

文章编号: 1005-0523(2003)03-0004-04

网络统计学的理论体系探讨

习勤

(华东交通大学 理工学院, 南昌 330013)

摘要: 从传播媒介的发展过程和网络经济的出现为背景, 阐述了网络统计的产生, 并结合统计理论、计算机技术和网络技术, 介绍了网络统计学的基本框架.

关键词: 网络; 统计学; 理论

中图分类号: F222.3

文献标识码: A

1 网络统计学的产生背景

科学技术发展, 尤其是传播媒介的发展也透视着统计的演变过程. 我们知道, 文字发明之前的原始统计是结绳记事、结绳计量; 人类第一次通信革命是发明文字, 它的出现使知识克服了时空的障碍, 因此, 有了用文字进行人口、粮食和土地的统计纪录; 印刷术是第二次通信革命, 它带来现代语言、民族国家、政治机构以及新闻报刊等, 由此出现了纸介质统计; 第三次通信革命是电信业的发展, 出现电话报、电报和传真报, 并与纸介质统计并存等; 第四次通信革命计算机和网络技术, 把文字、图形和声音转化为二进制的数字语言, 从而使传统统计发生了前所未有的变革, 磁介质统计替代纸介质统计, 并迅速发展成无介质统计. 电子统计, 即各部门和单位的统计数据借助联网计算机进行准确、及时地搜集、传输、处理与发布. 所以有人说, 现代计算机技术、网络通信技术的发展是网络统计产生的技术基础. 它不仅加速了统计技术现代化的步伐, 而且将引起统计管理体制、组织机构、工作方式的根本变革.

网络经济是与计算机网络、特别是与 Internet 有关的经济. Internet 就是众多计算机及其网络, 通过

电话线、光缆、通信卫星等连结而成的一个计算机网, 它将入网的不同类型的网络 and 不同机型的计算机互联起来, 构成一个整体, 从而实现了网络资源和信息的共享. 信息技术产业和信息服务产业是网络经济时代的主导产业, 统计属于信息服务产业中的一个重要组成部分. 全社会的信息化和网络化也必然推动统计工作的信息化步伐. 因此, 网络经济的出现为网络统计的产生起到催化作用.

2 网络统计学的理论基础

现代统计理论与方法是网络统计学的理论基础. 因为社会、经济、科技和生态环境之间的关系日趋密切、复杂多变, 传统统计理论在深入现象间关系研究中具有简单化和表象性, 难以揭示现象内在的矛盾性与统一性. 统计工作者在继承传统统计理论的基础上, 借助 新三论、老三论、模糊数学、灰色理论等的科学思维方法与原理原则, 推动着现代统计理论的建立和完善. 虽然其理论框架与内容是什么, 现在还不能准确阐述, 但现代统计理论发展的特征是非常明显, 那就是统计理论与方法的综合集成化和深度化. 所谓集成化是指, 统计理论是多种方法和工具以及多学科的综合集成; 所谓深度化是指, 统计数据加工不再限于表象分析, 而是对数

收稿日期: 2002-06-15

中国期刊网 <http://www.cnki.net> 习勤, 男, 江西宜春人, 副教授.

据(或信息)进行切片和旋转操作,以便从不同的角度考察数据、使用数据挖掘技术和数据导航技术,对大量企业的数据进行探索,揭示其中隐藏的规律性并进一步将之模型化等等。而这一切都要以电子计算机为手段、网络技术和网络资源为共享环境才得以实现。所以,现代统计理论是核心、基础。

与传统统计相比,网络统计具有以下特点:

第一互动性.所谓互动性是指统计工作过程是调查者与被调查者之间、统计信息供给者与统计信息需求者之间不再是一种单向的联系,而是始终保持着即时性的相互交流的过程,这种互动性是由网络超时空性质所决定的,尤其表现在调查过程中;

第二虚拟性.在大规模的统计调查中,利用网络虚拟技术,进行调查试点工作,检验调查方案的可行性,是一种费用省、时间快和效益好的方法。又如模拟企业内部的生产经营过程和企业外部环境因素变化,分析数据,为企业领导决策提供依据。

第三全球性.由于经济全球化趋势使得统计工作的着眼点不再是一个国家或地区范围内,而是要胸怀全球、放眼世界。网络统计平台利用 Internet 为我们认识世界政治、经济和社会发展的现状,掌握世界发展的大趋势开辟了一个大窗口。

第四共享性.网络统计提供的共享性不仅包括统计信息的资源的共享,还包括统计方法和统计应用程序以及各种网络资源的共享,如 E-mail 或 URL 或 IP 等,其内容之多,范围之大,已经不能与信息资源共享同日而语了。

第五营销性.统计信息是一种智力资源,它与资本、土地、劳动力一样是有价值的,是可以销售的。网络统计平台及统计网站,它是进行网络统计的载体,借助它来完成统计的设计、资料搜集、数据整理、数据分析、信息管理和信息发布以及销售等工作。一句话就是统计工作的推销站点。统计营销一方面表现在通过所占有的统计信息为企业经营决策提供咨询意见和建议;帮助企业进行市场调查项目的策划与实施;抽样调查理论与方法的应用性研究;调查设计与数据处理的技术服务;专业调查人员的技术培训等。

它是统计服务功能的延伸和发展。另一方面表现在网上定期发布国内外宏观经济信息和行业经济信息以及社会发展信息,或者通过统计网站与世界各国和本国重要的数据库链接。网络统计的营销性是统计工作的一种新理念,是发挥统计职能的具体体现。

第六安全性.安全和保密是网络统计中人们所关注的问题。由于网络应用的普及,促进了网络的安全技术的发展,常用数字签名、数据加密/解密技术、防火墙和 VPN 等。通过这些技术可以对一些敏感数据或重要数据的传输,提供数字签名和数据加密技术,防止数据被未经授权用户非法阅读,或者进行网络门户管理等。

3 网络统计学理论体系

网络统计学理论体系的建立,既要遵循统计工作的基本程序,又要考虑网络的一般特性进行设计,笔者认为它应由网络统计平台建设、网络统计设计、网络统计调查、网络统计整理与分析、网络统计数据库、网络数据仓库、网络统计环境等部分所构成。

3.1 网络统计平台建设

明确了网络统计的特点和作用,能够更好地对建设网络统计平台进行定位。网络统计平台是指,建立在 Web 平台基础上的,配之以相应的支持网络统计的应用软件和保证网络安全以及电子支付需要的软件。而网络统平台建设的关键是统计网站建设。它如同企业的销售网站一样,兼服务与营销两大功能。因此,统计站点建设应包括 ISP 的选择问题、选择接入 Internet 的方式问题、站点的设计和域名的申请问题等等。统计站点是宣传统计工作、提升统计形象、加强与上、下级统计部门或客户之间联系、提供统计信息和统计应用程序共享、交流工作经验、进行统计人员继续教育、从事统计科学研究的园地。网页设计的三要素即文本、图形和超链接要合理地、充分地加以使用,主页的导航设计中应包括三个层次,即政府统计工作、统计信息服务和统计营销。政府统计主要包括现行统计部门的工作;信息服务包括提供统计信息的检索、咨询、经验交流、专业人员继续教育等;统计营销包括有偿提供的所有的服务活动,如为企业作市场调研、进行企业决策咨询、培训统计人员等。

3.2 网络统计设计

网络统计设计是统计工作的基础环节,它由网络统计流程设计、网络化计算机程序设计、网络统计调查表设计和网络统计报表生成等部分组成。

1) 网络统计流程设计

网络统计一般程序

上级主管部门对下属机构的报表收集、统计工

作可以按照如下的工作方式进行:

(1)根据调查目的需要制定调查方案、设计调查表,并将它们放入服务器中;

(2)分布在各地的基层单位在电脑中填写一张张的调查表;

(3)在规定的时间内通过 Internet 将电脑中的调查表传送到统计主管部门;

(4)统计主管部门则根据收集到的信息由服务器产生统计报表;

(5)将统计报表通过网络有控制地发布到有关的部门.

2) 网络化计算机程序设计, 主要指在服务器/客户机环境下的计算机程序, 它包括通用程序和专用程序, 前者可以自行编制, 也可以用 SAS 或 SPSS 等通用程序进行开发; 后者根据特定的项目如人口普查、入户调查等分别设计, 其基本功能有: 编码、录入、编辑、计算和编表输出等.

3) 网络统计调查表设计

调查表是统计调查的基本工具, 网络调查表与一般调查表格式基本上没有差异, 所不同的有两点: 一是设计调查表的工具, 或者说利用网页制作的工具来实现一般调查表的设计, 例如, 问卷调查表设计中, 采取表单设计工具, 可以达到单项选择、多项选择、滚动文本、下拉菜单、单行文本等多种功能设计; 二是根据网络的特点, 调查问卷表不宜过长, 问句应简洁易懂, 定义清楚, 尽量采取选择答案方式, 简单、易答. 并使用文本、图形、表格、视频和音频等多种媒体组合, 使调查表丰富多彩. 因此, 要重点阐述调查表的概念、种类、设计原则和设计方法.

4) 网络统计服务工具(SFAQ)的设计

网络统计服务工具(SFAQ)页面是所有统计网站必有的项目, SFAQ 的英文全名是 Statistics Frequently Asked Questions, 即统计常见问题页面. 主要内容有: 调查方案解释、统计的指标解释、调查技巧问题、调查中的注意事项等等.

3.3 网络统计调查

网上调查是网络统计的重要内容, 它将阐述网上调查的概念、种类特点和开展网上调查的方法. 根据统计研究目的和任务的需要, 通过因特网进行统计资料搜集的方式和方法. 它是未来统计部门搜集统计原始资料的有效工具. 它与传统调查具有没有时空、地域限制, 组织简单, 费用低, 交互性好, 客观性强和时效性高以及便于质量控制等特点. 其调

查方法有站点法、电子邮件法、随机 IP 法、视讯会议法和其他调查法(计算机辅助调查), 这些方法中, 可以按照普查、重点调查、抽样调查等多种形式进行组织. 目前西方发达国家正在推广使用. 如美国劳工统计局最近提出一种基层单位调查数据搜集可供选择的方法. 利用在美国以及在国外迅速发展的电子邮件和 Internet, 将自动生成的调查表, 发往被调查者的电子邮件地址, 被调查者将在他们方便时, 通过电子邮件填写, 并回答表格问题, 对回答进行自动检查和编辑, 错误的回答通过电子邮件原路返回, 被调查者重新更正.

3.4 统计 EDI 系统

EDI 是英文 Electronic Data Interchange 的缩写, 中文可译为“电子数据交换”, 它是一种在公司这间传输订单、发票等作业文件的电子化手段. 它通过计算机通信网络将贸易、运输、保险、银行和海关等行业信息, 用一种国际公认的标准格式, 实现各有关部门或公司与企业之间的数据交换与处理, 并完成以贸易为中心的全部过程. 它是计算机、通信和现代管理技术相结合的产物. 而今 Internet 已渗透到各行各业, 为人类活动创造了一个信息传递既方便又低价的空间.

统计 EDI(SEDI)是指统计的电子数据交换, 通俗地讲就是标准化的统计数据文件在计算机之间从应用到应用的传输过程. 统计 EDI 适用于高度集中的统计体制. SEDI 系统建设可以促进统计工作的标准化、减少人为干扰统计数字、加快信息传递、可靠性强、安全和保密. SEDI 术语统计调查的部分, 由于它是一种应用前景非常好的调查方式和手段, 因此, 单独加以研究. SEDI 包括内容有, EDI 概念、EDI 的原理、SEDI 的特点、SEDI 的基本构成以及建立 SEDI 需要解决的关键问题等等. 荷兰统计局局长沃特·J·凯勒博士说: 到 2000 年, 电子数据交换将成为经济调查数据所选择的搜集方法. 至少对一些国家来说, 这是很可能成为现实. 美国普查局的威廉·L·尼克尔斯 II 指出: 电子数据交换在未来 10 年到 20 年的经济数据搜集中将肯定发挥越来越大的作用.

3.5 网络统计数据库

在网络环境下, 统计数据库是沟通统计部门与上级领导以及企业或调查者之间的联系的重要内容和手段, 也是网络统计调查、整理和分析的基础. 这里将阐述网络数据库的概念、类型、网络统计数据库的建立和运用中的注意事项等.

3.6 数据仓库与数据挖掘技术

数据仓库是面向多主题的、随时间变化的大量数据、硬件平台和有关的各种工具集成起来,用户从中运行查询、制作报表,进行数据分析,用以支持管理决策的工具。它具有丰富的数据采集、数据管理和数据分析的能力;具有丰富的信息描述和信息展现能力。数据挖掘技术是数据仓库中最好的决策支持工具,它能够帮助统计工作者和企业从堆积如山、势如洪水的大量数据中挖掘出隐藏着的规律性,为政府和企业决策者及时做出正确的宏观决策和经营决策提供依据。

数据挖掘技术应用领域非常广泛,包括宏观经济运行分析、市场营销分析、欺诈甄别分析、加载和更新、产品制造和科学研究等。数据挖掘技术的常用软件是SAS的SAS/EM,它是一个图形化界面的、菜单驱动的对用户非常友好的、功能强大的数据挖掘集成软件,其集成的工具有:数据获取工具、数据取样工具、数据筛选工具、数据变量转换工具、数据挖掘数据库、数据挖掘过程、多种形式的回归工具、人工神经网络等等。

3.7 网络统计的环境建设

中国电子商务网公布了其对近300家企业所做的网络安全问题抽样调查结果。调查数据表明,目前企业存在的主要安全隐患是,信息被篡改占44%,信息失窃占40%,被病毒破坏占40%,被恶意攻击占32%,每种情况都有的占4%;企业采取的主要信息安全措施有,72%的采用防火墙,48%的采用防病毒软件,采用安全网关的占20%,采用信息加密的占8%,采用其它措施的为0%;从调查显示的结果看,很多企业已经开始重视网络安全。随着网络统计的建立,网络安全同样成为统计工作中一个非常重要的问题。

1) 网络统计安全建设

网络安全是指计算机信息网络安全是指网络中的硬件、软件和数据不受自然和人为有害因素的危害,保证正常运行。网络安全是一个系统概念,它包括硬件安全、软件安全、数据安全和运行安全等几个方面。这里的内容包括网路统计安全的概念、网络安全的等级、网络安全体系、网络安全核心技术以及网络统计安全管理等部分。

2) 网络统计的法律规范

与网络安全密切相关的问题就是建立网络统计法律规范和健全统计法律体系。我们可以从两个方面看:一方面根据现有的有关法律进行管理。我国《刑法》中已经制定了一些有关惩治计算机犯罪方面的条款。比如第二百八十五条规定:违反国家规定,侵入国家事务、国防建设、尖端科学技术等重要领域的计算机信息系统的,处三年以下有期徒刑或者拘役。这些条款同样适用于对网上犯罪的惩治。另一方面在《统计法》中增加有关针对网络统计犯罪的专门条款等。

网络统计是一个新生事物,是统计理论的创新,其理论与方法的建立不仅需要统计工作者参与,而且需要计算机人员、网络技术人员和法律工作者等的共同参与才能够完成,缺少任何一个方面都不可能建立起科学的网络统计理论体系,这是网络统计与其他统计的根本区别,也是网络统计的一大特色,因此,可以称之为高新技术统计理论。

参考文献:

- [1] 张福德. 电子商务导论[M]. 北京: 科学出版社, 1999.
- [2] 王道平. 网络经济[F]. 石家庄: 河北人民出版社, 1999.
- [3] 张廷茂. 网络营销[F]. 石家庄: 河北人民出版社, 1999.
- [4] 佐尔坦, 肯尼西, 张泽厚等译. 统计的未来, 国际统计发展展望[F]. 北京: 中国统计出版社, 1998.

A Probe into the Theoretical System of Network Statistics

XI Qin

(School of Science and Engineering, East China Jiaotong Uni., Nanchang 330013, China)

Abstract: This paper, under the background of the appearance in network economy and the development course of spreading medium, elaborates on the development of network statistics and introduces the basic frame of network statistics through a combination of statistical theories, computer technology and network technology.

Key word: network; statistics; theory