巨大甚至是有关企业生死存亡的重大经营决策问题,如企业的扩张、新项目的上马、新产品的研制和投产等。而在日常的经营管理中,管理人员更时时面临各种各样的决策问题。利用统计决策模型,则可降低决策的不确定风险,做出明智的选择。市场管理人员运用统计方法可以进行市场调查和预测;生产管理人员运用统计方法可以进行产品的抽样检验和质量控制;仓管人员应用统计方法可以确定库存量;会计人员应用统计方法可以进行财务预测、损益分析、投资分析、内部稽核等。统计方法的应用使得企业管理富于效率,增强了企业的竞争能力。

统计方法在经济生活和经济管理中的作用还可以无穷无尽地列举下去,它的事半功倍之效令无数人人为之神往,要有效地从事经济研究和管理,统计学不可不学。

第五节 Excel 与统计学

随着统计学理论的发展,统计学能够刻画分析的数据规模越来越大,使用的分析工具也越来越复杂,渐渐超过了人们心算或笔算所能解决的范围。而计算机在数学工程学领域的广泛应用为统计学的发展提供了很多便利。现阶段,这种便利主要体现在两个方面。一方面,一些专业统计软件为统计学的专业学习和研究提供了更好的操作平台与实验环境,把统计学家从烦琐的计算中解放了出来;另一方面,一些非专业的统计软件为非统计学专业的各行各业人士应用统计学创造了更多的可能性。本书着重介绍的 Excel 软件就属于后者,该软件主要面向的客户群是商用客户而非科研人员,这也就决定了该软件不需要使用者具备太多的统计学专业知识和数理背景。相比较为专业的统计软件如 SAS 等,Excel 具有更富亲和力的人机界面,更简单直观的操作程序,对于缺乏统计学知识

的使用者,也相对容易上手。尽管 Excel 显得并不十分专业,甚至在很多人眼里它的功能无外乎绘制表格,但是,该软件的功能事实上对于一般的商务管理工作甚至初级的统计分析来说已经足够强大了。

Excel 能够解决的统计学问题包括描述统计、基础推断统计、线性回归分析等,它不仅能够准确计算出统计公式的结果、绘制出清晰的图表、给出详细的统计分析报表,甚至还能通过 VBScript 编程实现更多复杂的任务。除了制表功能和统计功能之外,Excel 还有强大的财务计算能力。因此,目前 Excel 在商业和金融领域运用尤其广泛,世界顶尖的投资银行、咨询公司的第一线分析员往往都把 Excel 作为最主要的分析软件。

本书的主要逻辑是介绍概率论和统计学的基础概念,以及初级统计分析的原理,在可能的情况下尽量不涉及过多的数学推理,并且尽可能把相应的计算工作交给 Excel 完成。因此,掌握统计学原理和熟悉软件操作对本书读者有同等重要的意义。如果读者到目前为止,对 Excel 的基本功能和简单操作还不太了解的话,建议参阅一下相关的 Excel 入门教程,本书在后面的内容中将着重介绍 Excel 的统计功能。

第六节 统计学中的常用概念及统计研究的程序

一、总体、总体单位、样本

总体和样本是统计学中两个最基本的概念,统计学研究的目的就是达到对总体数量特征的认识,而在绝大多数情况下,这个任务是通过对本体的研究完成的。

总体是指调研者所感兴趣的全部观察对象,由客观存在的具有相同性质的多个个体所形成的整体。例如,如果我们要了解自己所在学校的学生的手机使用情况,那么学校全体学生的集合就组成为一个总体。

总体单位是组成总体的每个个体,是调查项目的访问对象。在上例中总体单位即学校里的每个学生。

从总体中选取部分个体,所得到的来自总体部分对象的集合,称为样本。从总体中选取若干个个体的过程称为抽样,抽样的结果称为样本,样本中所含个体的数量称为样本容量。构成统计样本的每一个个体称为样本单位。继续前面的例子,如果我们要了解自己所在学校的学生的手机使用情况,通常很难做到把每个学生的情况都调查到,因此通过问卷调查或访问的方式,找到部分学生的手机使用情况,调查到的这部分同学的集合称为样本,所调查的每一个学生称为样本单位。

我们还将在本书的数理统计部分,对总体和样本这两个概念进行更多、更深入的讨论。

二、统计标志、统计指标与统计变量

统计标志,简称为标志,是说明总体单位所具有的属性或特征的名称。在学生使用手机情况的例子中,总体为全体在校生的集合,单位为每个学生,每个学生都拥有以下这些特征或属性:性别、年龄、身高、体重、家庭收入、籍贯等,这些名称都称为标志。所以可以看到,标志总是依附和说明总体单位的。

一个总体单位在某种特征上的具体表现即对标志的描述或取值,称为标志值或标志表现。标志表现反映总体单位的量的特征时,这样的标志为数量标志,用数值来表示。

标志表现反映总体单位的质的特征或属性时,这样的标志为品质标志。显然,年龄、身高、体重和家庭收入为数量标志,性别和籍贯为品质标志。数量标志和品质标志这两类标志在统计研究中都很重要,是形成统计指标的基础。

统计指标,简称为指标,是说明统计总体综合数量特征的名称。统计指标按所反映的总体特征的不同性质,分为数量指标和质量指标。数量指标能够直接或间接地说明总体的规模大小和数量多少,如某省总人口和粮食总产量,前者说明了某省人口总体的总规模,后者说明了该省粮食生产成果总体的总规模。质量指标能够说明总体内部数量关系或总体单位某种数量标志的标志值水平,如产品合格率和市场占有率。这两个指标反映了总体内部的数量关系。我们把平均成绩称为质量指标,因为它代表各单位标志值的一般水平。

本书的第三章第二节对统计指标进行了更详细的讨论。

统计变量是统计指标、数量标志和品质标志的总称,是说明现象某种特征的概念。变量的具体表现称为变量值。变量按表现形式不同可分为数量型变量(或定量变量)和属性变量(或定性变量)。如住房面积和收入为数量型变量,而文化程度和性别为属性变量。数量型变量又可以分为离散型变量和连续型变量。如果变量的取值可以一一列举,在相邻的两个数值之间不再其他的数值,这样的变量称为离散型变量,如在校学生人数和全国上市公司企业数。如果变量的取值不能一一列举,在任意两个数值之间都可以再取无限多的数值,这样的变量称为连续型变量,如人的年龄和股票的价格。

三、统计研究的程序

进行统计研究,是从调查与数据搜集开始的,经过数据的组织和整理,研究人员可以对数据进行列表和图示,从而对数据的总体情况进行描述统计分析。如果统计资料是作为总体搜集的,那么可以直接由对资料的分析得出关于总体的结论;如果统计资料只是作为样本搜集的,则在此基础上还需要进行推断统计分析,然后得出关于总体的结论。可以用图 1.1 来概括统计研究的程序。

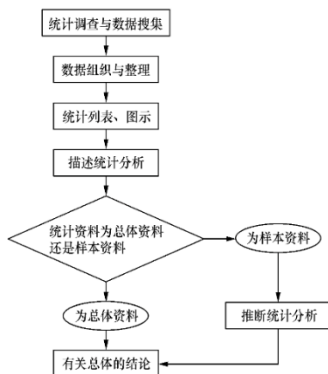


图 1.1 统计研究的程序

通过图 1.1 可以看到,统计研究的目的就是达到对总体的认识。数据作为统计研究的出发点,它的性质决定了统计研究的方法,即对于总体数据,描述统计分析就足够了;而对于样本数据,还需要借助推论统计工具,达到通过样本来认识总体的目的。在本书后面的章节,我们也将遵循统计研究的程序,逐步介绍每一个步骤中的工具和方法。

// 关键术语 //

计量描述 分析推论 总体 样本 抽样 数量特征 数量关系 描述统计学
推论统计学 理论统计学 数理统计学 应用统计学 概率论 国势学 政治算术
统计规律性 统计误差 Excel 指标 标志

习题》》

1. 什么是统计学? 统计学在研究对象和研究内容上有什么特点?
2. 试通过你所熟悉的事例谈谈搞经济理论研究的人为什么要学习统计学,从事经济工作的人又为什么要学习统计学。
3. 什么是描述统计学? 什么是推论统计学? 请各举一例说明。
4. “统计学研究的是现象的数量特征”这句话对吗? 为什么?
5. 请你根据最近媒体中对某一点问题的新闻调查,从统计学的角度回答下面的问题:
 - (1) 该调查感兴趣的研究对象是什么?
 - (2) 该调查对研究对象的哪些特征感兴趣?
 - (3) 该调查研究了样本单位的哪些属性和特征?
 - (4) 该调查有没有解释样本是如何得到的?
 - (5) 该调查有没有说明样本所包含的单位个数?
 - (6) 该调查构造或使用了什么指标?
 - (7) 该调查对总体特征做出了什么样的结论?
 - (8) 该调查有没有告诉你它所作推断的可信度?
6. 统计学目前已在多个学科领域得到极为广泛的应用和发展,你认为未来统计学会有哪些新的发展方向? 还可能在哪些新的领域得到怎样的应用?