

普通高等院校数学类规划教材

应用统计学简明教程

Applied Statistics

大连理工大学城市学院基础教学部 组编



大连理工大学出版社
DALIAN UNIVERSITY OF TECHNOLOGY PRESS

普通高等院校数学类规划教材

C8
148

应用统计学简明教程

Applied Statistics

大连理工大学城市学院基础教学部 组编

主编 曹铁川

编者 孙晓坤 王淑娟 佟小华



大连理工大学出版社
DALIAN UNIVERSITY OF TECHNOLOGY PRESS

图书在版编目(CIP)数据

应用统计学简明教程 / 大连理工大学城市学院基础
教学部组编. — 大连 : 大连理工大学出版社, 2014. 1
ISBN 978-7-5611-8502-5

I. ①应… II. ①大… III. ①应用统计学—教材
IV. ①C8

中国版本图书馆 CIP 数据核字(2014)第 015279 号

大连理工大学出版社出版

地址:大连市软件园路 80 号 邮政编码:116023

发行:0411-84708842 邮购:0411-84703636 传真:0411-84701466

E-mail:dutp@dutp.cn URL:<http://www.dutp.cn>

丹东新东方彩色包装印刷有限公司印刷 大连理工大学出版社发行

幅面尺寸:185mm×260mm	印张:12.25	字数:283 千字
2014 年 1 月第 1 版		2014 年 1 月第 1 次印刷

责任编辑:王 伟

责任校对:李 慧

封面设计:熔点创意

ISBN 978-7-5611-8502-5

定 价:28.00 元

前 言

应用统计学是一门通过数据的收集、整理、分析和解释,来探索客观事物内在规律的方法论科学。它为许多学科和社会经济管理的数据分析,提供了理论依据和分析方法。随着社会发展和科学技术的不断进步,数据形式更趋多样性,加之计算机技术的应用和普及,应用统计的理论与方法在社会生活中的作用愈发显得重要。

应用统计学的研究对象是统计数据。针对数据类型的特点,对数据进行分析研究,并得到相应的结论,是科技工作者和工程技术人员应具备的一种能力。根据普通高等院校,特别是应用技术型大学的实际需求,我们编写了《应用统计学简明教程》。该书借鉴了国内外优秀的统计教材,对统计学的知识体系进行了适当重组,特别强调了知识方法的多面性与实用性,注意选择具有实际背景的内容,弱化理论性推导,重在介绍数据的处理、分析、解释的方法。

本书第1章主要介绍统计学及其基本概念;第2~4章为描述统计学部分,包括统计数据的收集、整理与图示,统计数据的描述统计量、统计指数;第5~7章为概率论部分,包括概率论的基础知识、二项分布与泊松分布、正态分布;第8~11章为推断统计学部分,包括置信区间、假设检验、一元线性相关与回归、时间序列。

本书由大连理工大学城市学院基础教学部组织编写,曹铁川任主编并负责统稿,编者有佟小华(第1章、第2章)、孙晓坤(第3~7章、第9章)、王淑娟(第8章、第10章、第11章)。

由于作者水平有限,不妥之处在所难免,期待读者和同行批评指正。

编 者

2014年1月于大连

目 录

第 1 章 统计学及其基本概念 /1

1.1 两个引例 /1

1.2 统计学及其应用 /2

1.2.1 统计学的应用 /2

1.2.2 统计研究的基本过程 /4

1.3 统计学的基本概念 /4

1.3.1 总体和样本 /4

1.3.2 参数和统计量 /5

1.3.3 变量和数据 /5

习题 1 /6

第 2 章 统计数据的收集、整理与图示 /7

2.1 引 例 /7

2.2 统计数据的收集 /8

2.2.1 数据的来源 /8

2.2.2 原始数据的收集方法 /9

2.2.3 统计调查的方式 /10

2.2.4 调查方案的设计 /11

2.2.5 统计调查的工具 /12

2.3 统计数据的整理 /13

2.3.1 数据整理的步骤 /13

2.3.2 品质型数据的整理 /13

2.3.3 数值型数据的整理 /14

2.4 常见的统计图 /17

习题 2 /20

第 3 章 统计数据的描述统计量 /22

3.1 引 例 /22

3.2 离散型数据的描述统计量 /23

3.2.1 离散型数据的集中趋势 /23

3.2.2 离散型数据的离中趋势 /26

3.3 连续型数据的描述统计量 /30

3.3.1 连续型数据的集中趋势 /30

3.3.2 连续型数据的离中趋势 /33

3.4 其他描述统计量 /36

3.4.1 集中趋势 /36

3.4.2 离中趋势的相对度量

——离散系数 /38

3.4.3 数据对称状态的度量

——偏态系数 /39

习题 3 /40

附录:样本标准差计算公式的推导 /43

第 4 章 统计指数 /44

4.1 引 例 /44

4.2 统计指数的概念 /45

4.3 统计指数的编制方法 /47

4.3.1 简单指数 /47

4.3.2 加权指数 /51

4.3.3 价值指数 /54

4.4 指数的基期转移、拼接 /56

4.4.1 指数的基期转移 /56

4.4.2 指数的拼接 /56

4.5 几种常见的统计指数 /58

4.5.1 零售价格指数 /58

4.5.2 股票价格指数 /60

习题 4 /61

第 5 章 概率论的基础知识 /64

5.1 随机事件及其运算 /64

5.1.1 随机试验 /65

5.1.2 样本空间 /65

- 5.1.3 随机事件 /66
- 5.1.4 事件间关系与运算 /66
- 5.2 概率的定义及其运算 /69
 - 5.2.1 概率的统计定义 /69
 - 5.2.2 古典概型 /70
 - 5.2.3 概率的公理化定义及性质 /72
 - 5.2.4 条件概率及乘法公式 /73
 - 5.2.5 事件的独立性 /76
- 5.3 概率树 /77
 - 5.3.1 全概率公式与贝叶斯公式 /77
 - 5.3.2 相互独立事件的概率树 /79
 - 5.3.3 相关事件的概率树 /80

习题 5 /82

附录:排列组合的定义及计算公式 /85

第 6 章 二项分布与泊松分布 /87

- 6.1 随机变量 /87
- 6.2 离散型随机变量及其分布律 /88
- 6.3 二项分布 /89
- 6.4 泊松分布 /91
- 6.5 离散型随机变量的数字特征 /92
 - 6.5.1 数学期望 /93
 - 6.5.2 随机变量的方差 /94
 - 6.5.3 二项分布与泊松分布的期望与方差 /96
- 6.6 决策树 /97

习题 6 /100

附录:泊松分布逼近二项分布的证明 /103

第 7 章 正态分布 /104

- 7.1 连续型随机变量及其概率密度函数 /104
- 7.2 正态分布的概念 /105
- 7.3 标准正态分布 /106
- 7.4 正态分布的计算 /111
- 7.5 相互独立的正态分布随机变量的和与差的分布 /113
- 7.6 抽样分布 /114

7.6.1 抽样分布的概念 /114

7.6.2 单一样本均值的抽样分布 /114

7.6.3 其他统计量的抽样分布 /118

习题 7 /120

第 8 章 置信区间 /122

- 8.1 引 例 /122
- 8.2 两个基本概念 /122
- 8.3 单一总体参数的置信区间 /124
 - 8.3.1 单一正态总体均值(μ)的置信区间 /124
 - 8.3.2 单一总体比例(π)的置信区间 /128
 - 8.3.3 样本容量的确定 /129
- 8.4 两个总体参数的置信区间 /130
 - 8.4.1 两个独立总体均值差($\mu_1 - \mu_2$)的置信区间 /130
 - 8.4.2 两个总体比例差($\pi_1 - \pi_2$)的置信区间 /132

习题 8 /133

第 9 章 假设检验 /136

- 9.1 引 例 /136
- 9.2 假设检验的一般问题 /137
 - 9.2.1 假设的提出 /137
 - 9.2.2 假设检验的过程 /137
- 9.3 对单一总体均值的假设检验 /139
 - 9.3.1 正态总体方差 σ^2 已知或大样本抽样时的检验 /139
 - 9.3.2 正态总体方差 σ^2 未知且小样本抽样时的检验 /142
- 9.4 对单一总体比例的假设检验 /144
- 9.5 对两个总体均值差异性的假设检验 /145
 - 9.5.1 两正态总体方差 σ_1^2, σ_2^2 均已知,或大样本抽样时的检验 /146
 - 9.5.2 两正态总体方差 $\sigma_1^2 = \sigma_2^2$

均未知,且小样本抽样 时的检验 /148	11.2 时间序列的构成因素与模型 /168
9.6 对两个总体比例差异性的 假设检验 /149	11.2.1 时间序列的构成因素 /169
习题 9 /150	11.2.2 时间序列的模型 /170
第 10 章 一元线性相关与回归 /153	11.3 时间序列图 /171
10.1 引 例 /153	11.4 长期趋势的测定 /172
10.2 散点图 /154	11.4.1 线性回归趋势 /172
10.3 线性相关程度的测度 /157	11.4.2 移动平均趋势 /174
10.3.1 序数型数据 ——等级相关系数 /157	11.5 季节性因素的测定 /177
10.3.2 基数型数据 ——相关系数 /158	11.5.1 利用加法模型求 季节性因素 /177
10.4 可决系数 /159	11.5.2 利用乘法模型求 季节性因素 /179
10.5 一元线性回归分析 /160	11.6 时间序列的预测方法 /180
10.6 预测分析 /162	习题 11 /181
习题 10 /164	附 表 /184
第 11 章 时间序列 /168	附表 1 标准正态分布表 /184
11.1 引 例 /168	附表 2 t 分布表 /185
	参考文献 /186

(续表)

单词使用的频数	不同的单词数
722	8
822	7
922	6
1022	5
1122	4
1222	3
1322	2
1422	1

第1章 统计学及其基本概念

统计学是伴随着人类社会进步产生的一门学科,并在实践中不断加以发展。在当今社会经济管理许多领域中,统计学发挥着越来越重要的作用,是进行决策和科研的有力工具。本章将通过两个引例,简单介绍统计学的概念和一些应用,并介绍统计学的基本概念。

统计学是伴随着人类社会进步产生的一门学科,并在实践中不断加以发展。在当今社会经济管理许多领域中,统计学发挥着越来越重要的作用,是进行决策和科研的有力工具。本章将通过两个引例,简单介绍统计学的概念和一些应用,并介绍统计学的基本概念。

1.1 两个引例

【引例1】 第二次世界大战期间,统计学家 A. Wald 被咨询到这样一个问题:拟在战斗机机身上加强钢板,以确保能最大限度地抗打击,同时机身又不能太重,以保证战斗机灵活战斗,问应在什么位置上加强钢板?

A. Wald 经考虑后,画了一架战斗机模型,并将返航战斗机机身上的弹孔位置逐一标注在模型上。经过一段时间后,发现整个机身基本被填满了弹孔,但有几小块位置上没有弹孔。于是,他建议在少数这几小块无弹孔的位置上加强钢板。因为这几小块被击中的战斗机都没能安全返回。

【引例2】 1985年11月,研究莎士比亚的学者泰勒(G. Taylor)在英国 Bodleian 图书馆的收藏品中,发现了写在纸片上的九节新诗,该诗只有429个字,但没有记载作者是谁。

两个统计学者 Thisted 和 Efron 利用统计方法,对这首新诗进行研究,得到结论:这首诗用词的风格与莎士比亚的风格非常一致。其实,他们的研究方法纯粹基于统计学基础,其过程可简单描述如下。

已知莎士比亚所有著作的用词总数为884647个,其中不同的单词数为31534个。这些词出现的频数见表1-1。

表 1-1

单词使用的频数	不同的单词数
1	14376
2	4343
3	2292
4	1463
5	1043

(续表)

单词使用的频数	不同的单词数
6	837
7	638
⋮	⋮
大于 100	846
总数	31534

表 1-1 中所包含的信息说明:如果要求莎士比亚写一个含有一定数量单词的新作品,他会使用多少新单词(以前作品中未使用过的);在他以前所有的作品中,有多少单词他仅使用过一次,两次,三次等.这些数字都可以用费舍尔(Fisher)等提出的一种统计方法来预测(费舍尔曾用他提出的方法估计了未被发现的蝴蝶总数).

在这首新发现的含有 429 个单词的诗(简称新诗)中,有 258 个单词不同.新诗的观测值和预测值(基于莎士比亚的风格)的分布见表 1-2(最后两竖栏).从表 1-2 可以看到,在所期望的差值范围内,两个分布比较一致,这表示新诗的作者可能就是莎士比亚.

为了对比,表 1-2 中也给出了与莎士比亚同时代的其他几位诗人本·约翰逊(B. Johnson)、马洛(C. Marlowe)、多恩(J. Donne)的长度几乎相同的作品中,所使用单词的频数分布.总体上看,这些作者的作品中,单词的频数分布与新诗中单词的观测频数,以及与莎士比亚用词风格的期望观测频数之间,看起来多少有些不同.

表 1-2

莎士比亚作品中 单词使用的次数	其他诗人不同单词使用的频数			发现的 新诗	基于莎士比亚 作品的期望值
	本·约翰逊 (哀歌)	马洛 (四首诗)	多恩 (狂喜)		
0	8	10	17	9	6.97
1	2	8	5	7	4.21
2	1	8	6	5	3.33
3 ~ 4	6	16	5	8	5.36
5 ~ 9	9	22	12	11	10.24
10 ~ 19	9	20	17	10	13.96
20 ~ 29	12	13	14	21	10.77
30 ~ 39	12	9	6	16	8.87
40 ~ 59	13	14	12	18	13.77
60 ~ 79	10	9	3	8	9.99
80 ~ 99	13	13	10	5	7.48
不同单词数的单词总数	243	272	252	258	258
总数	411	495	487	429	

1.2 统计学及其应用

1.2.1 统计学的应用

在日常生活中,人们经常接触“统计”的内容.例如,企业每年都统计产量和产值;了

解股票的交易状况要统计有关成交额和股票指数。又如,人们也经常在报纸和杂志的文章中看到类似如下的报道:

- 调查发现,有 82.8% 的消费者赞同在餐厅划分吸烟区和无烟区,其中吸烟者中赞成的比例为 68.3%,不吸烟者中赞成的比例高达 90.7%。

- 数据显示,1~3月,全国商品房销售面积同比增长 8.2%,销售额同比增长 23.1%;其中商品住宅销售面积增长 8.7%,商品住宅销售额增长 24.7%。

- 有关部门公布的数据显示,去年第一季度,国内汽车产和销分别为 256.76 万辆和 267.88 万辆,产量同比增长 1.91%,销量同比增长 3.88%。

在上述报道中,与数据有关的资料,都称为统计资料,用于研究统计资料的科学称为统计学(statistics)。

统计作为一种社会实践活动已有悠久的历史。据历史记载,我国在西周朝代就已建立了统计报告制度。统计一词,英文为 statistics,最早源于拉丁文的“国会”和意大利文的“国民或政治家”,是代表“研究国家的科学”,因此可以说,自从有了国家,就有了统计实践活动。统计最初只是作为一种计数活动,用于为统治者治理国家而收集资料,为国家管理提供人力、物力和财力的依据。随着社会经济和科技的发展,以及统计学自身的进步,其应用领域不断扩大。

统计学与人类社会生活关系密切,它可以帮助我们认识客观世界的发展规律,测度事物之间的数量关系,准确描述事物发展状态。统计学通过大量的试验,观察,调查等方式来收集数据,并且利用适当的数据分析方法,揭示研究对象的本质、特征及发展趋势,从而做出相应的决策。

那么什么是统计学?它所研究的是什么问题?

简而言之,统计学是关于现象数据的收集、整理和分析的方法论科学,其目的是探究数据内在的数量规律性,以达到对现象的科学认识。

由于各个领域都要研究现象的数量信息,因此统计学已渗透到各个学科领域,上至国家的宏观管理、企业的微观调控,下到百姓的日常生活都离不开统计学。

下面简单介绍统计学的一些实际应用。

(1) 经济预测

经济学家经常要对经济或市场状况进行预测,在进行这种预测时,需要使用各种统计信息。比如说,经济学家要进行通货膨胀预测,就要参考相关的统计资料,如生产价格指数、零售价格指数和失业率等,把这些相关因素输入到预测通货膨胀的相应模型中,就可以进行预测。

(2) 市场研究

企业在激烈的市场竞争中若要取得优势,就必须要了解市场。比如,超市的零售结账柜台的电子扫描系统,可以收集市场调研需要的数据,通过分析这个数据能更好地对企业的进货和库存进行管理。

(3) 财务分析

企业的财务顾问要根据各种各样的统计信息,来指导企业进行合理投资。比如在股票市场中,可根据上市公司所提供的财务数据等资料给出买、卖或持有股票的建议。

(4) 质量控制

统计质量控制图可用于监测生产过程的产出。

(5) 人力资源管理

利用统计学方法对企业员工的年龄、性别、教育程度、工资水平等相关资料进行管理和分析,为制订企业工资计划和惩罚制度提供依据。

1.2.2 统计研究的基本过程

统计研究工作一般是从收集数据入手,在得到大量的原始数据之后,根据研究的目的和要求,对原始数据进行必要的整理,再据此对现象进行分析和解释。因此统计活动就是采用统计方法,对某一现象进行分析研究的过程。

1. 数据的收集

统计数据的收集,是指为研究某一现象,搜集大量相关数据的过程,它是统计工作的基础。统计数据的质量直接影响统计工作的质量,只有科学有效地获取数据资料,才能为统计分析与决策提供有力的依据。统计数据的收集可以采用统计调查、观察或实验等方式来获得。

2. 数据的整理

统计数据的整理,是指对收集到的大量的、零散的数据进行整理,得到系统的、有条理的数据的过程。数据整理的过程包括数据的审核、数据的分组、数据的汇总、数据的显示等步骤。

3. 数据的分析

数据的分析是统计学的核心内容,是指应用统计学的方法,对数据的特性进行定量的分析,以达到了解数据本身规律的目的。

4. 数据的解释

数据的解释,是利用分析出来的结论,对研究的对象进行解释,并对以后的工作起到一定的参考作用。

1.3 统计学的基本概念

统计学中的概念众多,下面介绍几个经常用到的基本概念。

1.3.1 总体和样本

1. 总体

统计学是研究客观现象数量特征和数量关系的科学,因此,首先对研究对象的全体要有一个明确的认识。

所谓总体(population),是根据分析目的确定的研究对象全体的集合。总体中每个对象称为个体。总体和个体可以是人、物、机构等。例如,对某公司员工的收入进行统计,该公司的所有员工就构成总体,其中每一个员工就是一个个体;某工厂质检部门对某一批产品

的质量进行检验,该批产品构成总体,每一个产品是一个个体。

总体中个体的数量叫做总体容量,记作 N 。总体的容量可以是有限的,也可以是无限的。例如,研究某公司员工的收入时,由全体员工构成的总体是有限总体;分析工厂流水线产品时,由于产品的数量在不断变化,由流水线生产的产品所构成的总体是无限总体。再如,要检验一批灯泡的寿命,这批灯泡构成的集合就是总体。但是,在分析的时候,我们关心的其实是每只灯泡的寿命,而不是灯泡本身,所以,实质上总体是由每只灯泡的寿命值构成的集合,也就是一些实数构成的集合。通常,如果总体由有限个实数构成,为有限总体;如果总体由无限个实数构成,则为无限总体。

2. 样本

统计学研究的是总体的数量特征,但当总体中个体数量很多甚至无限多时,就很难直接对所有个体进行调查分析。这时可按一定的原则,从总体中抽取一部分个体,作为总体的代表加以研究。

所谓样本(sample),即指从总体中选出一定数量的个体构成的集合。

从总体中抽取样本,研究样本的信息,再据此去推断总体的特征,是统计学研究中经常用到的方法。例如,想了解某城市家庭的年收入,可抽取一定数量的家庭作为样本加以推断。

除了对容量较大的总体需抽取样本以外,若为了解总体的特征所做的试验为破坏性试验,也必须采用这种抽样调查的方式。例如,某工厂欲了解一批灯泡的合格率,需对受检灯泡进行寿命亮度等检测,而这种检测是一种破坏性的试验,所以必须采取抽样的方式进行调查。

样本中个体的数量,称为样本容量,记作 n 。样本的容量究竟应取多少,要根据实际问题决定。

1.3.2 参数和统计量

1. 参数

参数(parameter),或称总体参数,是用来描述总体数量特征的概括性值。常见的参数有总体的均值,记为 μ ;总体的标准差,记为 σ ;总体的比例,记为 π ,等。

在进行总体推断时,总体数据通常是不完全的,所以参数一般表现为一个未知的常数。要了解参数的真实值,就要进行抽样,根据样本信息去推算总体的参数值。

2. 统计量

统计量(statistic),又称样本统计量,是用来描述样本数量特征的另一个重要概念。

统计量是样本的函数,可以根据样本计算得到。常见的统计量有样本的均值 $\bar{X}$,其值记为 $\bar{x}$;样本的标准差 S ,其值记为 s ;样本的比例 P ,其值记为 p ,等。

1.3.3 变量和数据

1. 变量与数据的概念

变量(variable),是用于说明个体某种特征或属性的名称。

例如,“性别”、“身高”、“商品的销售额”、“产品的质量等级”等,都是变量。变量的具体取值,即变量值,称为统计数据,简称数据(data)。

例如,性别的取值为男,女;身高(单位:cm)的取值可以为120,135,158,176,181,...

2. 变量与数据的类型

根据具体情况,变量可以分为品质型变量和数值型变量。

(1) 品质型变量与数据

品质型变量,又称定性变量,用于反映事物品质特征,是可用文字来表述的变量,多表现为类别。

例如,“性别”和“产品的质量等级”均是品质型变量。

按其类别情况的不同,品质型变量可分为分类变量和顺序变量。分类变量主要用于区别各类别的不同,例如,“性别”就是分类变量;又如,“企业经济类型”也是分类变量,表现为国有企业、集体企业、私营企业、合资企业、独资企业等。顺序变量不但区分出各类别,还体现出各类别之间的优劣或等级,例如“产品的质量等级”就是顺序变量。

品质型变量的取值,为品质型数据,表现为文字形式。例如,性别的取值为“男”和“女”,产品质量等级的取值为“一等品”、“二等品”、“三等品”、“次品”。

(2) 数值型变量与数据

数值型变量,又称定量变量,是用于反映事物数量特征的变量,通常用具体的数值来表述。

例如,“身高”和“商品的销售额”都是数值型变量。

数值型变量按照其取值是否连续,可以分为离散型变量和连续型变量。

如果数值型变量的取值是有限个,或是无限可列个,则为离散型变量。例如,某个班级32名学生的期末考试成绩,就是离散型变量。

如果数值型变量可以在一个或多个区间中取任何值,则为连续型变量。例如,某种灯泡的寿命、人的身高、体重等,都属于连续型变量。

数值型变量的取值,称为数值型数据,其表现为具体的数值。例如上述身高(单位:cm)的取值,又如商品销售额(单位:万元)的取值可以是10,20,...。根据数值型变量的类别,数值型数据分为离散型数据和连续型数据。特别地,连续型数据以区间的形式表述。

习题 1

1-1 什么是统计学?统计学研究的基本过程是什么?

1-2 举例说明总体、样本、参数、统计量、变量和数据的概念。

1-3 变量可以分为哪几类?指出下列变量的类型。

(1) 年龄;

(2) 考试成绩;

(3) 体重;

(4) 购买商品的支付方式(现金、信用卡、支票)。

1-4 某城市的研究机构欲了解该城市家庭的年平均收入,随机抽取1000个家庭的年收入作为样本。

(1) 描述总体和样本;

(2) 指出参数和统计量。

第2章 统计数据的收集、整理与图示

统计数据是统计分析的基础,只有科学有效地获取数据,并对数据进行科学的分类、汇总和显示,使之清晰地反映总体特征,才能为分析决策提供正确的依据。本章主要介绍统计数据的收集方法、统计数据的整理以及用统计表和统计图显示整理结果的方法等。

2.1 引例

某学院某年对大学生消费的问卷调查如下:

调查目的:了解学生的日常生活消费水平

调查对象:学院的全体学生

调查方式:抽样调查

样本容量:200人

调查工具:调查问卷

问卷题目如下:

1. 你每年的学费、住宿费是多少?

(一般在 5000 ~ 8000 元)*

2. 你每月的固定生活费大约是多少?

A. 500 元以下(80 人)

B. 500 ~ 800 元(96 人)

C. 800 ~ 1100 元(16 人)

D. 1100 元以上(8 人)

3. 你每月生活费的消费基本构成:

基本生活消费(衣食住行等): (平均约 400 元)

学习消费(书费、考证等): (平均约 50 元)

休闲娱乐消费(旅游、休闲等): (平均约 80 元)

人际交往消费(人情、交友等): (平均约 120 元)

其他: (平均约 120 元)

4. 你每月生活费的来源为

A. 父母供给(160 人)

B. 依靠自己的能力来支付(30 人)

* 选项后面的标注是抽样调查的汇总结果。

- C. 靠社会救助(2人) D. 靠助学贷款(8人)
5. 你是否拥有一部自己的手机? ()
- A. 是(162人) B. 否(38人)
6. 当同寝室其他人都有手机时,你会 ()
- A. 考虑买一部(88人)
- B. 目前我还用不到,等工作再说(72人)
- C. 经济状况不允许,同时也承担不了使用费用(40人)
7. 购物时,你首先考虑的因素是什么? ()
- A. 品牌(36人) B. 价格(50人)
- C. 自己的爱好(96人) D. 个性(18人)
8. 你的金钱观是 ()
- A. 金钱至上(8人)
- B. 够用就行(64人)
- C. 视金钱如粪土(4人)
- D. 钱不是万能的,但没有钱是万万不能的(124人)
9. 你的家庭状况属于 ()
- A. 贫困(20人) B. 一般(130人)
- C. 中等偏上(35人) D. 富裕(15人)
10. 你每月花费占全家月支出的比例为 ()
- A. 30%以下(75人) B. 30%到50%(110人)
- C. 50%以上(15人)
11. 你是否参加过勤工助学? ()
- A. 从来没有(15人) B. 偶尔一两次(165人)
- C. 经常(20人)

2.2 统计数据的收集

用统计方法分析实际问题离不开数据,因此,统计数据的收集是统计研究的基础工作之一。

2.2.1 数据的来源

从统计数据本身的来源看,最初都是来自于直接的调查或实验.但从使用者的角度看,数据的来源主要是两种渠道.一是直接的调查或科学实验.用这种方式获得的数据称为原始数据(primary data),又称一手数据.例如,引例中调查问卷得到的数据就是原始数据.数据来源的另一种渠道是使用现有的数据,也就是使用别人调查或实验的数据,用这种方法获得的数据称为间接数据(secondary data),又称二手数据,常见的如《中国统计年鉴》.在我国,间接数据主要来自国家和地方的统计部门,以及各种报刊媒体公开出版的统

计数据。随着计算机网络技术的发展和普及,网络也成为获得间接数据的一种重要渠道。

下面主要从使用者的角度介绍原始数据的收集方式和方法。

2.2.2 原始数据的收集方法

原始数据主要来源于统计调查,也可以由观察或实验获得。统计调查是取得社会经济数据的重要手段,多以询问的方式实现。而实验则是取得自然科学数据的主要手段。

1. 询问调查

询问调查是由调研人员事先制订好调查提纲,再通过被调查者回答相关的问题,来收集资料获得信息的一种方法。具体包括访问调查、邮寄调查、电话调查、电脑辅助调查、座谈会、个别深访等形式。

(1) 访问调查又称派员调查,是调查者与被调查者通过面对面交谈,得到所需资料的调查方法,多依赖于调查问卷等工具进行。其优点是能够对调查过程加以控制,从而获得比较可靠的调查结果,这种方法在市场调查和社会调查中应用广泛。

(2) 邮寄调查是通过邮寄或宣传媒体等方式,将调查表或调查问卷送到被调查者手中,由被调查者填写后返回到调查者手中的一种方法。其特点是调查人员和被调查者没有直接的语言交谈,信息的传递完全依赖于调查表或调查问卷。这种方法常用于统计部门的统计报表调查,或市场调查机构进行的问卷调查。

(3) 电话调查是调查人员利用电话同受访者进行交流,从而获取信息的一种数据收集的方法。其优点是时效性强,费用低,但用于电话调查的问题要明确,问题量不宜过多。

(4) 电脑辅助调查也称电脑辅助电话调查,是在电话调查时,借助电脑实现电话拨号,呈现调查问卷、答案,记录调查结果,完成数据处理的一种方法。该方法由各种电脑辅助电话调查系统(computer assisted telephone interviewing system,简记为 CATIS)实现,在它所保存的数据库中,不仅包括由访问员直接录入的问卷数据,也包括访问员在访问过程中的录音文件等资料,以备随时对数据的质量进行审核。

(5) 座谈会是将一组被调查者集中在调查现场,让他们就调查的主题发表意见,从中获取数据的方法。这种方法多用于收集与研究课题有关人员的意见和倾向。

(6) 个别深度访问是一种每次只有一名受访者参加的、特殊的定性研究。调查人员要运用大量的提问技巧,尽可能让受访者自由发挥,从受访者表达的想法和感受中收集数据。

2. 观察与实验

观察法是调查者有目的地边观察、边记录被调查对象行为或状态的一种数据收集方法。

例如,为了合理地设置城市交通信号灯,设计者根据安装在道路上的自动计数器,测定每天 24 小时内的交通流量,并将这些流量排序,选择适当的标准,建立合适的模型,来控制信号灯的交替时间。这里自动计数器就起到了观察的作用。

实验法是一种特殊的观察调查方法,是在所设定的特殊实验场所、特殊状态下,对调查对象进行试验,以获得所需资料的一种调查方法。

例如,某药厂研发了一种新药品,为了解该种药品的效果,需要对该药品进行临床实

验,检验其效用。

2.2.3 统计调查的方式

统计数据的获得主要是通过统计调查实现的。统计调查是指调查者按一定的计划,有组织地收集原始数据的方法。统计调查主要包括以下几种方式。

1. 统计报表制度

统计报表制度(statistical report forms)是按照国家有关法规的规定,自上而下统一布置,并按统一的格式、指标、时间和程序实施,再自下而上逐级定期地提供基本统计数据的调查方式,多用于调查国民经济或社会发展的基本统计资料,是我国国家机关和地方政府部门统计数据的主要来源。

统计报表制度能够取得比较完整准确的统计资料,便于连续定期地进行观察。这种调查方式调查范围广,耗费的人力、物力和财力较多,多为国家机关团体、国有企业、事业单位采用。

2. 普查

普查(census)是为某一特定目的而专门组织的全面调查,如人口普查、经济普查或农业普查等。

普查是专门组织的全面调查,可以全面深入地收集到统计数据。由于需要动用大量的人力、物力、财力,耗费大,耗时长,因此适用领域有限,目前多用于重要的国情国力数据的收集。通过在全国范围内组织普查,可以全面了解某项重要的国情,为政府制定相关的方针政策提供依据。

3. 抽样调查

抽样调查(sampling survey)是从总体中随机抽取一部分个体作为研究对象,并根据这一部分个体的调查结果来推断总体特征的一种数据收集方法。

抽样调查是应用最广泛的一种调查方法,具有经济性好(即节约人力、物力、时力和财力),时效性强,适用面广,准确性高等优点。

抽样调查有不同的抽取方法。根据抽取原则的不同,抽样方法主要分为概率抽样和非概率抽样。概率抽样又称随机抽样,即样本的抽取按一定的概率原则,个体被抽中与否不取决于研究人员的主观意愿,而是取决于客观的概率值。非概率抽样也称非随机抽样,此时样本的选取取决于研究人员的意向,样本中个体的抽取不再是随机的。

抽样推断通常建立在随机抽样的基础上,因此重点介绍随机抽样的方法,主要包括简单随机抽样、分层抽样、整群抽样和系统抽样等方法。

(1) 简单随机抽样

如果总体中每个个体被抽中的概率相同,从总体中随机抽取一定数量的个体构成样本,这样得到的样本称为简单随机样本,这种抽样方法称为简单随机抽样。例如,日常生活中的“抽签”就是简单随机抽样。又如,引例中样本的获取可采取简单随机抽样。

简单随机抽样按抽选时每个个体是否允许被重复抽中,分为重复抽样和不重复抽样。

重复抽样是指从总体中抽取容量为 n 的样本,每取出一个个体,记录下其特征后,将其放回总体参加下一次抽选,直至抽满 n 个个体。重复抽样也称放回抽样或还原抽样。