

商务统计学

杨国忠 郑连元 编著

清华大学出版社

北 京

内 容 简 介

本书紧密结合实际研究中出现的新问题,系统介绍商务统计学的基本理论和方法,着重阐述具体应用和案例分析,以培养学生运用商务统计理论和方法分析与解决实际问题的能力。

全书共分为 11 章,讲述了统计的相关概念、统计数据的收集、统计整理、统计数据特征的测度、抽样分布与参数估计、假设检验、方差分析、相关与回归分析、时间序列分析、统计指数、统计决策等。每章有引例、案例分析和习题。附录部分列出了相关统计分布表。

本书编写的基本指导思想是:以经济与管理等相关专业学生为主要读者对象,力求通俗易懂,内容相对全面。理论与实际相结合,在介绍统计理论和方法的基础上,结合具体案例,阐述如何应用统计方法去实际问题。应用 SPSS 统计软件解决实际问题。本书不单独介绍统计软件的使用,而是在各章中运用 SPSS 统计软件进行复杂的计算,使读者掌握软件的应用。各章附有习题,以便读者在学习过程中通过思考和练习,掌握和巩固所学知识,进一步加深对相关内容的理解。

本书适合作为经济管理类本科生、研究生的教材,也可供从事商务管理和经济分析的各类人员参考。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

图书在版编目(CIP)数据

商务统计学/杨国忠,郑连元编著. —北京:清华大学出版社,2019

(21 世纪经济管理精品教材·经济学系列)

ISBN 978-7-302-51259-2

. 商... . 杨... 郑... . 商业统计学-高等学校-教材 . F712.3

中国版本图书馆 CIP 数据核字(2018)第 212064 号

责任编辑:左玉冰

封面设计:李召霞

责任校对:王凤芝

责任印制:宋 林

出版发行:清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址: 北京清华大学学研大厦 A 座 邮 编: 100084

社 总 机: 010-62770175 邮 购: 010-62786544

投稿与读者服务: 010-62776969, c-service@tup.tsinghua.edu.cn

质 量 反 馈: 010-62772015, zhiliang@tup.tsinghua.edu.cn

课 件 下 载: <http://www.tup.com.cn>, 010-62770175 转 4506

印 装 者: 清华大学印刷厂

经 销: 全国新华书店

开 本: 185mm×260mm 印 张: 24.5 字 数: 563 千字

版 次: 2019 年 1 月第 1 版 印 次: 2019 年 1 月第 1 次印刷

定 价: 52.00 元

产品编号: 080858-01

前言

统计学作为一门研究收集、整理和分析统计数据的方法论科学，目前在各学科领域和各行各业都有着非常广泛的应用，成为当代最活跃的学科之一。与此相适应，也出现了很多统计学教材。商务统计学主要是为经济与管理等相关专业学生而编写的。本书紧密结合实际研究中出现的新问题，系统介绍商务统计学的基本理论和方法，着重阐述具体应用和案例分析，以培养学生运用商务统计理论和方法分析与解决实际问题的能力。

本书编写的基本指导思想是：①以经济与管理等相关专业学生为主要读者对象，力求通俗易懂，内容相对全面。②理论与实际相结合，在介绍统计理论和方法的基础上，结合具体案例，阐述如何应用统计方法去解决实际问题。③应用 SPSS 统计软件解决实际问题。本书不单独介绍统计软件的使用，而是在各章中运用 SPSS 统计软件进行复杂的计算，使读者掌握软件的应用。④各章附有习题，以便读者在学习过程中通过思考和练习，掌握和巩固所学知识，进一步加深对相关内容的理解。

本书由中南大学杨国忠和郑连元编著，编写工作分工如下：第 1 章、第 5~8 章、第 11 章以及附录部分由杨国忠负责，第 2~4 章、第 9 章、第 10 章由郑连元负责，全书由杨国忠统稿、整理与审核。

本书适合作为经济管理类本科生、研究生的教材，也可供从事商务管理和经济分析的各类人员参考。

本书在编写过程中参阅了大量国内外相关教材和文献，在此谨向相关作者表示诚挚的谢意。本书获中南大学 2017 年精品教材立项。清华大学出版社的编辑左玉冰老师对本书的出版给予了大力的支持和帮助，在此深表谢意。由于编者水平和掌握资料所限，本书内容必有不足之处，敬请读者批评指正。

编 者

2018 年 5 月

目 录

第 1 章 绪论	1
1.1 统计与统计学	1
1.1.1 统计学的产生与发展	1
1.1.2 统计与统计学的含义	3
1.1.3 统计学的研究对象及特点	4
1.1.4 统计学的分科	5
1.2 统计的几个基本概念	6
1.2.1 总体与总体单位	6
1.2.2 样本	7
1.2.3 指标与标志	8
1.2.4 变异与变量	9
1.3 商业世界中的统计学	10
1.4 统计数据的计量尺度与类型	11
1.4.1 统计数据的计量尺度	11
1.4.2 统计数据的类型	12
1.5 统计研究方法过程	14
1.5.1 统计研究方法	14
1.5.2 统计研究过程	15
1.6 案例：高考志愿	16
1.6.1 案例背景	16
1.6.2 案例分析	17
习题 1	17
第 2 章 统计数据的收集	18
2.1 统计设计	18
2.1.1 统计设计的概念和意义	18
2.1.2 统计设计的种类与内容	19
2.1.3 统计指标和指标体系设计的内容	21
2.1.4 统计指标和指标体系设计的原则	23
2.1.5 统计指标体系的构成	24

2.2	统计调查的方式与方法	25
2.2.1	统计调查的意义与基本要求	25
2.2.2	统计资料的收集方式和方法	27
2.2.3	统计数据的收集方法	37
2.3	统计调查方案与调查问卷的设计	41
2.3.1	统计调查方案的设计	41
2.3.2	统计调查问卷的设计	44
2.4	统计调查误差与控制	54
2.4.1	统计调查误差的概念和种类	54
2.4.2	统计调查误差的产生原因	55
2.4.3	防止与减少统计调查误差的办法	56
2.5	案例：我国人口统计数据的调查与发布	56
2.5.1	案例背景	56
2.5.2	案例分析	57
	习题 2	62
第 3 章 统计整理		63
3.1	统计整理概述	64
3.1.1	统计整理的概念与意义	64
3.1.2	统计整理的程序	65
3.1.3	统计整理的原则	65
3.2	资料预处理	66
3.2.1	资料审核	66
3.2.2	资料筛选	68
3.2.3	资料排序	68
3.3	统计资料的整理	68
3.3.1	统计分组的概念与作用	68
3.3.2	分组标志的种类及选择	70
3.3.3	统计分组参数的确定	72
3.3.4	统计资料汇总	77
3.4	统计整理结果的显示	79
3.4.1	分布数列	79
3.4.2	统计表	83
3.4.3	统计图	87
3.5	案例：变量数列的编制	100
3.5.1	案例背景	100
3.5.2	案例分析	101
	习题 3	103

第 4 章 统计数据特征的测度	106
4.1 总体数量特征的测度	106
4.1.1 总体绝对数量的测度	106
4.1.2 总体相对数量的测度	109
4.2 集中趋势的测度	117
4.2.1 集中趋势的含义	117
4.2.2 算术平均数	118
4.2.3 调和平均数	123
4.2.4 几何平均数	127
4.2.5 众数	128
4.2.6 中位数与四分位数	130
4.2.7 平均指标的比较	134
4.3 离散趋势的测度	135
4.3.1 变异指标	135
4.3.2 极差	136
4.3.3 平均差	137
4.3.4 方差与标准差	139
4.3.5 变异系数	143
4.3.6 应用平均指标、变异指标的注意问题	144
4.4 分布形态的测度	145
4.4.1 偏度	145
4.4.2 峰度	147
4.5 案例：某地区水稻产量	148
4.5.1 案例背景	148
4.5.2 案例分析	148
习题 4	151
第 5 章 抽样分布与参数估计	153
5.1 抽样与抽样分布	153
5.1.1 几个基本概念	153
5.1.2 常用抽样方法	155
5.1.3 抽样分布	161
5.1.4 几种与正态分布有关的分布	163
5.1.5 正态总体的常用统计量分布	165
5.2 参数估计的含义与类型	165
5.2.1 参数估计的含义	165
5.2.2 参数估计的类型	166

5.3 参数的点估计	167
5.3.1 矩估计法	167
5.3.2 特征数法	168
5.3.3 极大似然估计法	169
5.3.4 估计量优劣的衡量标准	172
5.4 参数的区间估计	172
5.4.1 单正态总体均值的区间估计	172
5.4.2 单正态总体方差的区间估计	175
5.4.3 两个正态总体均值差的区间估计	176
5.4.4 两个正态总体方差比的区间估计	178
5.4.5 单个正态总体比例的区间估计	179
5.4.6 两个总体比例之差的区间估计	181
5.4.7 单侧置信区间	182
5.5 样本容量的确定	183
5.5.1 估计总体均值时样本容量的确定	184
5.5.2 估计总体比例时样本容量的确定	185
5.6 案例: 大学生自习时间的差异	185
5.6.1 案例背景	185
5.6.2 案例分析	186
习题 5	188
第 6 章 假设检验	190
6.1 假设检验的基本思想	190
6.1.1 假设检验的概念	190
6.1.2 小概率原理	190
6.1.3 假设检验的推理方法	190
6.1.4 假设检验的基本步骤	191
6.1.5 双侧检验与单侧检验	193
6.1.6 假设检验的两类错误	194
6.2 一个正态总体参数的检验	194
6.2.1 总体均值的检验	194
6.2.2 总体方差的检验	196
6.2.3 总体比例的检验	197
6.3 两个正态总体参数的检验	198
6.3.1 两个总体均值之差的检验	198
6.3.2 两个总体方差之比的检验	199
6.3.3 两个总体比例之差的检验	201
6.4 案例: 广告效果检测	202

6.4.1 案例背景	202
6.4.2 案例分析	203
习题 6	204
第 7 章 方差分析	205
7.1 方差分析基本思想	205
7.1.1 方差分析的基本概念	205
7.1.2 方差分析的基本原理	206
7.1.3 方差分析的基本方法	206
7.2 单因素方差分析	207
7.2.1 单因素方差分析的数据结构	207
7.2.2 单因素方差分析的基本步骤	208
7.2.3 单因素方差分析中的不等重复问题	211
7.3 双因素方差分析	211
7.3.1 双因素方差分析的类型	211
7.3.2 双因素方差分析的数据结构	212
7.3.3 离差平方和的计算	212
7.3.4 应用实例	213
7.4 案例: 员工生产效率的差异	215
7.4.1 案例背景	215
7.4.2 案例分析	215
习题 7	222
第 8 章 相关与回归分析	224
8.1 变量的相关分析	224
8.1.1 相关关系的概念	224
8.1.2 相关关系的种类	225
8.1.3 相关关系的显示	226
8.1.4 相关关系的度量	227
8.2 一元线性回归分析	231
8.2.1 一元线性回归模型	231
8.2.2 一元线性回归模型参数的估计	232
8.2.3 一元线性回归模型的检验	234
8.2.4 一元线性回归模型预测	238
8.3 多元线性回归分析	239
8.3.1 多元线性回归模型的形式	239
8.3.2 多元线性回归模型的估计	239
8.3.3 多元线性回归模型的检验	241

8.3.4 多元线性回归预测	243
8.4 非线性回归分析	244
8.4.1 非线性回归分析的意义	244
8.4.2 非线性函数形式的确定	244
8.4.3 非线性回归模型的估计	246
8.5 案例: 国内生产总值与全社会固定资产投资及价格指数关系	247
8.5.1 案例背景	247
8.5.2 案例分析	248
习题 8	249
第 9 章 时间序列分析	252
9.1 时间序列的基本问题	252
9.1.1 时间序列的概念与意义	252
9.1.2 时间序列的种类	253
9.1.3 时间序列的编制原则	255
9.2 时间序列的水平指标	256
9.2.1 发展水平	256
9.2.2 平均发展水平	257
9.2.3 增长量	263
9.2.4 平均增长量	264
9.3 时间序列的速度指标	266
9.3.1 发展速度	266
9.3.2 增长速度	267
9.3.3 平均增长速度	269
9.3.4 平均发展速度	269
9.4 时间序列趋势分析	273
9.4.1 时间序列趋势分析的原理	273
9.4.2 长期趋势分析	276
9.4.3 季节变动分析	283
9.4.4 循环变动与不规则变动分析	287
9.5 案例: 我国人口增长趋势	289
9.5.1 案例背景	289
9.5.2 案例分析	290
习题 9	295
第 10 章 统计指数	299
10.1 统计指数概述	299
10.1.1 统计指数的概念与作用	299

10.1.2	统计指数的性质	301
10.1.3	统计指数的分类	302
10.2	统计指数的编制	303
10.2.1	指数编制的基本问题	303
10.2.2	综合指数的编制	308
10.2.3	平均指数的编制	311
10.2.4	指数体系	315
10.2.5	指数数列	319
10.3	因素分析法	322
10.3.1	因素分析法的概念	322
10.3.2	因素分析的步骤	322
10.3.3	总量指标因素分析	323
10.3.4	相对指标因素分析	326
10.3.5	平均指标变动的因素分析	327
10.4	常用统计指数	330
10.4.1	居民消费价格指数	330
10.4.2	工业品出厂价格指数	332
10.4.3	股票价格指数	334
10.4.4	空间指数	339
10.4.5	工业生产指数	340
10.5	案例：辉腾公司效益影响因素	342
10.5.1	案例背景	342
10.5.2	案例分析	342
习题 10		345
第 11 章	统计决策	348
11.1	统计决策概述	348
11.1.1	决策的概念	348
11.1.2	决策的类型	349
11.1.3	统计决策的概念	352
11.2	决策树方法	352
11.2.1	决策树的构成与决策方法	352
11.2.2	决策树方法的应用	353
11.3	贝叶斯决策方法	356
11.3.1	贝叶斯公式	356
11.3.2	贝叶斯决策的概念和步骤	356
11.3.3	贝叶斯决策方法的应用	357
11.3.4	贝叶斯决策的利弊	359

11.4 效用概率决策方法	360
11.4.1 效用概率决策的概念	360
11.4.2 效用决策的准则	361
11.4.3 效用决策的应用	361
11.5 马尔可夫决策方法	361
11.5.1 基本概念和性质	361
11.5.2 马尔可夫决策方法的应用	362
11.6 案例: H 公司生产产品的选择	364
11.6.1 案例背景	364
11.6.2 案例分析	364
习题 11	365
附录 常用统计表	367
参考文献	377

第 1 章

绪 论

引 例

2 元钱的经济纠纷^①

2017 年 9 月 26 日下午, 宁波海曙江厦派出所的廖警官, 出了一次警——一笔“涉案金额”为 2 元的小纠纷。根据当事双方的要求, 他很快处理好了这次纠纷。回来后, 他细想下来, 总觉得哪里不对劲: 似乎这道“数学题”不是这么算的!

这是怎么回事? 到底是什么数学题, 这么费脑筋呢?

报警电话是一名小卖部的老板打来的, 说是“有人买东西不给钱”。廖警官到场后, 了解了情况, 大概是这么回事: 有名小朋友在小卖部里买了一支雪糕, 花了 2 元钱。后来, 小朋友的外婆觉得雪糕有点儿变形, 怀疑是变质的, 于是到店里来要求退钱。小卖部老板当然不同意退。这时, 一名女孩过来, 买了一瓶水, 也是 2 元。外婆就抢过她递给店老板的 5 元钱, 然后找了对方 3 元钱。女孩也没说什么, 拿了水就走了。对小朋友的外婆来说, 这笔账似乎就这么了了。可在店老板看来, 他没收到矿泉水的钱啊。两人就为这 2 元钱起了纠纷, 于是报了警。

“那你们双方什么意见?” 廖警官问道。“她把那 5 元钱还给我!” “那他得把买雪糕的 2 元钱退给我!” 双方争先恐后地说。民警在场, 两人倒是很快协调好了, 各自拿了钱。当事双方没有多想, 对这一协调结果表示满意。于是, 大家都散了。但廖警官在返回所里的路上隐隐觉得: 这道数学题不是这么算的吧? 回去后, 他特意把每次交易行为都一一列了下来, 得出了结论——有人亏钱了。

1.1 统计与统计学

1.1.1 统计学的产生与发展

统计作为一种人类认识自然、改造自然的实践活动, 在原始社会就已经产生。统计学是伴随着资本主义古典哲学、古典政治经济学和空想社会主义的产生而发展起来的。一般认为, 统计学从 17 世纪 60 年代算起, 到现在已有近 400 年的历史。统计学的发展

^① 案例来源于《宁波晚报》。

过程大体可分为以下四个阶段。

第一个阶段称为“城邦政情”阶段。“城邦政情”阶段始于古希腊的亚里士多德写的“城邦政情”或“城邦纪要”。他一共撰写了 150 余条纪要，其内容包括城邦的历史、行政、科学、艺术、人口、资源和财富等社会及经济情况的比较与分析，具有社会科学特征。“城邦政情”式的统计研究延续了一两千年，直到 17 世纪中叶才逐渐被“政治算术”这个名词所替代，并且很快被演化为“统计学”(statistics)。统计学依然保留了“城邦”(state)这个词根。

第二个阶段称为“政治算术”阶段。“政治算术”阶段与“城邦政情”阶段并没有明显的分界点，两者也没有根本的区别。“政治算术”的特点是统计方法与数学计算和推理方法相结合，分析社会经济问题的方式更加注重运用定量分析方法。可以用 1690 年英国古典政治经济学的奠基人威廉·配第(William Petty, 1623—1687)出版《政治算术》一书作为这个阶段的起始标志。威廉·配第在书中强调了自己的立论的方法，“用数字、重量和尺度来表达自己的问题”。用数字、重量和尺度量化社会经济现象的方法是近代统计学的主要特征。因此，威廉·配第的《政治算术》被后来的学者评价为近代统计学的来源，威廉·配第本人也被评价为“近代统计学之父”。构成“政治算术”方法的核心是“数字”。配第在书中使用的数字有三类：第一类是对社会经济现象通过统计调查和经验观察得到的数字，第二类是运用某种数学方法推算出来的数字，第三类是为了进行理论性推理而采用的示例性数字。配第把这种运算数字和符号进行推理称为“代数的算法”。

另外，法国的帕斯卡(B.Pascal, 1623—1662)和皮埃尔·德·费马(Pierre de Fermat, 1601—1665)将赌博中出现的各种具体问题归结为一般的概率原理，为概率论和统计学的发展奠定了重要的基础。特别值得提出的是比利时统计学家凯特勒(Quetelet L. A. J., 1796—1874)，他对统计学理论最大的影响是把概率与统计学相结合，从而提出了关于统计学的新概念，并努力使人们把统计学理解为一种可用于任何学科的一般研究方法。在“政治算术”阶段出现的统计学与数学的结合趋势逐渐发展形成了“统计分析科学”。

第三个阶段称为“统计分析科学”阶段。19 世纪末，欧洲大学里开设的“国情纪要”或“政治算术”等课程名称逐渐消失，取而代之的是“统计分析科学”课程。当时的“统计分析科学课程”的内容仍然是分析研究社会经济问题。“统计分析科学”课程的出现是现代统计学发展阶段的开端。

经过历代统计学家的努力和发展，到 19 世纪末建成了古典统计学的基本框架。1908 年，英国的威廉·希利·戈塞特(William Seely Gosset, 1876—1937)用笔名 Student 发表了关于 t 分布的论文，这是一篇在统计发展史上划时代的文章。他创立了小样本代替大样本的方法，开创了统计学的新纪元。后来统计学家罗纳德·艾尔默·费希尔(Ronald Aylmer Fisher, 1890—1962)给出了 F 统计量、极大似然估计、方差分析等；内曼(J. Neyman, 1894—1981)和皮尔逊(Egon S. Pearson, 1895—1980)提出了置信区间估计和假设检验；沃尔德(A. Wald, 1902—1950)发明了序贯抽样和统计决策函数；到 20 世纪 50 年代，构筑了现代统计学的基本框架。20 世纪 50 年代后，统计学进入一个全新的发展阶段，新的研究领域层出不穷，像多元统计分析、非参数统计、现代时间序列

分析和线性统计模型等。

第四个阶段称为“统计学帝国主义”阶段。从世界范围看,自20世纪60年代以后,统计学的发展有三个明显的趋势:第一,随着数学的发展,统计学依赖和吸收数学方法越来越多;第二,向其他学科领域渗透,或者说以统计学为基础的边缘学科不断形成;第三,随着统计学应用的日益广泛和深入,以及受计算机和新兴科学的影响,统计学越来越依赖计算技术,成为数量分析的方法论科学。这一时期统计学的研究和运用范围越来越广。美国的《Science》有一篇文章列出近百年来最有用的科学,统计学位居前10名。

1.1.2 统计与统计学的含义

“统计”一词,在不同的场合,所代表的含义也有所不同。一般来讲,“统计”有三种含义:统计工作、统计资料和统计学。

1. 统计工作

统计工作是指对社会现象或自然现象的总体数量方面进行收集、整理和分析的实践活动,它是一种社会调查研究活动。统计工作在人类历史上出现得比较早。随着历史的发展,统计工作逐渐发展和完善起来,并成为国家、企业和个人及科研单位认识与改造客观世界和主观世界的一种有力工具。统计工作一般包括统计设计、统计调查、统计整理、统计分析、统计资料的提供和管理等环节。

2. 统计资料

统计资料是指可以用以推导出某项结论的一些事实或数字。统计资料一般由元素、变量和观测值三部分组成。

(1) 元素。我们所研究的对象由各元素组成,统计资料就是关于各元素特征的信息。如统计某公司员工基本情况时,各员工就是元素。

(2) 变量。变量是关于元素的一种属性和特征。年龄是各员工的一个特征,这些特征对于不同元素取不同的结果,所以称为变量。其中结果用数字表示的,称为定量变量;不可用数字表示的,称为定性变量。例如,年龄、身高、体重、公司服务年限和受教育年限是定量变量,性别、民族、籍贯、学历是定性变量。

(3) 观测值。观测值是指通过测量或测定所得到的样本值。如统计每小时通过一个高速公路收费站的汽车数量就是一个观测值。

3. 统计学

统计学是以收集、整理、分析和研究等统计技术为手段,对所研究对象的总体数量关系和数据资料去伪存真、去粗取精,从而达到显示、描述和推断所研究对象的特征、趋势和规律性的目的。简言之,统计学是以少量的数据(称为样本)所提供的信息来推断欲研究对象(称为总体)特征的一门科学。

统计工作、统计资料和统计学的相互关系如图 1-1 所示。

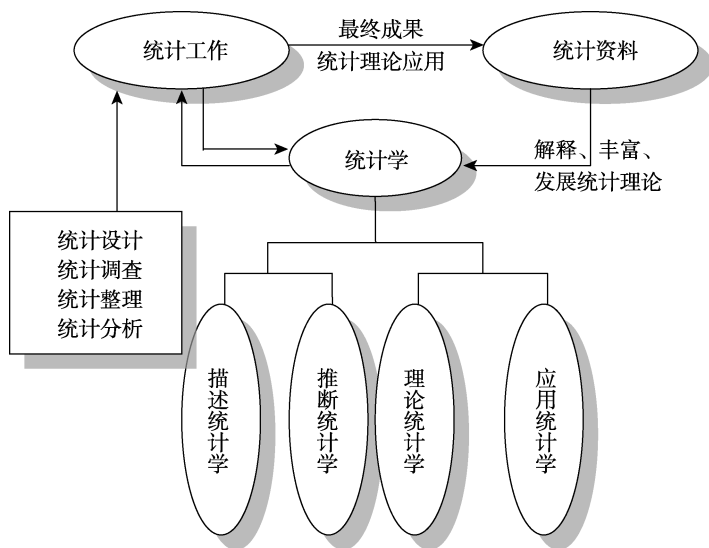


图 1-1 统计工作、统计资料和统计学的相互关系

1.1.3 统计学的研究对象及特点

统计学的研究对象是客观事物的总体数量特征、数量关系和数据资料，以反映其发展过程及规律性。统计学的特点可以归纳为以下五个方面。

1. 数量性

统计学的特点首先表现为以准确的事实为基础；同时，这些事实用数字加以表现，具有简短性和明显性。数量性是统计学研究对象的重要特点，这一特点也可把其他实质性的社会科学（如政治经济学）区别开来，如各种统计年鉴就是用大量数字资料说明事物的规模、水平、结构、比例关系、差别程度、普遍程度、发展速度等。应当注意，统计学不是单纯地研究社会现象的数量方面，而是在质与量的密切联系中研究社会现象的数量方面。

2. 总体性

统计学研究社会现象的数量方面指的是总体的数量方面。从总体上研究社会现象的数量方面，是统计学区别于其他社会科学的一个主要特点。社会现象是各种社会规律相互交错作用的结果，它呈现出一种复杂多变的情景。统计学对社会现象总体数量方面的调查研究，用的是综合研究方法，而不是对单个事物的研究，但其研究过程是从个体到总体，即必须对足够大量的个体进行登记、整理和综合，使它过渡到总体的数量方面，从而把握社会现象的总体规模、总水平及其变化发展的总趋势。

3. 具体性

统计研究对象是自然和社会经济领域中具体现象的数量方面。它不是单纯数量的研究,而是具有明确的现实含义,这一特点是统计学与数学的分水岭。数学是研究事物的抽象空间和抽象数量的科学,而统计学研究的数量是客观存在的、具体实在的数量表现。正因为统计的数量是客观存在的和具体实在的数量表现,它才能独立于客观世界,不以人们的主观意志为转移。统计资料作为主观对客观的反映,必然是存在第一性,意识第二性,存在决定意识,如实反映具体的已经发生的客观事实,才能为我们进行统计分析研究提供可靠的基础,才能分析、探索和掌握事物的统计的规律性。统计研究对象的这一特点,也正是统计工作必须遵循的基本原则。

4. 社会性

统计学研究社会现象,这一特点与自然技术统计学有所区别。自然技术统计学研究自然技术现象(如天文、物理、生物、水文等现象),自然现象的变化发展有其固有的规律,在其变化过程中,通常表现为随机现象可能出现也可能不出现的现象。而统计学的研究是人类社会活动的过程和结果,人类的社会活动都是人们有意识、有目的的活动,各种活动都贯穿着人与人之间的关系,除了随机现象而外,还存在确定性的现象,即必然要出现的现象。所以统计学在研究社会现象时,还必须注意正确处理好这些涉及人与人之间关系的社会矛盾。

5. 广泛性

统计学研究的数量方面非常广泛,指全部社会现象的数量方面。广泛性这一特点,可区别研究某一特定领域的其他社会科学(如政治学、经济学、社会学、法学)。统计学研究的领域包括整个社会,它既研究生产关系,也研究生产力以及生产关系和生产力之间的关系;它既研究经济基础,也研究上层建筑以及经济基础和上层建筑之间的关系。此外,它还研究生产、流通、分配、消费等社会再生产的全过程以及社会、政治、法律、军事、文化、教育等全部社会现象的数量方面。

1.1.4 统计学的分科

从统计方法的构成看,统计学可分为描述统计学和推断统计学;从统计方法研究和应用角度看,统计学可分为理论统计学和应用统计学。

1. 描述统计学和推断统计学

描述统计学,研究任何反映客观现象的数据并通过图表形式对所收集的数据进行处理,进而通过综合、概括与分析得出反映客观现象的规律性的数量特征的方法。内容包括统计数据的收集方法、数据的加工整理方法、数据分布特征的概括与分析方法等。它是整个统计学的基础。

推断统计学,研究根据样本数据去推断总体数量特征的方法,它是在对样本数据进

行描述的基础上,对统计总体的未知数量特征作出以概率形式表述的推断。推断统计学是现代统计学的主要内容。

描述统计学和推断统计学的划分,既反映了统计方法发展的前后两个阶段,也反映了统计方法研究和探索客观现象内在数量规律性的先后两个过程。从描述统计学发展到推断统计学,既反映了统计学发展的巨大成就,也是统计学发展成熟的重要标志。

2. 理论统计学和应用统计学

理论统计学主要是阐明统计学的一般原理和一般方法,主要内容包括概率论、随机化原则理论、各种估计原理、假设检验原理和一般预测决策原理,其侧重点在于统计学原理的推导与证明。

应用统计学主要研究如何应用统计方法去解决实际问题。统计方法的应用几乎扩展到所有的科学研究领域。统计方法在经济领域的应用就形成了经济统计学及其若干分支:在管理领域的应用就形成了管理统计学,在商务领域的应用就形成了商务统计学,在生物学中的应用就形成了生物统计学,在医学中的应用就形成了医疗卫生统计学,等等。这些应用统计学的不同分支,其所应用的基本统计方法是相同的,即都是描述统计和推断统计的主要方法,只是由于各应用领域的特殊性形成了各分支间的一些不同的特点。

1.2 统计的几个基本概念

1.2.1 总体与总体单位

总体是在某种共性的基础上由许多个别事物结合起来的整体。构成总体的个别事物称为总体单位,总体单位也称个体。总体和总体单位都是客观存在的事物,是统计研究的客体。例如,在工业普查中,“工业企业”就是一个总体,它是由所有从事工业生产活动的企业所组成的,其中的每一个工业企业就是一个总体单位。总体必须同时具备三个性质:同质性、大量性和差异性。

1. 同质性

同质性是形成统计总体的必要条件,是统计总体各个单位在某一点上具有共同的属性和特征,否则就不能形成一个统计总体。同质性是确定统计总体的基本标准,它是根据统计的研究目的而定的。研究目的不同,则所确定的总体不同,其同质性的意义也随之变化。例如,研究城镇居民的生活状况,城镇居民就构成了统计总体,凡是城镇居民都是同质的。

2. 大量性

大量性是指总体由许多单位组成,仅个别和少数单位不能构成总体。这一点是由统计研究的目的所决定的。统计研究的目的是要揭示自然和社会经济现象的规律性,我们只有通过大量事物的观察、分析和研究,才能发现从其普遍联系中表现出来的规律性。

例如,要研究某市职工的工资水平,我们只观察少数几个职工的工资是达不到目的的,因为这少数几个职工的工资不能代表该市全体职工的工资水平,它可能偏高或偏低。如果我们观察许多职工的工资,就可降低或抵消偶然性的偏差,计算他们的平均工资,就可以反映该市职工工资的一般水平。每个职工工资差别越大,需要观察的职工人数就越多。如果要提高观察职工工资的代表性,就需要增加观察的职工人数。可见,统计总体的大量性是一个相对概念。总体的大量性与各单位标志表现的差异性密切联系在一起。同时,总体的大量性也与研究目的的要求有关,精确度要求越高,总体单位数也越大。

3. 差异性

差异性是指构成统计总体的各个总体单位除了某一方面是同质的以外,在其他方面又要有差异,否则就没有必要进行统计分析了。以某企业职工总体为例,这里每位职工除了工作单位相同外,在姓名、性别、年龄、学历、职称和工龄等许多方面都存在差异。因此,事物存在差异性在进行统计研究的前提。

总体可分为有限总体和无限总体。如果一个总体中包含的总体单位数是有限的,称其为有限总体;如果一个总体中包含的总体单位数是无限的,称其为无限总体。例如,我们要研究海洋中的鱼类、某地区的空气污染情况,则海洋中的鱼类、某地区的空气污染情况就是无限总体。但社会经济现象总体大多是有限的,如全国工业企业总体、某企业职工总体、某商场库存商品总体以及某地交通事故总体等。

总体和总体单位具有相对性,它们随着研究目的不同是可以变换的。例如,如果要研究某地区工业企业的生产经营情况,则该地区全部工业企业构成总体,而每一个工业企业是总体单位;如果要研究该地区某一个企业的生产经营情况,那么该企业就成了总体,该企业下属的各个职能部门就是总体单位。由此可见,一个工业企业由于研究目的不同,既可以作为一个总体单位来研究,也可以作为一个总体来研究。

1.2.2 样本

样本是指从统计总体中抽取出来作为代表这一总体的部分单位组成的集合体。与样本相对应的统计总体称为全及总体。由于样本是从总体中抽取出来并代表总体的,全及总体也可称为母体,而样本则称为子体。样本也有自己的几个特点。

1. 代表性

抽取样本的目的是用来推断总体,这就必然要求样本能够代表总体。样本代表总体的程度越高,样本计算的抽样指标与总体指标的误差就越小。因此,抽样推断时,总是要求样本具有较高的代表性。

2. 客观性

从总体中抽取样本时,必须排除主观因素的影响。抽取样本可以用随机抽样的方法抽取,也可以用非随机抽样的方法抽取,但都必须保证取样的客观性,保证样本的选择不受调查者和被调查者的主观影响。

3. 随机性

一个总体可以抽取许多不同的样本，至于到底抽取的样本是哪一个，完全取决于样本的随机性。随机性是用样本数据资料推断总体特征性质的基本原则。

4. 排他性

样本单位必须取自总体内部，而不能抽取总体外部的单位。因为统计推断是利用样本作为总体代表，用样本的数据资料来推断总体的数据资料。

1.2.3 指标与标志

1. 指标

指标也称统计指标，其含义一般有两种理解：一种是指反映总体现象数量特征的概念。例如，国内生产总值、国民生产总值、商品销售额、人口出生率等。这种理解的统计指标包括三个构成要素：指标名称、计算方法及计量单位。这种含义一般在进行统计设计或理论研究时所使用，它是仅有数量概念而没有具体数字的统计指标。另一种是由反映总体现象数量特征的概念（指标名称）和具体数值构成的统计指标。例如，《湖南统计年鉴 2017》显示，2016 年长沙市支出法地区生产总值 9 356.91 亿元。从完整的意义上讲，指标由六个要素构成：时间限制、空间限制、指标名称、指标数值、计量单位和计算方法。这种含义的统计指标是统计工作中经常使用的。上述例子对应的六要素依次是：时间限制是 2016 年、空间限制是长沙市、指标名称是地区生产总值、指标数值是 9 356.91、计量单位是亿元、计算方法是支出法。

指标通常分为数量指标和质量指标。数量指标是反映现象总规模、总水平和工作总量的统计指标，如人口总数、企业总数、工资总额、国内生产总值等。由于数量指标反映现象的总量，所以也称总量指标，并且由于用绝对数表示，也称统计绝对数。质量指标是反映现象相对水平或工作质量的统计指标，如人口密度、出生率、出勤率、劳动生产率、职工平均工资等。质量指标通常是由两个总量指标相比而派生的指标，用相对指标或平均指标来表示，反映现象之间的内在联系和对比关系。

2. 标志

标志是说明总体单位属性和特征的名称。例如，某企业全体职工作为一个总体，每一位职工是总体单位，职工的性别、年龄、籍贯、民族等是说明每一位职工的特征的名称，都是标志。

标志按其性质不同可分为品质标志和数量标志。品质标志是表明总体单位品质属性或特征的名称，它不能用数值表示，只能用文字说明。例如，企业职工性别、籍贯、民族、文化程度就是品质标志。数量标志是表明总体单位数量方面的特征，它用各种不同的数值表示。例如，企业职工的年龄、工龄就是数量标志。数量标志的具体表现为标

标志。例如，A公司某职工年龄为28岁，工资为6500元/月。

3. 指标和标志的区别与联系

指标和标志，两者既有区别又有联系。指标和标志的区别有以下四点。

第一，标志是说明总体单位特征的，而指标是说明总体特征的。

第二，标志有能用数值表示的数量标志和不能用数值表示的品质标志，而指标不论是数量指标还是质量指标，都是用数值表示的。

第三，指标数值是经过一定的汇总取得的，而标志中的数量标志不一定经过汇总，可直接取得。

第四，标志一般不具备时间、地点等条件，但作为一个完整的统计指标，一定要包含时间、地点、范围等信息。

指标和标志的联系有以下两点。

第一，统计指标的数值是从总体单位数量标志的标志值进行直接汇总或间接计算而来的。例如，某企业职工的月工资总额是该企业所属职工月工资额汇总而来的，而职工的月平均工资则是通过进一步计算得到的。

第二，指标与数量标志之间存在着变换关系。由于研究目的的不同，原来的统计总体变成了总体单位，则相应的统计指标也就变成数量标志了；反之亦然。例如，在研究某公司职工情况时，该公司的全部职工是总体，该公司的工资总额是统计指标。而在研究该公司所属的集团公司职工工资时，该公司就是总体单位，则该公司的工资总额为数量标志，具体的工资总额数值为标志值。于是，该公司的工资总额由统计指标相应变为数量标志了。

1.2.4 变异与变量

统计中的变异是普遍存在的，一般意义上的变异是指标志在总体单位之间的不同具体表现，但严格地说，变异仅指品质标志的不同具体表现，如性别表现为男、女，民族表现为汉族、满族、蒙古族、回族、苗族等。而数量标志的不同具体表现则称为变量值或标志值，如某职工的年龄为35岁、工龄为14年、月工资为8500元等。可变的数量标志称为变量。品质标志的变异最后表现为综合性的数量时，如按职工的性别，汇总计算出男、女各多少人，才构成统计研究的对象。观察登记总体各单位的品质标志和数量标志的变异与变量，是统计研究的起点。

变量按其取值是否连续，可分为离散变量和连续变量。凡变量值只能以整数出现的变量，称为离散变量；凡变量值可作无限分割的变量，称为连续变量。例如，一个地区的人口数、工厂数、一个工厂的机器台数等，都只能是整数而不可能带小数，这些就是离散变量。但人的身高、体重、工厂的产值、利润等却是可以带小数而且其数值是可以无限分割的，在两个数字之间还可以有连续不断变化的其他数字，这种变量就是连续变量。

变量按其所受因素影响的不同，可分为确定性变量和随机性变量。确定性变量是具

有某种（或某些）起决定性作用的因素致使其沿着一定的方向呈上升、下降或水平变动的变量。例如，随着人们生活水平的提高以及医疗卫生条件的完善，这些确定性因素的影响，使得人们的期望寿命这个变量的变量值不断提高。期望寿命就是确定性变量。随机性变量是指变量值的变化受不确定性因素的影响，变量值的变化没有一个确定的方向，有很大的偶然性。例如，在同一台机器设备上加工某种机械零件，其尺寸大小总是存在差异的。造成这些差异的因素可能有原材料质量、电压的不稳定，气温和环境的变化以及操作工人的情绪波动等。这些影响该种机械零件尺寸变动的因素都是随机发生的，是不确定的。这里的机械零件尺寸就是一个随机性变量。

1.3 商业世界中的统计学

统计学作为一门基础学科，其方法适用于所有学科领域。统计分析方法是一种通用数据分析方法。可以说，只要有数据的地方就会用到统计方法。随着人们对定量研究的日益重视，统计方法已被应用到自然科学和社会科学的众多领域。几乎所有的研究领域都要用到统计方法，如政府部门、学术研究领域、日常生活、公司或企业的生产经营管理等。下面简要介绍统计在商业中的一些应用。

1. 企业发展战略

发展战略的重要性对一个企业而言是不言而喻的。发展战略的制定需要把握整个宏观经济和市场的状况及发展变化趋势。同时，还要对本企业进行合理的市场定位，以把握企业自身的优势和劣势。所有这些都离不开统计。它不仅需要统计提供可靠的数据，还需要利用统计方法对数据进行科学的分析和预测。

2. 市场研究

市场竞争是激烈的。企业要在激烈的市场竞争中获胜，首先必须了解市场。为此，企业需要作广泛的市场调查，取得所需的信息，并对这些信息进行科学的分析，以便作为生产和营销的依据，这些都需要统计的支持。

3. 产品质量管理

质量是企业的生命，是企业持续发展的基础。质量管理中离不开统计的应用。质量控制已成为统计学在生产领域中的一项重要应用。各种统计质量控制图被广泛应用于监测生产过程。当今 6 σ 准则已成为一种重要的产品质量管理理念。

4. 财务分析

企业的财务分析不可避免地使用统计分析方法。一方面，企业自身的投资，离不开对财务数据的分析，其中要用到大量的统计方法；另一方面，投资者往往根据上市公司提供的财务报表和相关的统计数据进行分析，以此作为投资决策的参考依据。

5. 经济预测

企业经营者必须对未来的市场状况进行预测。在进行预测时要使用各种统计信息和统计方法。例如,企业要对产品的市场潜力作出预测,以便及时调整生产计划,这就需要利用市场调查取得数据,并对数据进行统计分析。

6. 人力资源管理

企业人力资源管理人员利用统计方法对企业员工的年龄、性别、受教育程度、工作业绩、工资等进行分析,并作为企业制订工资计划、奖惩制度的依据。

当然,统计并不是仅仅为了管理才有用,它是为自然科学、社会科学的多个领域而发展起来的,为多个学科提供了一种通用的数据分析方法。从某种意义上说,统计仅仅是一种数据分析的方法。与数学一样,统计是一种工具,它是一种数据分析的工具。利用统计方法可以简化繁杂的数据,如用图表展示数据,建立数据模型。有人认为统计的全部目的就是让人看懂数据,其实这仅仅是统计的一个方面。统计更重要的功能是对数据进行分析。它提供了一套分析数据的方法和工具。统计不是万能的,它不能解决你所面临的所有问题。统计可以帮助分析数据,并从分析中得出某种结论,但对统计结论的进一步解释,则需要你的专业知识。例如,吸烟会使患肺癌的概率增大,这是一个统计结论,但要解释吸烟为什么能引起肺癌,这就不是统计学家所能解释的,而需要有更多的医学知识才行。

商务统计学应用的广泛性,在于它的分析方法的有效性,其中就包括统计模型。统计模型用统计学方法描述经济活动。根据所采用的数学方法不同,对经济活动揭示的程度不同,构成各类不同的统计模型。主要有相关回归分析模型、时间序列分析模型以及统计决策模型等,相关知识我们将在后续章节相应介绍。

1.4 统计数据的计量尺度与类型

1.4.1 统计数据的计量尺度

要对客观对象进行计量,就要弄清楚统计数据的计量尺度问题。根据对研究对象计量的不同精确程度,人们将计量尺度由粗略到精确分为四个层次:定类尺度、定序尺度、定距尺度和定比尺度。

1. 定类尺度

定类尺度是最粗略、计量层次最低的计量尺度。它是按照客观现象的某种属性对其进行分类,不能进行加、减、乘、除等数学运算。这一场合所使用的数值只是作为各种分类的代码,并不反映各类的优劣、量的大小或顺序。例如,人口按性别分为男、女,用“1”表示男性,“0”表示女性。

2. 定序尺度

定序尺度也称顺序尺度,是对现象之间的等级或顺序差别的一种度量。利用定序尺度不仅可以研究对象分为不同的类别,还可以反映各类的优劣、量的大小或顺序,但不能进行加、减、乘、除等数学运算。例如,学生成绩可分为优、良、中、及格和不及格五类,产品质量可分为一、二、三等,一个人对某件事情的态度可以分为非常赞同、赞同、不赞同、坚决反对等。

3. 定距尺度

定距尺度是对现象类别或顺序之间间距的度量。定距尺度不仅可以用数表示现象类别的不同和顺序大小的差异,而且可以用确切的数值反映现象之间在量方面的差异。定距尺度的计算结果表现为数值,可以进行加、减等数学运算。反映现象规模水平的数据必须以定距尺度计量,如产品产量、人口数、企业数、国内生产总值等都以定距尺度来计量。定距尺度在统计数据中占据重要地位,统计中的总量指标也是用定距尺度来计量的。

4. 定比尺度

定比尺度是在定距尺度的基础上,确定相应的比较基数,然后将两种相关的数加以对比而形成相对数(或平均数),用于反映现象的结构、比重、密度、速度等数量关系。定比尺度的计量结果也是数值,可以进行加、减、乘、除等数学运算。例如,将一个企业创造的增加值与该企业的职工人数对比,计算全员劳动生产率,以此反映该企业的生产效率。在统计的对比分析中,要广泛地运用定比尺度进行计量。

1.4.2 统计数据的类型

统计数据是对现象进行计量的结果。例如,对经济活动总量的计量可以得到国内生产总值的数据,对股票价格变动水平的计量可以得到股票价格指数,对人口性别的计量可以得到男或女这样的数据,等等。由此可见,统计数据不仅仅是数字,也可以是文字的。

按照所采用的计量尺度不同,可以将统计数据分为分类数据、顺序数据和数值型数据。统计数据还可以从其他角度进行分类,如按照统计数据的收集方法,可以将其分为观测数据和实验数据;按照被描述的对象与时间的关系,可以将统计数据分为截面数据和时间序列数据。

1. 分类数据、顺序数据和数值型数据

(1) 分类数据。只能归于某一类别的非数字型数据,称为分类数据。分类数据是对事物进行分类的结果,数据则表现为类别,是用文字来表述的。例如,人口按照性别分为男、女两类,学位按照等级分为博士、硕士和学士三类。

在分类数据中,各类别之间是平等的并列关系,无法区分优劣或大小,各类别之间的顺序是可以任意改变的。它仅具有等于或不等于的数学特性。虽然分类数据只是表现为某种类别,但为了便于统计处理,特别是为了便于计算机识别,我们可以对不同类别用不同的数字或编码来表示,如用“1”表示男性人口,“0”表示女性人口;用“1”表示国有企业,“2”表示集体企业,“3”表示私营企业;等等。这些数字只是给不同类别的一个代码,并不意味着这些数字可以区分大小或进行任何数学运算。

(2) 顺序数据。只能归于某一有序类别的非数字型数据,称为顺序数据。顺序数据也是对事物进行分类的结果,但这些类别是有顺序的。例如,学生考试成绩可以分为优、良、中、及格、不及格,一个人的受教育水平可以分为小学、初中、高中、大学及以上;一个人对某一事物的态度可以分为非常同意、同意、保持中立、不同意、非常不同意,等等。

顺序尺度的计量结果虽然也表现为类别,但这些类别之间是可以比较顺序的。很显然,顺序数据要比分类数据精确一些,它除了具有等于或不等于的数学特性外,还具有大于或小于的数学特性。

(3) 数值型数据。按数字尺度测量的观察值,称为数值型数据。数值型数据是使用自然或度量衡单位对事物进行计量的结果,其结果表现为具体的数值。例如,考试成绩用百分制度量,重量用克度量,长度用米度量,等等,其结果都表现为具体的数值。因此数值型数据可以进行加、减、乘、除运算。

分类数据和顺序数据说明的是事物的品质特征,通常是用文字来表述的,其结果均表现为类别,因而也可统称为定性数据或品质数据;数值型数据说明的是现象的数量特征,通常是用数值来表现的,因此也可称为定量数据或数量数据。我们所处理的数据大多为数值型数据。

2. 观测数据和实验数据

观测数据是通过调查或观测收集到的数据,这类数据是在没有对事物人为控制的条件下而得到的,有关社会经济现象的统计数据几乎都是观测数据。实验数据是在实验中控制实验对象而收集到的数据。例如,对一种新药疗效的实验,对一种新的农作物品种的实验等,自然科学领域的大多数数据都为实验数据。

3. 截面数据和时间序列数据

不同对象在同一时期或时点发生的数据,称为截面数据。截面数据所描述的是现象在某一时刻的变化情况。例如,2017年我国34个省市的国内生产总值数据就是截面数据。

同一对象在不同时期或时点发生的数据,称为时间序列数据。时间序列数据所描述的是现象随时间而变化的情况,如2001—2017年我国的国内生产总值数据就是时间序列数据。

下面我们给出统计数据分类的框图,如图1-2所示。

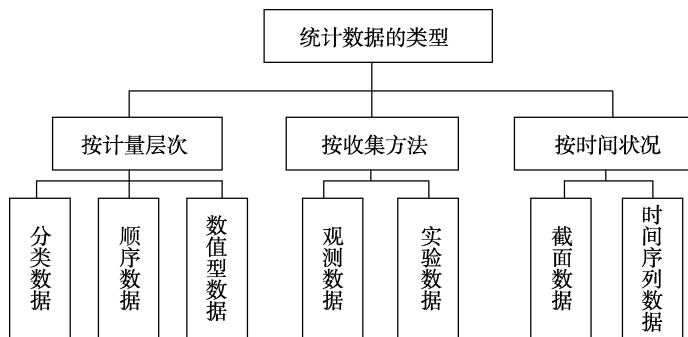


图 1-2 统计数据的分类

区分数据的类型是十分重要的，因为对不同类型的数据，我们将采用不同的统计方法来处理和分析。适用于低层次测量数据的统计方法，也适用于较高层次的测量数据，因为后者具有前者的数学特性。例如，在描述数据的集中趋势时，对分类数据，我们通常是计算众数，对顺序数据通常是计算中位数，但对数值型数据同样也可以计算众数和中位数；反之，适用于高层次测量数据的统计方法，则不能用于较低层次的测量数据，因为低层次测量数据不具有高层次测量数据的数学特性。例如，对于数值型数据我们可以计算平均数，但对于分类数据和顺序数据则不能计算平均数。理解这一点，对于选择统计分析方法是十分有用的。值得注意的是，对于 SPSS 等统计软件，必须定义变量的类型，否则无法进行统计运算。

1.5 统计研究方法过程

1.5.1 统计研究方法

统计学研究客观现象总体的数量特征与规律性所涉及的方法多种多样，这些统计研究方法构成了统计研究方法体系。在统计研究方法体系中，最重要、最基本的研究方法有以下几种。

1. 大量观察法

大量观察法是根据总体中足够多数的单位进行统计研究的方法。由于在研究的总体中，总体单位受各种因素的影响往往具有差异性，总体单位不能反映总体的一般特征和规律性。这就要求统计必须对足够多的总体单位进行综合研究，以清除偶然因素的影响，反映出总体的必然性、数量特征和规律性。大量观察法可以是对总体所有单位进行研究，如普查；亦可以是对能表现现象本质特征和规律性的部分单位进行非全面统计，如抽样调查。

2. 统计分组法

统计分组法是对所研究的客观现象按照一定的分类标准或标志，把研究的全部总体

单位划分为不同的组别,用以区别现象的各种不同类型,揭示总体的内部结构及其分别特征,反映现象之间的相互关系,达到认识客观现象的本质特征和规律性的目的。统计分组法既是统计资料整理的基本方法,也是一种重要的统计分析方法。

3. 综合指标法

综合指标法是统计研究中不可缺少的基本方法。综合指标是表明客观现象总体数量特征的各种数量化的概念及其数值。常用的综合指标有总量指标、相对指标、平均指标等。综合指标法是统计整理、统计分析的基本方法,其他各种统计分析方法均以它作为基础,如时间序列法、指数法、抽样法、相关法等都离不开综合指标的对比研究。

4. 归纳推断法

归纳法是从个别到一般的推理方法。综合指标法是对个体的数值综合汇总成总体的数值,以概括反映总体的一般数量特征,就是归纳法的具体体现。统计分组法是将总体各个个体划分为不同的组别,以研究总体的内部结构和分别特征,也是一种归纳法。推断法是根据样本数据来推断总体特征的方法,常用的推断法有抽样推断、参数估计、假设检验、产品质量检测、统计预测决策法等。

1.5.2 统计研究过程

统计研究是一个复杂的过程,主要包括以下几个步骤。

1. 统计设计

统计设计是根据统计研究的目的和要求,对统计研究的对象、内容、方法及程序所做的通盘考虑和安排。统计设计的结果表现为各种设计方案,如统计调查方案、统计指标体系、分类目录、统计报表制度、统计整理与分析方案等。它是统计工作在实践中展开前所做的计划性安排,既包括统计方案的全面设计,又包括人、财、物等各种资源的安排。一项统计工作能否顺利进行与统计设计是否全面、周密有很大关系。

2. 统计调查

统计调查是根据统计设计的要求收集统计数据的阶段,是定量认识的起点。统计调查与其他调查活动的不同点在于,精心设计的统计调查方法更能保证调查活动收到事半功倍的效果。统计调查的方法分为普查、统计报表制度、重点调查、典型调查和抽样调查,这五种方法各有特点。普查收集的资料最全面、详细、准确,但缺点也很明显,那就是费钱、费时且费力。重点调查和典型调查是我们短期迅速获得数据的好方法,但容易产生无法控制的误差,且随意性很大。抽样调查是最经常使用的一种方法,原因在于它具有时效性强、低成本、高质量和科学性的特点,虽然抽样调查的结果存在误差,但这个误差是我们可以接受和控制的。抽样调查有一整套科学方法,是统计学的重要分支。

3. 统计整理

统计整理是对通过统计调查收集到的原始数据、统计资料,根据研究的需要进行分类、汇总等。统计整理是一个对统计调查结果去粗取精、去伪存真的过程,就是要对统计调查收集到的原始数据进行把关。统计整理后的结果就是我们通常看到的各种统计资料。

4. 统计分析

统计分析是对统计整理的数据进行再加工和深加工的过程,主要是采用各种分析方法,计算各种分析指标,以揭示被研究对象的总体数量特征和规律性,从而达到统计研究的最终目的。这一阶段充分体现了统计学的精华,主要表现为对各种统计分析方法的应用,如统计指标、时间序列分析、统计指数和回归分析等。

统计研究过程框架如图 1-3 所示。

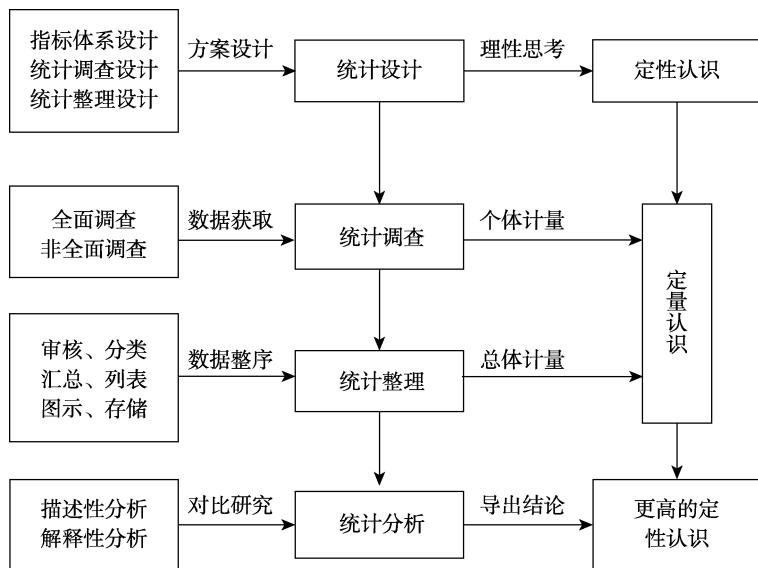


图 1-3 统计研究过程框架

1.6 案例：高考志愿

1.6.1 案例背景

湖南的赵大爷一家祖祖辈辈都是地地道道的农民,全家节衣缩食供唯一的孙女读书,希望知识改变命运。孙女芳芳刻苦努力,不负众望,2017 年高考成绩 588 分(理科)。在填高考志愿之际,赵大爷问在中南大学工作的杨老师:能否填报中南大学?

1.6.2 案例分析

1. 高考志愿填报的基本步骤

考生该如何填报高考志愿呢？一般而言，应经过以下几个基本步骤。

（1）熟悉规则。每一位考生可以填写 10 个平行志愿——10 所大学，每个志愿下可填写 6 个专业，高考录取遵循“分数优先、志愿优先”的原则。

（2）收集资料。查询历年（尤其是近 3 年）国家最低控制分数线和各高校的抛档分数线。

（3）数据分析。根据当年的最低控制线和历年的最低控制线，预测各高校当年的抛档分数线。

（4）统计决策。根据考生当年的高考分数和预测的高校抛档分数线，选择学校。根据选择的学校，进一步查询该校历年各专业的录取线，结合当年实际分数和自己的兴趣，选择相关专业。

2. 芳芳的高考志愿填报

2017 年芳芳高考理科成绩 588 分，湖南省本科一批理科国家最低控制线是 505 分，在一本院校选择。而中南大学属于一本，可以进一步分析被中南大学录取的可能性。中南大学理工科 2014—2016 年最低控制分数线依次为 522 分、526 分、517 分，抛档分数线依次为 612 分、614 分、610 分。抛档分数线高出最低控制分数线 90 分、88 分和 93 分，平均高出约 90 分。预测 2017 年中南大学理工科抛档分数线至少 595 分，芳芳不宜填报中南大学。事实上，2017 年中南大学理工科实际抛档分数线为 601 分。

习 题 1

- 1.1 怎样理解统计学与统计数据的关系？
- 1.2 统计数据分为哪几种类型？比较不同类型数据的特点。
- 1.3 什么是变量？变量有哪些类型？
- 1.4 举例说明连续变量与离散变量。
- 1.5 举例说明总体、样本、参数、统计量等概念。
- 1.6 指标与标志的区别是什么？请举例加以说明。
- 1.7 举例区分不同的统计尺度。
- 1.8 结合实际问题的论述统计研究过程。
- 1.9 常见的统计研究方法有哪些？请举例加以说明。
- 1.10 结合实际问题的谈谈商务统计的应用。

第 2 章

统计数据的收集

引例

《2017 年国民经济和社会发展统计公报》

2018 年 2 月 28 日国家统计局发布的《2017 年国民经济和社会发展统计公报》显示了 2017 年国民经济和社会发展情况。我国国内生产总值 82.7 万亿元,按年平均汇率折算超过 12 万亿美元,占世界经济的比重 15%左右;年末内地总人口 139 008 万人、城镇新增就业 1 351 万人、全年全国一般公共预算收入 172 567 亿元、粮食产量 61 791 万吨、全社会固定资产投资 641 238 亿元、社会消费品零售总额 366 262 亿元、货物进出口总额 277 923 亿元、在学研究生 263.9 万人、在校普通本专科生 2 753.6 万人、新改建高速公路里程 6 796 千米、新建高速铁路投产里程 2 182 千米、农村贫困人口比上年末减少 1 289 万人、年末国家外汇储备余额 31 399 亿美元……笔笔数据记录了过去一年全国上下拼搏奋进的累累硕果,展示了我国发展取得的历史性成就、发生的历史性变革,经济社会发展主要预期目标实现,开启了高质量发展时代新征程。

为什么国民经济与社会发展统计公报中选择这些指标,指标数值是如何调查获得的?

2.1 统计设计

2.1.1 统计设计的概念和意义

统计设计是根据统计研究的目的和研究对象的特点,筹划和安排统计工作的各个方面和环节,制订各种具体实施方案的工作阶段。统计设计阶段是统计工作过程不可缺少的环节,制订的各种统计工作方案是统计工作的指导依据。统计设计阶段制定统计指标体系、统计分类目录、统计报表制度、统计调查方案、统计汇总或整理方案、统计分析提纲或方案等方面的工作方案,也是统计设计的基本任务。

统计设计是整个统计工作过程的第一阶段,是统计调查、统计整理、统计分析进行之前的必要准备阶段,统计设计制订的各种工作方案是统计工作实施的基本依据和重要保证。从认识角度看,对客观现象和事物的认识从定性认识开始,没有定性认识作为基础,就无法进行定量分析。统计设计是对统计总体的定性认识和定量认识的连接点,没有统计设计阶段的筹划安排,就不知道去调查什么和怎么调查,也不知道去研究什么和

怎么研究,更不知道如何组织。从工作角度看,统计是一项高度集中统一的工作,需要把认识对象作为一个整体来进行全面综合的反映和研究,必须经过事先的通盘考虑、统筹安排才能制订出科学合理、行之有效的统计设计工作方案,才能为后续统计工作质量提供保证。从统计实践来看,加强和重视统计设计工作对完成整个统计工作,保证统计工作的质量是必需的,是使整个统计工作协调有序顺利进行的重要保证。

2.1.2 统计设计的种类与内容

1. 横向设计与纵向设计

统计设计的内容是对统计研究对象的内容和统计工作过程通盘规划与统筹规划的综合,统计工作的研究目的不同,统计设计的内容就存在差异。按设计的角度不同,统计设计分为横向设计和纵向设计两个方面。

1) 横向设计

横向设计是统计研究对象的内容设计,就是对统计工作涉及的各个方面或组成部分的设计,可分为整体设计和专项设计。整体设计是将统计研究内容或认识对象作为一个整体,对其整个工作进行全面设计。如为调查某企业生产经营状况、某区域社会经济发展状况,对该企业整个生产经营工作或该区域社会经济发展的整体统计工作的设计就是整体设计。专项设计是对研究对象的某一个方面或组成部分的具体内容的设计。如对某区域统计调查工作中农业、工业、贸易、教育、科技、金融等工作中的某一方面设计,对某企业的人力资源、物资、资金、生产、供应、营销等工作中的某一方面设计,等等。横向设计中,整体设计是主要的,专项设计在整体设计的基础上进行,并且服从整体设计的统一安排。

不同的横向设计的内容不同,但主要的基本内容是具有共性的,包括以下几个方面。

(1) 统计指标和指标体系的设计。统计指标和统计指标体系是统计设计首先要解决的问题和主要内容。统计指标是统计工作研究客观现象总体数量特征所运用的最基本、最重要的手段。一系列相互联系的统计指标组成统计指标体系可全面系统地反映和体现总体主要特征。

(2) 统计分类或分组的设计。统计分类或分组的设计是统计指标和统计指标体系紧密相联系的另一个重要的设计,但侧重于对象分类,是深入研究客观现象数量特征的重要方法。

(3) 统计调查方式和方法的设计。对客观事物的数据进行收集可采用不同的方式、方法,无论采用哪一种都要从多方面进行考虑,使得调查的效果最好。

(4) 统计分析方法的设计。统计分析是统计工作最重要的阶段,是统计工作出成果、出结论的阶段,必须选择合理的统计分析方法。对总体的本质特征作出定性和定量分析,才可能获得符合实际的结果。

(5) 统计工作各个部门和各个阶段的协调与关联。统计工作中涉及很多部门,要经过多个阶段,只有建立了有力的协调与沟通机制,才能确保统计工作高效有序地进行。

(6) 统计力量的组织与安排。统计工作的开展离不开一定的人力、物力和财力,统

计设计需要对统计人员的培训与使用、对有关经费的筹集、分配与使用和对物资、设备的调拨使用等作出妥善的安排。

2) 纵向设计

纵向设计是统计工作过程的设计,是对统计工作的各个环节的设计,可分为全阶段设计和单阶段设计。全阶段设计是对统计工作全过程的设计,也就是对统计工作中经历的各个阶段所进行的全面设计,既包括从确定统计内容、统计指标体系开始到分析研究的全过程通盘安排,也包括统计工作阶段所有的设计方案及工作阶段间的协调统一。单阶段设计则是对统计工作过程中某一具体阶段所进行的设计。一般体现某一阶段具体的统计设计方案,如统计调查方案设计、统计整理方案设计、统计专题分析方案设计等。全阶段设计是主要的,单阶段设计在全阶段设计的基础上进行,并且服从全阶段设计的安排。

纵向设计在主要内容上是具有共性的,以全阶段设计为例,其主要内容包括以下几个方面。

(1) 明确规定统计研究的目的。统计设计中首先确定的是统计研究的目的,它是统计设计其他内容的基础,必须根据调查组织单位的实际情况与要求确定。

(2) 制订统计调查方案。制订统计调查方案是指在确定统计工作的目的的基础上,明确统计总体、样本、总体单位、填报单位的范围;规定统计的空间标准和时间标准;根据统计目的和分析研究的要求,制定出调查登记的项目和指标;设计好获取资料的方式方法;等等。

(3) 明确统计整理的基本任务。统计整理中要考虑如何对资料进行审核、筛选和排序,如何明确分组目的、选择分组标志与分组方法、确定组数组距和组限、对资料分类或分组,如何选择统计整理指标和确定资料汇总的方法,如何编制的统计图表,等等。

(4) 选择恰当的统计分析方法。常见的统计分析方法有抽样推断、假设检验、相关分析与回归分析、主成分分析与因子分析、聚类分析与判别分析、时间序列分析、统计指数分析、贝叶斯分析等。

(5) 统计工作全过程的组织工作。统计工作组织中要成立统计工作机构,妥善安排各方面统计人员的招聘、培训与使用,规定各个阶段的工作进度、时间安排,对有关经费的筹集与分配、物资设备的准备与使用等,统筹规划各个工作阶段的联系和各阶段的工作基本方法,制订保证统计资料准确性、及时性、完整性的措施。

2. 长期设计、中期设计和短期设计

按设计时期的长短不同,统计设计可分为长期设计、中期设计和短期设计三类。

1) 长期设计

长期设计是对较长时期的统计工作的设计,如对未来五年以上的统计工作的规划。

2) 中期设计

中期设计所包含的时期长度介于长期设计与短期设计之间,一般是对一年以上到五年之内的统计工作的设计。

3) 短期设计

短期设计是对较短时期的统计工作的设计,时间长度通常在一年以内,如对一年、

半年或一个季度的统计工作的安排。

通常统计设计的时期越长,设计的内容越简略、概括;统计设计的时期越短,设计的内容越详细、具体、明确。

2.1.3 统计指标和指标体系设计的内容

统计指标和指标体系设计的内容包括以下几方面。

1. 确定统计指标体系的框架

统计指标体系的框架主要确定指标体系包括哪些指标,哪个指标是指标体系中的核心指标,各个指标之间具有什么样的联系,等等。核心指标是统计指标体系中为主的指标,核心指标的确定取决于许多因素,如统计对象的性质、统计总体的范围、统计研究的目的等;核心指标不是固定不变的,它可以随着客观情况的变化而变化。目前,我国国民经济核算指标体系以国内生产总值作为核心指标,工业企业生产经营指标体系通常以利润作为核心指标。

核心指标确定以后,一是应当围绕核心指标从不同的角度设计各种相互依存、相互联系的指标体系。例如,以国内生产总值指标作为国民经济统计指标体系的核心指标,考虑到国内生产总值的形成、国民经济中的各种事物和现象总量之间的主要比例关系等,如农业、工业和服务业状况,人口及就业,收入支出,国内外贸易与投资,社会保障,教育科技文化体育卫生,资源环境等方面社会经济指标及其变化发展。这样设计各种统计指标,对各种比例的数量进行观察。二是根据影响核心指标的活动成果因素进行设计。例如,以利润指标作为工业企业统计指标体系的核心指标,而利润的形成依赖于企业各种物资的供应和消耗状况、设备的利用和运转情况、工艺技术水平、劳动生产率高低等各种因素,可设计指标对各种因素的数量进行观察,从而形成工业企业统计指标体系。三是根据核心指标的活动成果构成进行设计。例如,以财政总收入和财政总支出作为财政部门统计指标体系的核心指标,考虑到财政收入的来源有各项税收、企业收入、债务收入、其他收入等,财政支出有经济建设投资、科教文卫事业费用、国防支出、行政管理费用、其他支出等,这样可设计各种指标分别对财政收支的各组成部分的数量进行观察,从而形成财政部门统计指标体系。

2. 确定各项指标的名称、内涵和外延

统计指标是表明总体数量特征的概念和数值,必须首先有一个明确的概念,才能据此计算出指标数值。统计指标名称的确定既要考虑研究对象本身的特点,又要依据一定的理论、规定和惯例。有些概念具有明显测度功能,如价格、成本、工资、利润等直接体现了测度的特征,可以直接用作指标名称。有些概念不具备测度功能,如企业规模、产品质量等,必须通过合适的方法选择能体现本质特征、具有测度功能作为统计指标,通常有:一是从与性质相近的具有测度功能的特征中选择最合适的作为统计指标,如企业规模是不可度量的,可选择与其性质近似的职工人数、资产总额、产量、销售收入等

特征作为统计指标；二是从变动的原因或形成的结果等具有测度功能的特征中选择最能体现本质特征的作为统计指标，如技术水平可选择接受教育年限、学历、职称、技术等级等特征作为统计指标；三是根据性质与运动规律设计专门的测量手段作为统计指标，如设计智商（儿童达到的智力年龄与实际年龄之商）作为观察儿童智力的统计指标；四是根据品质差异设计顺序或等级，如将产品质量按标准划分为特等品、一等品、二等品等。

任何一个统计指标的概念都包括指标的实质含义（内涵）和所属范围（外延）两方面，指标的内涵是对事物本质属性的反映，而指标的外延则是指标计算的范围，它们是设计统计指标的第一个要点。也就是说，设计任何指标首先要明确确定它是什么，界限划在什么地方，什么内容应该计算在内，什么内容不应该计算在内。指标体系的各个指标应该统一口径，其口径范围应以核心指标的口径范围来确定，指标的不可比给统计分析带来困难。

3. 确定各项统计指标的计量单位

计量单位是统计指标的组成部分之一，指标的计量单位有不同的表现形式。有些统计指标可以有若干种计量单位，可在实物单位、价值单位、劳动单位之间进行选择，而且在同种类型里也可以有多种选择。不恰当的计量单位，给统计的综合带来很大的不方便，需要从若干种计量单位中确定一种比较适合的、统一的计量单位。对于自然形成的计算单位可以依据事物和现象的计量标准来确定，如汽车产量以辆计量、销售收入以元计量等。对于利用两个及以上特征通过一定形式结合而成的统计指标，其计量单位应是结合相应特征，多数使用复合单位计量或无名数计量的。

4. 确定各项统计指标的计算方法

一个指标可能有多种计算方法，不同方法计算的结果有可能产生相当程度的差异。而不同指标的计算方法难易程度差别很大，有些指标如职工人数、产品产量的计算方法等比较简单，通过点数、测量、登记和汇总等；有些指标如国内生产总值、居民消费价格指数等的计算都很复杂，依据一定的理论，如何选择计算方法特别重要。当一个指标有几种可供选择的计算方法时，既要考虑研究对象的特点，也要考虑工作条件、工作惯例与有关规定，还应比较更符合这些理论和实践的要求决定计算方法的取舍。

5. 确定统计指标的计算时间和空间范围

统计指标的计算时间就是统计指标所属的时间，也叫作统计指标的时间标准。统计指标的时间标准具体表现形式包括时期标准和时点标准两种，其中时期标准就是指标计算的起止期限，用一段时间表示，如一个月、一个季度、一年等；时点标准则是指标值所属的标准时点，就是某一具体时刻，如以某年的年初、年末、月初、月末等为时间标准。计算时间是由统计指标性质、特点、需要和可能来决定的。

统计指标的空间范围就是统计指标的空间标准，是指地区范围和组织系统范围，主要是地区范围。

2.1.4 统计指标和指标体系设计的原则

统计指标和指标体系设计的原则包括以下几方面。

1. 科学性原则

统计指标和指标体系的科学性原则包含三重含义：一是指标体系的设计要符合客观现象本身的性质特点，反映客观事物总体的真实情况、内部及相互之间的数量关系；二是指标体系的设计必须以正确的、科学的理论为指导，根据各种社会经济理论对总体进行深刻的定性分析，以便使设计的核心指标、指标数量、指标口径、计算时间、计算方法和计量单位等都符合科学原则的要求；三是符合我国的国情和社会经济，从实际出发。

2. 目的性原则

统计指标和指标体系的设计要依据统计研究的目的或目标。指标是目标的具体化、行为化和操作化，必须充分地反映目标，与目标相一致。同一研究对象具有多方面的属性，一组指标体系反映一个方面的属性。例如，对国民经济活动进行统计核算，反映生产过程时要设置一套指标体系，反映分配过程时又有一套指标体系。要保证指标体系内各条具体指标的目标的一致性，不能把两条相互冲突的指标放在同一指标体系中。

3. 整体完备性原则

指标体系内各项指标之间是相互联系、相互制约、相互依赖、相互补充的关系，彼此之间协调一致，它们完整地反映被研究对象的基本特性，合在一起表现出一定的整体功能。指标体系不应遗漏任何一个重要指标，才能全面地、毫无遗漏地再现和反映目标。统计指标和指标体系的设计，要从整体上考虑各个指标之间的联系，指标口径、时间、空间和计算方法的确立要从全局出发，考虑到彼此间的联系。

4. 结构层次性原则

指标体系内的各个指标间不能是杂乱无章的，而应形成一定的结构和层次，即要求明确何者为核心指标，何者为次要指标，指标间如何联系，等等。各指标间只有形成一定的结构层次，才能使指标显示出一定的整体功能。

5. 可行性原则

可行性原则又称可能性原则，包括四个方面内容：第一有足够的信息可利用，第二有足够的人力、物力和财力可利用，第三有切实可行的量化方法可利用，第四信息是可能获得的。

6. 可比性原则

要使设计的统计指标和指标体系能满足管理的需要与科学研究的要求，要注意不同地区、不同单位、不同时期统计指标和指标体系的可比性，这样便于统计指标和指标体

系的相互比较, 便于随着社会经济的发展, 改革和充实统计指标与指标体系, 有利于各个指标在不同时期的相互衔接和相对稳定, 以便于分析、研究事物发展变化的规律性。

7. 相互独立性原则

指标体系是由一组相互紧密联系的指标结合而成的, 但各个指标必须是相互独立的, 同一层次各指标必须不存在任何包含与被包含的关系、不存在因果关系、相互不重叠。原因主要有二: 一是不独立指标实质上反映了同一事物的同一方面, 它的存在对整个指标体系没有贡献, 还加大了统计工作量; 二是不独立指标, 重复计算加大了在指标体系中的指标权重。

8. 统一性原则

统一性原则要求指标体系中各指标在分类标准、指标口径和计算方法等方面协调一致。统计指标和指标体系的设计要力求与计划、会计和业务核算相统一, 即设计时必须考虑到计划、会计、业务核算的实际情况和统计的需要, 尽可能地使各种核算的原始记录统一、计算方法一样, 统计范围、经济内容相同, 起止时间一致, 等等。

2.1.5 统计指标体系的构成

指标必须与目标相一致, 可以通过分解目标的方式来形成指标体系, 这是建立指标体系的基本途径。对于复杂的系统, 还可以在目标与指标之间设置若干中间过渡环节, 中间环节通常称为次级目标。我们一般可以把指标分解为总体指标、结构指标、单项指标三个基本层次, 分别对应零级指标、一级指标、二级指标, 有时甚至分解为四级指标、五级指标。指标体系包含指标层级的多少, 一般应由研究问题的复杂性决定, 运用时要根据实际情况确定。

多级指标体系有树形多级指标体系(图 2-1)和非树形多级指标体系(图 2-2)两种表现形式, 两种指标体系中至少包含三级指标, 同一指标层中的不同指标之间存在并列

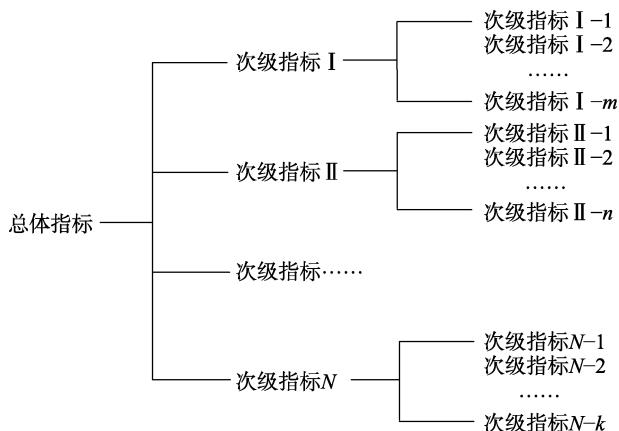


图 2-1 树形指标体系的构成

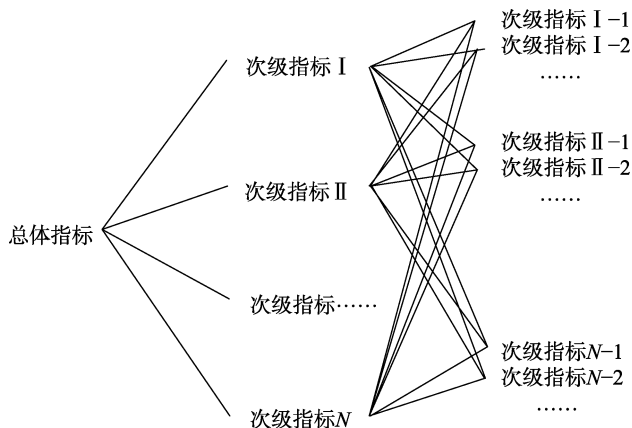


图 2-2 非树形指标体系的构成

关系，彼此之间不存在隶属关系。但是在相邻的两级指标中，树形多级指标体系中下一级的任一指标只隶属于与之相邻的上一级指标中的一个指标；非树形多级指标体系中下一级的某一指标可以同时隶属于上一级中的多个指标。

2.2 统计调查的方式与方法

2.2.1 统计调查的意义与基本要求

1. 统计调查的意义

统计调查就是统计资料的收集，是根据统计研究的目的和要求，有组织、有计划地向调查对象收集统计资料的过程。

统计调查过程中收集的统计资料，分为原始资料和次级资料。原始资料又称初级资料，指为特定目的而进行实地观察或通过对党政机关、企事业单位和其他组织或团体调查而获得的第一手资料。它是直接向调查对象收集各个总体单位的资料，是没有经过分组汇总、需要由总体单位过渡到总体的统计资料。原始资料是资料的直接来源，主要有调查资料和实验资料。其中，调查资料是通过统计调查方法获得的资料，通常是对社会现象而言的，一般取自有限总体；实验资料是通过实验方法得到的资料，通常是对自然现象而言的，也被广泛运用到社会科学中，如经济学、管理学、心理学、教育学、社会学等。作为一个相对独立的工作阶段，统计调查的主要任务是收集原始资料。

次级资料是指已经经过加工整理的资料，如统计公报、统计年鉴、会计报表、研究报告、报纸杂志的摘引等。次级资料是资料的间接来源，可分为系统外部资料和系统内部资料。其中系统外部资料主要包括政府部门尤其是统计部门公布的有关资料，如各类统计年鉴、统计公报等；各类经济信息中心、信息咨询机构、专业调查机构等提供的资料；各类专业期刊、报纸、书籍所提供的资料；各种会议，如博览会、展销会、交易会及专业性、学术性研讨会上交流的有关资料；从互联网或图书馆查阅到的相关资料；等

等。而系统内部资料,主要包括业务资料,如与业务活动有关的各种单据、记录;活动过程中的各种统计报表;各种财务、会计核算和分析资料;等等。

次级资料已加工整理成型,内容广泛、收集容易、采集成本低,而且作用广泛,可以用于分析所要研究的问题、提供研究问题的背景、帮助研究者更好地定义问题、检验和回答某些疑问与假设、寻找研究问题的思路和途径等,缺点是时效性、可靠性相对较差。但是在使用时注意进行评估,考虑资料是谁收集的、为什么目的而收集的、资料是怎样收集的、什么时候收集的等因素,从而大大提高次级资料的准确性,很多情况下收集次级资料在研究中可以优先考虑。

统计工作过程就是收集统计资料、进行加工整理数据,而后开展分析研究的过程。在整个统计工作过程中,统计调查担负着提供基础资料的任务,所有的统计计算、研究和分析都是建立在统计资料收集的基础上的。因此,统计调查是整个统计工作过程的基础,是统计整理、统计分析、预测和决策的前提,关系到整个统计工作的成败。

统计调查是决定整个统计工作质量的首要环节,调查工作的好坏、取得资料是否完整与正确,将直接影响到以后各个阶段工作的好坏,影响整个统计工作任务的完成。只有搞好统计调查,才能保证统计资料整理及统计分析结论的正确性,进而认识客观事物的规律性。

2. 统计调查的基本要求

统计调查在统计工作过程中起着举足轻重的作用,这对统计调查提出了一些要求,具体表现在以下几个方面。

1) 准确性

统计资料的准确是统计的第一生命。要求统计调查所得到的统计资料符合客观实际、真实可靠,避免人为因素可能造成的消极影响。只有基于准确可靠的统计信息,才有可能得到对客观事物的正确认识。

2) 及时性

自古以来人们对于信息的时效性一直有很高的要求,它是人们运筹帷幄、决胜千里的基本保证。统计调查的目的是获取和提供满足各级各类决策者需要的统计信息,其时效性特别关键,要求各项统计资料必须迅速、及时,做到及时反映、及时预报,才能收到“雪中送炭”的功效;不能满足时效的资料犹如“雨后送伞”,失去了它的价值。在信息化的今天,统计资料的迅速、及时显得尤为重要。

3) 完整性

统计调查取得的资料必须系统和完整,该调查的单位就必须调查,该有的调查项目就必须有,并且应保证各调查项目之间有必要的逻辑联系,不仅达到统计调查资料最大的开发利用价值,而且要确保达到对社会经济现象总体特征的正确认识目的。

4) 经济性

在满足资料质量要求的前提下,尽可能降低成本,要力争最大限度地降低人力、物力、财力消耗,以达到尽可能好的效果。

对统计调查资料的准确、及时、完整、经济等方面的要求,是相互联系、相辅相成

的,不能孤立地、片面地去理解。强调统计资料的及时性,更要强调它的准确和完整,任何虚假的、残缺不全的“统计资料”都是没有价值的,甚至是有害的;而准确、完整的统计资料必须及时提供才能发挥其应有的作用,为此应避免“放马后炮”;过高的成本会使统计资料失去应有的效用与价值。统计调查要注意把四者辩证地结合起来,以准为基础,做到准中求快、准中求全、力求节约。

2.2.2 统计资料的收集方式和方法

在组织统计调查时,应根据不同的调查对象和调查的目的、任务,采用不同的调查方式。

1. 全面调查和非全面调查

按调查对象的范围,统计调查分为全面调查和非全面调查。全面调查是对调查对象中的全部总体单位进行调查,如普查、全面统计报表等。全面调查能比较准确、完整地掌握统计资料,了解事物和现象的全貌,但它往往调查对象范围较广、涉及面较宽、调查单位较多,需要花费较多的人力、物力、财力和时间,操作难度大、组织工作复杂,不利于对事物做深入细致的调查和研究。由于全面调查对所有个体进行调查,其调查内容不宜太多,适合总体规模相对较小或者调查特别重要的基础资料的场合。非全面调查是对调查对象中的部分总体单位进行调查,如抽样调查、典型调查、重点调查和非全面统计报表等。非全面调查的调查单位少,可以集中力量做深入细致的调查,能够调查更多的指标;通过非全面调查可以了解客观事物的详细情况及其发展变化的原因;同时实施灵活简便,能及时经济地取得统计资料,并可根据调查资料估计或推断全部总体的情况。非全面调查在实践中应用非常广泛,它既适用于有限总体,也适用于无限总体,一般能采用非全面调查满足要求就不必进行全面调查。

2. 经常性调查和一次性调查

调查时间的连续性,统计调查分为经常性调查和一次性调查。经常性调查是对调查对象进行连续不断的登记和调查。经常性调查的目的是取得反映事物和现象随时间变化在数量上变动情况的资料,如企业的产品产量、原材料燃料和动力消耗、商品销售量、工业增加值就是逐日、逐月、逐季地连续登记获得的,是采用各种统计报表来进行的。经常性调查适合于时期指标,时期指标数值与时间长短有直接关系,经过一段时间才会发生较大的变化量,只有连续不断地登记和调查才能获得事物与现象正确、完整、连续的资料。一次性调查又称一时调查,是对调查对象在某一时刻的状况进行不连续的调查。一次性调查是间隔一段时间对事物和现象进行调查,适用于时点指标,由于时点指标在一段时期内的变化量通常相对较小,不需连续不断地登记或调查,如对人口资料、固定资产和商品库存量等的调查。一次性调查可以定期或不定期进行,定期调查是每隔一个固定间隔期而进行的调查,如每月末登记职工人数、每五年一次的投入产出调查等;不定期调查是根据调查目的和需要,在间隔不等的时间上所进行的调查。随着科技的进步,

很多一次性调查的事项可以进行连续不断的登记、监测,如借助电子计算机对企业或单位各类商品库存量进行连续监控。

3. 统计报表和专门调查

按调查的组织方式不同,统计调查可分为统计报表和专门调查。

1) 统计报表

统计报表是按统一规定的指标体系、表格形式、报送程序和报送时间,自上而下统一布置,定期地自下而上报送基本统计资料的一种调查方式。统计报表主要用于收集全面的基本情况,也常为重点调查等非全面调查所采用。

相对于其他统计调查方式,统计报表具有以下特征。

(1) 统一性。在统计报表实施范围内,所有单位按照国家或地方统一规定的指标体系、表格形式、报送程序和报送时间等报送统计资料,统计报表的指标含义、计算方法、计算口径是统一的。

(2) 强制性。统计报表一般由国家或地方政府授权的有关部门根据统计法规统一设计、布置、报送和管理,具有法律的强制力。弄虚作假、不及时报送会受到行政的、法律的制裁。

(3) 周期性。周期性也称经常性,统计报表必须严格按照规定的时间和程序上报,规定范围内的单位要定期收集、报送国民经济和社会发展基本情况的某些方面的资料,而且填报的项目相对稳定,可以完整地积累形成一系列时间序列资料。

(4) 相对可靠性。统计报表是根据国民经济和社会发展宏观管理的需要而周密设计的统计信息系统,是建立在基层单位日常业务的原始记录和核算数字基础上的,通过分门别类地系统积累和总结形成统计台账,再到统计报表,使调查资料具有可靠的基础。

统计报表按照不同分类标准有不同的分类。

(1) 按调查范围,统计报表可分为全面统计报表和非全面统计报表。全面统计报表由调查对象的全部单位填报,非全面统计报表按要求由调查对象的一部分单位填报。

(2) 按填报单位,统计报表可分为基层统计报表和综合统计报表。基层统计报表是由基层企业、事业单位填报的报表,为有关管理部门提供各单位基础资料;综合统计报表是由管理部门或单位根据基层报表逐级汇总填报的报表,汇总后得到各级的基本统计指标。

(3) 按报送周期长短不同,统计报表可分为日报、旬报、月报、季报、半年报和年报。日报、旬报、月报、季报和半年报合称为定期报表。日报和旬报称为进度报表,主要用来反映生产、工作的进展情况,也用于报告重大或紧急统计项目或指标状况;月报、季报、半年报用于检查各月、季、半年的生产经营等工作情况,反映各月、季、半年国民经济重要基础数据进展状况、经常性计划和合同执行情况检查;年报用于年度总结和分析,全面反映全年社会经济活动的成果,检查年度国民经济与社会发展计划的方针、政策和计划贯彻执行情况等。

(4) 按报表实施范围,统计报表可分为国家统计报表、部门统计报表和地方统计报表。国家统计报表是国民经济基本统计报表,由国家统计部门统一制发,用以收集全国

性的经济和社会基本情况,包括农业、工业、基建、物资、商业、外贸、劳动工资、财政等方面最基本的统计资料;部门统计报表是为了适应各部门业务管理需要而制定的专业技术报表;地方统计报表是适应地区特点而补充制定的地区性统计报表,是为本地区的计划和管理服务的。

(5) 按报表内容,统计报表可分为基本统计报表和专业统计报表。基本统计报表是提供国民经济各部门、各行业最基本的统计资料,作为国家和地方宏观决策与调控依据;专业统计报表是根据部门业务特点设计,为部门、企业和事业单位管理提供资料。专业统计报表比基本统计报表内容要多些,详细些,但不得与其相抵触。

统计报表数据的来源主要包括以下几个方面。

(1) 原始记录。原始记录是基层单位通过一定的表式,用数字或文字对事物和现象的过程与结果所作的最初记录,是未经过任何加工整理的初级资料。原始记录形式多种多样,如一个工业企业有:生产进度、产量、品种、质量等产品生产的原始记录;职工的基本情况、变动情况、考勤、工时利用等劳动方面的原始记录;设备增减、运转、利用、维修、完好状况等使用的原始记录;原材料燃料的收入、拨出、领用、消耗、利用、储备等原始记录;等等。从原始记录的形式看,包括综合性原始记录和专用性原始记录。综合性原始记录是一张原始记录上记录了调查对象的几个方面情况,如以产品为对象的综合性原始记录的施工单、加工路线单等反映工序的产量、质量和工时利用情况;以生产者对象的综合性原始记录的个人和生产班组的生产日记,反映产量、质量和工时利用;以生产设备为对象的综合性原始记录反映设备的操作情况和生产成果。专用性原始记录是在一张原始记录上只记录调查对象的某个方面情况,如入库单、质量检验单、废品通知单、领料单、退料单等反映某项生产经营管理工作。

(2) 统计台账。统计台账是根据统计报表和单位管理的需要,以一定的表格形式,按时间顺序分类登记和定期汇总统计资料的一种记录表册。是基层单位根据统计报表和管理需要而设置的一种系统积累统计资料的整理性的记录表册,是对原始记录的初步整理核算。统计台账有综合台账和专用台账两种形式,综合台账按时间顺序,同时登记若干有关指标的数值;专用台账按时间顺序,同时登记各单位某项指标的数值。统计台账的记载必须与原始记录数据衔接,做到准确、及时、连续、完整、清晰,查询方便。

(3) 内部报表。内部报表是反映事物和现象在一段时间内的发展变化状况与综合成果的报表。如企业内部报表反映企业内部各层级在一定时期内生产、劳动、设备原材料和财务成本情况与综合成果,又为编制企业统计报表、制订计划、指导生产提供依据。内部报表是为了向本单位领导提供情况而编报的,主要反映产品产量、产品工时、产品成本、原材料消耗、设备运转、工时利用、技术革新及主要技术经济指标等方面的报表。内部报表一般不需要统一规定的格式,也没有统一的指标体系,具有灵活性、及时性、针对性和网络性等特点,可以及时满足内部管理的需要。内部报表应严格控制,宁缺勿滥,内容要力求精简。

(4) 对外报表。对外报表是按上级规定或为了其他需要而编报的报表。企业的对外报表是企业向外提供的,供投资人、债权人、政府部门、其他企业和个人了解有关信息的统计报表,常见的是财务报表,如资产负债表、现金流量表、利润与利润分配表、资

产减值准备明细表、股东权益增减变动表、分部报表和其他有关附表等。

统计报表有专门的制度。统计报表制度是统计报表按照国家统计法规制定、实施和管理的一整套办法。执行统计报表制度,是各地区、各部门、各单位按照国家的法律规定必须向国家履行的一种义务。统计报表制度分为国家统计报表制度、部门统计报表制度和地方统计报表制度。国家统计报表制度是各级政府统计部门实施国家统计调查项目的业务工作方案,由国家统计局单独制定或者与国务院有关部门共同制定,包括周期性普查制度、经常性调查制度和非经常性调查制度。周期性普查制度是由国务院组织,每隔一段时间对我国社会经济发展状况进行一次普查的统计调查制度;经常性调查制度是年度和定期地经常性统计的统计调查制度,我国现行的共 30 项;非经常性调查制度是在一定时期内持续实施或一次性实施的专项调查、试点调查等临时性统计调查制度。部门统计报表制度是国务院有关部门实施部门统计调查项目的工作业务方案。地方统计报表制度是指县及县以上地方人民政府及其部门实施地方统计调查项目的工作业务方案。统计报表制度根据《中华人民共和国统计法》的有关规定制定,遵循统一规范化的原则和要求,按照统一的统计标准,以便于统计数据自动化处理,相互调用,实现统计信息共享。

我国统计报表制度的基本内容有以下几个方面。

(1) 指标和指标体系的确定。统计报表中一系列指标和指标体系的设置,必须以理论为指导、从实际出发、从当前社会经济发展需要出发,做到既能满足国家又能满足各级主管部门的需要。

(2) 报表表式的设计。报表表式是由国家统计部门根据研究的任务与目的而专门设计制定的统计报表表格,是统计报表制度的主体,设计应简明扼要,清晰易懂。报表表式包括主栏项目、宾栏项目以及补充资料项目、表名、表号、填报单位、报告期别、报送日期、报送方式、单位负责人及填报人签署等,并分为基本表式和专业表式。

(3) 报表的实施范围。报表的实施范围是指填报单位(报告单位)、汇总单位(报告单位)、各级统计部门与主管部门的范围等,一般是应执行国家统一规定的报表制度的企事业单位。因为我国经济类型的多样化,报表的实施范围要从实际出发,对所有制形式不同、规模大小不同、经营管理水平不同的企业,应根据不同情况区别对待。

(4) 报表的报送程序和报送日期。报表的报送要规定统计报表的报送程序,填报单位填报报表的份数、方式和受表单位,报送日期。

(5) 填表说明。填表说明的重点是指标解释,指标解释是对列入表的统计指标的含义、计算方法、计算价格、计量单位以及其他有关问题的具体说明,有时是对统计报表的统计范围、分类目录等作出的规定。

(6) 统计目录。统计目录是指统计报表中主栏项目的一览表,是主栏中应填报项目的分类。统计目录大体可分为两类:一类是主栏中填报的统计分组用的目录,如工业部门分类目录等;另一类是主栏中填报的具体项目的目录,如工业产品目录等。

统计报表制度是一个庞大的组织系统。它不仅要求各基层单位有完善的原始记录、统计台账和内部报表等良好的基础,而且要有一支熟悉业务的专业队伍,要很好地发挥统计报表制度的积极作用,必须严格按照统计法规办事,实行系统内的有效监督和管理。

报表制定时应坚持适用原则；精简原则即少而精，可要可不要的表式指标栏目，坚决不要；统一配套，并保持相对稳定；国家、部门、地方的统计报表制度必须适当分工，互相配合，尽可能减轻基层单位的填报负担。既要防止多、乱、滥发报表，又要防止虚报、瞒报和漏报，才能保证统计数字的质量，降低统计的社会成本。

统计报表是一种收集基本统计资料的基本的、传统的方法，由于存在内容较固定且中间环节较多，从经济和效率的角度考虑，在可能的情况下尽量减少统计报表，更多采用专门调查满足资料收集的需要。

2) 专门调查

专门调查是为研究和解决某种问题而临时组织的一次性调查。专门调查包括普查、重点调查、典型调查、抽样调查等种类，具有灵活多样、适应性强、因事制宜等优点。专门调查可以针对某专项内容进行，又可以补充统计报表的不足，应用的领域和范围很广，特别是社会经济运行的热点、难点问题；跟踪国家政策执行情况；各级党政领导关心、人民群众关注的问题、社会经济生活中出现的新情况和新问题；企事业单位和人民生活的需要等诸多方面。

(1) 普查。普查是普遍调查的简称，是为了某些特定目的而专门组织的一次性全面调查。普查主要用来调查不能够或不适宜用定期全面统计报表来收集的资料，作为了解和掌握基本国情国力资料的重要途径，在社会经济管理中占据重要地位。国家组织的农业普查、经济普查、人口普查按国家规定的统一组织、统一标准、统一方法、统一时间、统一表式和统一内容。普查数据一般较准确、规范化程度也较高，可以为抽样调查或其他调查提供基本依据。普查也有其局限性。普查要花费大量的人力、物力、财力和时间，因而间隔的时间通常较长，统计普查也不能经常进行。

普查一般适用于调查一定时点上的社会经济现象的总量，也可以调查某些时期现象的总量乃至一些并非总量的指标。我国的普查多半是在全国范围内进行的，而且所要收集的是经常的、定期的统计报表所不能提供的更为详细的资料，特别是重要的国情国力数据。目前我国实施周期性的普查制度，其明确规定人口普查和农业普查每10年进行一次，人口普查在逢0年份实施、农业普查在逢6年份实施；经济普查每5年进行一次，在逢3和8的年份实施。

普查是一次性调查，用来调查时点指标的，所取得的资料是反映事物或现象在一定时点上的总量和总体的各种构成情况的。由于普查工作量很大，涉及调查单位很广，调查项目较多，调查员队伍庞大，调查经费开支较大，国家普查需要差不多一年才能得到最终数据资料，因而根据实际需要和调查对象本身的特点，每隔较长一段时间才进行一次，一般这个间隔为5年或10年。普查是全面调查，对调查对象中包括的所有调查单位无一遗漏地进行调查，如经济普查就需对我国所有第二产业、第三产业的经济单位的发展规模、结构和效益等方面情况进行全面调查。普查数据标准化程度高，有着较高的准确性，可以为其他调查提供基本依据。

普查的组织方式一般有两种：一种是建立专门的普查机构普查。这种方式配备大量的专业普查人员对调查单位进行直接登记并开展普查工作。我国的人口普查、经济普查和农业普查从中央到地方，建立各级普查领导小组和工作小组，制订普查方案；组织普

查人员和物力财力, 对全国范围内的调查单位进行直接登记, 再汇总和公布资料, 最后总结普查工作。如我国第二次全国农业普查历时 3 年, 直接调查的原始指标 600 多个, 全国近 700 万普查指导员和普查员, 共调查了 2.3 亿普查户、40 万个农业生产经营单位、65 万多个村和 4 万多个乡镇, 填报普查表近 5 亿张。另一种以调查单位填报开展普查。它颁发调查表, 利用调查单位的原始记录和核算资料、结合盘点清查, 由报告单位填报, 如物资库存普查等。这种方式不设置专门的普查机构而利用基层单位本身的组织系统, 但需有专门的组织和人员从事该项普查工作。这种方式适用于内容比较单一、涉及范围较小的情况, 特别是为了满足某种紧迫需要而进行的“快速普查”。

普查时要注意: 一是规定统一的普查标准时点。规定统一的普查标准时点是指对被调查对象登记时所依据的统一时点。调查资料必须反映调查对象的这一时点上的状况或这个时点前的一段时间的状况, 以避免调查时因情况变动而产生重复登记或遗漏。二是规定统一的普查期限。普查期限是指普查中对调查对象进行登记的时间。在普查范围内各调查单位或调查人员应尽可能同时进行登记, 并在最短的期限内完成, 以便在方法和步调上保持一致, 保证资料的准确性和时效性。三是规定统一的普查项目和指标。普查时必须按照统一规定的项目和指标进行登记, 不准任意改变或增减, 以免影响汇总和综合, 降低资料质量。同一种普查, 每次调查的项目和指标应力求一致, 便于历次普查资料的对比分析和观察事物与现象发展变化情况。

(2) 重点调查。重点调查是指在调查对象中选择一部分重点单位进行调查, 以了解总体基本情况的一种非全面调查。重点调查的关键是选择重点单位。重点单位就是在总体中具有举足轻重的地位或单位数虽少但其标志值在总体标志值中占有很大比重的调查单位。通过对重点单位某种标志的观察和了解, 就可以从数量方面把握整个总体在该标志方面的基本情况。例如, 要掌握我国移动通信运营的基本情况, 只需对中国移动、中国电信和中国联通三个移动通信运营商进行调查就可以了, 它们占据我国移动通信的绝大部分业务。由于重点调查所涉及的调查单位较少, 项目较细, 内容相对较多, 可以节省较多的人力、物力、财力, 以较高的时效取得调查总体基本情况的资料。

重点调查一般有两种做法: 一是采用一次性调查方式, 对重点单位的某些标志值组织专门机构进行调查; 二是采用经常性调查方式, 向重点单位布置定期统计报表, 定期观察这些重点单位的主要指标完成情况及其变动。

重点单位的确定要考虑: 第一, 根据调查任务来确定重点单位, 不同的调查任务有不同的重点单位。一个调查单位只可能在一个或几个方面的标志值占有较大比重, 符合重点单位的条件, 只有选择其调查符合重点单位条件的方面。第二, 重点单位数量应根据调查总体的实际情况确定。选出的重点单位数应尽可能少些, 其标志值在总体标志值中所占比重应尽可能大些, 若调查单位标志值集中度高则调查单位数可以少一些, 反之则多一些。第三, 选中的单位应是管理健全、统计基础工作较好的单位, 以有利于统计调查的实施。第四, 适时调整重点单位。重点单位是可以变动的, 不同问题的重点调查或同一问题不同时间、不同方面内容的重点调查有不同的重点单位, 要随着情况的变化而调整重点单位。

重点调查适用于只要求掌握调查对象的基本情况, 总体中要有重点单位即总体中部

分单位能较集中地反映要研究的问题。如对我国石油行业生产情况进行调查,就可采用重点调查,因为我国石油产量比较集中在中石油、中石化和中海油三家生产单位。只有存在重点单位时,才采用重点调查了解总体基本情况;对于没有重点单位的总体现象的认识,就不能采取重点调查了。由于重点单位在全体调查对象中只占一小部分,调查的标志量在总体中却占绝大的比重,因面对这部分重点单位进行调查所取得的统计数据能够反映事物和现象发展变化的基本趋势,不能用以推断总体,因而也只是一补充性的调查方法。

(3) 典型调查。典型调查是根据调查的目的和要求,在对调查对象全面分析的基础上,有意识地选择部分有代表性或典型意义的单位进行调查,以认识同类事物和现象的发展变化规律及本质的一种调查方法。典型调查只调查部分有代表性的单位,属于非全面调查。

典型调查有以下几个特点:一是根据调查者对总体分析和判断来确定调查单位。典型单位是根据调查目的和任务有意识地选取的,受主观因素的影响很大,调查者对调查单位的了解情况、思想水平和判断能力对选择典型的代表性起着决定作用。二是典型调查主要是定性调查。它主要依靠调查者深入典型单位进行调查,对调查对象直接剖析,取得第一手资料,能够透过现象发现本质和发展规律,是一种以定性研究为主的调查方式。三是典型调查是一种由点及面,由个别到一般的认识方法。典型调查的方法论基础是辩证唯物论中关于从特殊到一般的认识原理,通过对少数典型单位的了解来指导一般,典型调查的结果可以用来大体上估计总体,但不能严格推及总体。四是典型调查是一种深入细致的调查研究。典型单位较少、设置指标可多一些、内容可细一些、使用调查工具不多,便于对事物和现象的状况深入分析与研究。运用灵活方便,节省人力、物力、财力和时间。

典型单位的选择方式主要包括以下三种:一是“解剖麻雀”。即对个别典型单位进行调查和研究,通过典型单位特征来说明事物的一般情况或事物发展的一般规律性。当调查单位之间情况差异较少,可选择若干个典型进行“解剖麻雀式”调查。二是划类选典。当调查对象的各单位间差异较大、涉及问题较复杂时,在了解总体大致情况的基础上把总体划分成若干类型,各类型单位按总体单位中所占比重从单一类型中选出若干典型单位进行调查,以从数量上推断整个总体。三是“抓两头”法。从调查目的和要求出发,分别从先进单位和落后单位中选择典型,以便总结成功经验和吸取落后教训,带动中间状态的单位,推动整体的发展。

选择典型单位是典型调查的首要问题,选择时要考虑以下方面:一是依据研究的目的和事物的特点选择典型。根据统计调查的目的和调查对象的特点选择典型,正确划分社会经济现象的类型和确定所选典型的类别。如果研究的目的在于探讨事物发展的一般规律,就应选取各方面发展较全面、较完善的单位作为典型。如在加强环保、美化城市方面,可选择全国优秀且有较成熟经验的大连、珠海等城市作为典型;如果要研究新生事物,应选择刚涌现的事物作为典型,如新一轮农村土地改革、产业发展智能化等问题;如研究问题的目的是表彰先进、树立榜样,当然要选择先进单位为典型。二是从事物的联系中去挑选典型。从事物的联系和共性出发,选择各类型单位中代表性强的单位

作为典型单位。如果是先进典型则优先考虑选择那些最有特色、最突出、最有借鉴意义的方面的单位；中间水平的典型单位可以考虑选择最普通、最一般、最有共性的对象作为典型单位；后进典型单位可能要有最值得改进或最需要吸取教训的方面的单位好一些。三是典型单位数目依据总体单位之间的差异程度确定。总体各单位之间差异程度大的，典型单位应该多选择一些，才能保证代表性；总体各单位之间差异程度小的，典型单位可以少选择一些。

只要客观地、正确地选择典型单位，通过对典型单位的深入细致的调查，既收集详细的第一手数字资料，又掌握生动具体的情况，就可以获得对总体本质特征的深刻认识，特别是对一些复杂的社会经济问题的研究，典型调查可以了解得更深入、更具体、更详尽。

典型调查是面对面的直接调查，主要依靠调查者深入基层与调查对象直接接触与剖析，通常有开调查会、个别访问、现场观察、查阅资料等典型调查的方法，其中开调查会是最常见的方法，这样才能对事物和现象的内部机制与变化过程了解得比较清楚、资料比较系统全面。

典型调查的适用有以下几个方面情形：一是研究新事物、了解新情况、发现新问题。改革和发展的年代，新生事物不断涌现、新情况和新问题不断产生，要全面了解它们的情况比较困难，通过对典型单位深入细致的调查，可以及时发现新情况、新问题，探测事物发展变化的趋势，形成科学的预见，并及时研究尚未充分发展、处于萌芽状况的新生事物，发现和解决出现的新问题或某种倾向性的社会问题。二是通过调查典型单位指导一般。由于典型调查的调查单位少，可以集中精力和时间深入调查典型单位，详细了解现象的具体情况，将事物的来龙去脉调查研究得清清楚楚，从而认识现象发展的一般规律，并以此指导和推动面上的工作。三是补充其他调查的不足。在总体内部差别不大，或分类后各类型内部差别不大的情况下，典型单位的代表性很显著，可用典型调查收集统计报表和其他专门调查无法取得的统计资料、收集不能用数字反映的各种情况、验证全面调查数字的真实性。四是一定条件下，可用典型调查资料结合基本统计数字估计总体指标数值。典型调查与全面统计相结合，可以补充全面数字的不足。全面资料只能反映总体事物的综合发展状况，为了更生动地说明问题，往往用典型调查所取得的具体资料加以补充。

（4）抽样调查。抽样调查的概念有广义和狭义之分。从广义上看，抽样调查就是非全面调查，它是指从调查对象中按一定的原则抽取部分单位作为样本进行观察研究，并根据样本结果来认识总体的一种调查方法。抽样调查按抽样方法不同，可分为随机抽样和非随机抽样两种。

随机抽样也叫概率抽样。一般所讲的抽样调查，大多数是指随机抽样，就是狭义的抽样调查，它是按照随机原则从调查对象中抽取部分单位进行调查，并根据所获得的样本数据对总体数量作出估计和推断的调查方法。随机原则是根据大数定律的要求，在抽取调查单位时，应保证总体中各个单位都有同样的机会被随机地抽中。

随机抽样具有以下几个基本特点：第一，按照随机原则抽选调查单位。样本单位的抽取不受任何主观因素及其他系统性因素的影响，总体的每个单位都有一定的机会被抽

选为样本单位。第二,对部分单位调查的目的是推断总体指标。抽样调查中的样本指标和对应的总体指标之间存在内在联系,提供了用实际调查部分信息对总体数量特征进行推断的科学方法,用样本统计量对总体参数进行估计。第三,抽样误差可以事先计算并加以控制。以样本资料对总体数量特征进行推断,不可避免会产生代表性误差,但抽样调查的代表性误差是可以根据有关资料事先计算并进行控制的,故可以保证推断结果达到预期的可靠程度。但随机抽样比较复杂,对调查人员的专业技术要求高,调查中需要抽样框,而构建和维护一个高质量的抽样框费用很高,抽样单位可能非常分散,而且不能轻易更换样本单位,增加了调查费用。

同全面调查相比较,随机抽样有以下优点:第一,随机抽样节约费用、快速及时,而且比较灵活。随机抽样的调查单位少,收集整理和汇总工作量少,尤其对急需的有关信息可以及时地获得结果,有时可以根据需要增加调查内容。第二,随机抽样能够提高资料的准确性和可靠性。抽样误差通过抽样单位数和抽样组织形式来调节与控制。全面调查只有工作误差而没有代表性误差,而随机抽样则两种误差全有,在全面调查过程控制得好时,随机抽样相对不准确。但是不能忽略了随机抽样和全面调查中两种误差的大小,全面调查的调查单位多、涉及面广,参加人员多、水平不齐,发生工作误差的可能性就大;随机抽样的调查单位少、参加人员少,可以严格培训,因而发生工作误差的可能性小,因而随机抽样有可能比全面调查的结果更准确。第三,随机抽样和全面调查是不能互相代替的。随机抽样只能提供说明整个总体情况的统计资料,而不能提供总体各层级状况中各个组成部分的资料,难以满足各级管理部门的要求。应根据资料的性质和调查对象的不同,采用不同的统计调查形式。一些重要的国情、国力的统计资料,必须组织全面调查。很多国家的人口普查与人口随机抽样同时进行,一些项目通过普查取得资料,另一些项目通过随机抽样取得资料,这样既节省了调查费用和时间,又丰富了调查内容。

随机抽样具有以下几个作用:第一,不可能进行全面调查的场合而又需要了解全面资料,必须采用随机抽样的方法。调查所使用的手段具有破坏性、损伤性或消耗性,如产品质量检验、食品防腐期限试验等都不可能将所有产品进行检查和实验;对无限总体或总体规模非常大的场合进行调查,如要测定湖水的含氧量、调查大气的污染程度、研究太空中行星运动规律等只能采用随机抽样的方法。第二,没有必要进行全面调查但需要掌握总体全面情况。实际上没有必要或很难办到要采用随机抽样,如要了解全国城乡人民的家庭生活状况,从理论上讲可以挨门逐户进行全面调查,但是调查范围太大、调查单位太多,实际上难以办到、也没有必要。第三,随机抽样的结果可以对全面调查的结果进行检查和修正。全面调查涉及面广、工作量大、参加人员多,容易产生漏查、重查、填错、算错等调查性差错。而随机抽样单位少、有准确度高及抽样误差可以控制等优点,在全面调查之后进行抽样复查,根据抽查结果计算差错率,并以此为依据检查和修正全面调查结果。第四,用于资料时效性要求很强的场合。例如,大量农产品收割前一般要进行产量预计,以便及时地组织仓储和运输设备;对工厂连续大量生产的产品质量检验和控制等都对调查结果有很强的及时性要求,一般均需采用抽样方法。第五,随机抽样可以用于工业生产过程的质量控制。在工业产品成批或大量连续生产过程中,

利用随机抽样可以检验生产过程是否正常,进行质量控制,保证生产质量稳定。便于采取措施,预防废品的发生。

随机抽样包括简单随机抽样、类型抽样、等距抽样、整群抽样和多级抽样等形式。简单随机抽样又称纯随机抽样,是事前对总体数量不作任何分组排列,完全凭偶然的机遇从总体中抽取样本进行调查的方法。简单随机抽样一般可采用直接抽样、抽签、摇号或查随机数表等方法抽取样本。采用这种抽样方式比较适合于总体单位之间差异较小的状况。等距抽样亦称机械抽样或系统抽样,它要求先将总体各个单位按照空间、时间或某些与调查无关的标志排列起来,然后等间隔地依次抽取样本单位。抽样间隔则等于总体单位数除去样本数所得的商。这种抽样方法在被调查总体数量较多时更为方便。类型抽样亦称分层抽样或分类抽样,先将总体单位按其差异程度或某一特征分类、分层,然后在各类或每层中随机抽取样本单位。它实际上是科学分组、分类与随机原则的结合,适用于总体量大、差异程度较大的情况。分层抽样有等比抽样和不等比抽样之分,当总数各类差别过大时,可采用不等比抽样。除了分层或分类外,其组织方式与简单随机抽样和等距抽样相同。整群抽样是按照某一标准将总体单位分成“群”或“组”,并从中抽选,然后把被抽出的“群”或“组”的所有单位都定为样本单位,最后利用抽中“群”或“组”的调查结果推断总体。抽取“群”或“组”可以采用随机方式或分类方式,也可以采用等距方式来确定;“群”或“组”内的调查采用普查方式进行。前四种抽样方法均为一次性直接从总体中抽出样本,称为单阶段抽样。还有多阶段抽样,又称多级抽样,它是将抽样过程分为几个阶段,结合使用上述方法中的两种或数种。

非随机抽样又称非概率抽样,是指调查者根据自己的认识和判断,从调查对象中有意识地抽取部分单位作为样本,进行观察研究,以认识总体的统计调查方法。一是方便抽样。方便抽样就是调查过程中由调查人员依据方便原则,自行确定样本单位的调查方法。其样本往往限于总体中易于抽到的一部分,最常见的是偶遇抽样,研究者将在某一时间和环境中所遇到的每一总体单位均作为样本成员。如调查者在街头、公园、商店等公共场所进行拦截调查,厂家在出售产品柜台前对路过顾客进行的调查。方便抽样具有容易实施、节约时间、调查成本低等优点;但是样本单位的确定带有随意性,样本无法代表有明确定义的总体,样本代表性因受偶然因素的影响太大而得不到保证,调查结果不宜推断总体。二是自愿抽样。自愿抽样是指公开征求自愿接受调查或实验对象的抽样方法。被调查者自愿参加、成为样本中的一分子,向调查人员提供有关信息,如参与媒体调查问卷活动等属于自愿抽样。自愿抽样的样本只能反映特定人群的情况,是有偏颇的,不能依据样本的信息推断总体。实际中可能出现自愿者的教育程度或社会经济地位比非自愿者高、女性多于男性的情况。三是判断抽样。判断抽样又称立意抽样,研究人员根据研究的目的和对研究对象的了解,按照自己的经验、判断,从总体中选择最能代表总体的单位作为样本的抽样方法。判断抽样有重点抽样、典型抽样等方式。判断抽样是主观的,样本选择的好坏取决于调研者的判断、经验、专业程度和创造性。当研究者对自己的研究领域十分熟悉,对研究总体比较了解时采用这种抽样方法,可获代表性较高的样本。这种抽样方法多应用于总体小而内部差异大的情况,以及在总体边界无法确定或因研究者的时间与人力、物力有限时采用,具有抽样成本比较低、容易操作的特点;

但样本是人为确定的、没有依据随机原则,调查结果不能用于推断总体。四是滚雪球抽样。根据研究的目的和要求,先选择一组具有所需特征的调查单位并对其调查,依靠它们提供符合要求的调查对象……依此类推,如同滚雪球般将调查过程持续进行。滚雪球抽样多用于总体单位的信息不足或观察性研究的情况,特别是稀少群体和特定群体研究。做到容易找到那些属于特定群体的被调查者,调查的成本也比较低。五是配额抽样。配额抽样也称定额抽样,先将总体按一定的标志分为若干类(层或群),按照各类型样本数与总体单位数成比例的原则主观抽取样本。配额抽样操作简单,节省费用,选择过程短,能够较快地取得调查结果,可以保证总体中不同类别的单位都包括在样本中,并使得样本结构和总体结构类似。配额抽样类似于分层抽样,有两点重要区别:首先,配额抽样的被调查者不是按随机原则抽选出来的,而分层抽样必须遵守随机抽样的原则;其次,在分层抽样中用于分类的指标应联系研究目标选择,而配额抽样无此要求。

非随机抽样比随机抽样容易实施、成本低、省时间,可根据样本调查结果,从一定程度上说明总体的性质、特征。但由于无法排除抽样者的主观性,不知道也无法控制和客观地测量样本的代表性,不能估计出抽样误差,非随机抽样的结果不能也不应该推算总体,也无法正确地说明样本统计值在多大程度上适合于总体。非随机抽样对精确性要求不严格的调查有相当大的吸引力,多用于探索性研究和预备性研究以及总体边界不清、难以实施随机抽样的研究。在实际应用中,非随机抽样往往与随机抽样结合使用,以取得更准确的结果。

2.2.3 统计数据的收集方法

任何一种调查都必须采用一定的调查方法去收集原始资料,即使调查的组织形式相同,其调查方法也可以是不同的。应根据调查目的与被调查对象的具体特点,选择合适的调查方法。

根据调查对象、调查单位的特点和调查内容不同,统计数据的收集方法分为以下几种。

1. 实验调查法

实验调查法也称试验调查法,就是通过某种实验或者实践活动的验证去收集有关资料的一种调查方法。实验调查法通过实验方法得到的数据,就是实验者按照一定实验假设,通过改变某些实验环境的实践活动来认识实验对象的本质及其发展规律的调查。将一组随机抽取的实验对象随机分配到两种或多种处理组,观察比较不同处理的效应,这种研究称为实验新方法或新方案使用效果的数据资料。实验设计能有效地控制误差,节省人力、财力、物力,提高效率,但须遵循以下四个基本原则。

(1) 对照原则。对照原则即实验要设立对照,除实验因素外,对照组与实验组其余因素保持一致。

(2) 重复原则。重复原则即研究对象要有一定的数量,或者说样本含量应足够,根据每个具体研究,可有不同的方法来进行样本含量估计。

(3) 随机原则。随机原则即应保证每个实验对象有同等机会进入实验或接受某种处理。

(4) 均衡原则。均衡原则即各组的非实验因素条件基本一致,以消除其影响。

实验调查法通常是对自然现象而言的,现在广泛运用到社会科学中。社会经济现象采用实验调查法是以社会为“实验室”,以实践为基础。常用于收集测试某一新产品、新工艺或新方法使用效果的资料的方法。如了解某新产品的质量和提高产品质量应采取的措施,可免费将新产品赠送给用户或消费者个人试用,以征求用户或消费者个人的意见;了解某新产品的销路和顾客对产品质量与价格的意见,可举办新产品展销会;为进行某项改革和推行某项新政策,先在小范围内进行试验,以了解改革的新政策的社会效应。

实验法要注意以下几个问题。

(1) 人的意愿。研究的对象是人的时候,在划分实验组和对照组时的随机原则将面临挑战。

(2) 心理问题。人们对被研究非常敏感,这使得他们更加注意自我,从而走到事物的另一个极端。

(3) 道德问题。当某种实验涉及道德问题时,人们会处于进退两难的尴尬境地。

2. 直接观察法

直接观察法又称现场观察法,是由调查人员到现场对调查对象进行直接点数和计量的一种调查方法。如要了解商场中商品库存量,调查员亲自到商场和仓库中现场观察、点数、计量;要测量农作物的产量,调查员到地块中进行实割实测;等等。这种方法能够保证所收集的统计资料的准确性,但经济性差,要花费较多的人力、物力和时间。调查机关派出调查人员到调查所涉及的事物所在的场所进行现场观察、点数、计量或用仪器测量和记录现场情况的调查方法。

直接观察法适用于非全面调查、收集表象资料、了解现实情况。由于受到观测、计量等方面能力和手段的限制,只能观察到表面现象或现场实际数量,无法了解被观察对象的内部状态,如对企业生产经营状况进行调查,如果仅通过现场观察企业的生产或销售等活动,可以统计当时发生的生产和销售等现状数据,但无法了解生产者、中间商、消费者等各种被观察者的心理活动,也无法对历史情况进行研究。

3. 采访法

采访法是指调查人员向被调查者提出问题,根据被调查者的回答以取得资料的调查方法,可根据具体情况,采用个别询问或以座谈会的形式收集资料。采访法分为口头访问、书面访问、个人访问、集体访问等形式。

现实中较常见的采访法有问卷调查法、座谈会和深度访问等。

1) 问卷调查法

问卷调查法也称问卷法,是采访法中的书面访问和集体访问形式的结合,是调查者运用统一设计的问卷向被选取的调查对象了解情况或征询意见的调查方法。它是为了特定的目的,选择合适的目标对象群,以问卷形式提问、由被调查者自愿自由回答的一种采集资料的方法。调查问卷由一连串写好的小问题组成,编制成书面的问题表格,收集

被访问者的意见、感受、反应及对知识的认识等,然后收回整理分析,从而得出结论。

2) 座谈会

座谈会也称集体访谈法,就是将一组被调查者集中在调查现场,让他们对调查的主题发表意见,从而获取资料的方法。参加座谈会的被调查者应是所调查问题的专家或有经验者,人数不宜太多,通常为6~10人,研究人员应对被调查者进行严格的甄别、筛选。讨论方式主要看主持人的习惯和爱好。这种方法能获取其他方法无法取得的资料,因为在彼此交流的环境里,被调查者相互影响、启发、补充,不断修正自己的观点,这就有利于研究者从中获得较为广泛深入的想法和意见,而且座谈会不会因为问卷过长而遭到拒访。

3) 深度访问

深度访问是一种一次只要一名被调查者参加的特殊的定性研究,简称“深访”。“深访”暗示着要不断深入到被调查者的思想中,努力发掘其行为的真实动机。“深访”是一种无结构的个人访问,调查者运用大量的追问技巧,尽可能让被调查者自由发挥,表达他的想法和感受。“深访”常用于动机研究,如消费者购买某种产品的动机等,以发掘被调查者非表面化的深层意见;最适用于研究个人隐私问题、敏感问题或政治性问题等。对于那些不同人之间观点差异极大的问题,用小组讨论可能会把问题弄糟,也可采用“深访”。

采访法在实施中,调查人员和被调查者直接交谈,调查人员可向被调查者说明调查的目的和要求,从而打消被调查者不必要的顾虑,也可当场解释调查中出现的问题、解答被调查者的各种疑问;在集体访问中,调查人员与被调查者可以相互讨论,相互补充,从而可保证收集资料的质量。但是采访法同时也具有耗费过大的缺点,在调查经费紧张时不宜采用这种方法。

采访法顺利实施需具备一定的条件,主要包括能激发被调查者的兴趣,愿意给予支持和合作;问题明确,易于回答;对调查结果不要求十分精确;不需要所有的调查对象都给答案;不属于个人保密情况。

4. 通信法

通信法指调查者利用网络、电视、电话、移动通信、信函、传真等通信中介手段将调查表格传递给被调查者,并请被调查者按调查内容反馈调查者的一种调查方法。常见的几种通信法有以下几种。

1) 网络调查

网络调查是一种利用互联网作为媒介的调查方式。将问卷在网上发布,被调查对象通过 Internet 完成调查。网上问卷调查一般有三种途径:第一种是将问卷放置在 WWW 站点上,等待访问者访问时填写问卷。第二种是通过 E-mail 方式将问卷发送给被调查者,被调查者完成后将结果通过 E-mail 返回;第三种是主题调查法,通过视频会议、新闻组、BBS、聊天室进行。它充分利用了互联网的速度快、互动性、实时性、方便性、调查空间大、调研群体大、费用低廉、调研内容设置灵活、易获得连续性数据等优点。局限性在于调查对象只能是互联网用户,代表性、保密性和可信度都较低。

2) 电话调查

电话调查是调查人员利用电话同被调查者进行语言交流,从而获得信息的一种调查方法。该方法具有时效快、费用低等特点。随着电话的普及,电话调查也越来越广泛。电话调查可以按照事先设计好的问卷进行,也可以针对某一专门问题进行电话采访。电话调查所提问题要明确,且数量不宜过多。

3) 短信调查法

短信调查法是调查者以短信的方式向被调查者提出问题,通过被调查者自愿回答来获得信息的一种调查方法。短信调查往往借助短信调查系统,调查者通过调查系统设置调查内容、调查流程以及调查过程中对被调查者进行自动引导,能根据被调查者的设置自动分析调查结果,等等。短信调查系统有两种调查方式:一种是对指定调查对象进行调查,调查系统根据调查者设置向预先拟定好的调查对象发送调查内容,根据被调查者的回复内容引导调查者完成整个调查过程。另一种是对所有自愿接受调查的调查对象进行调查,调查系统根据调查者设置接收所有被调查者的调查请求,将引导调查者完成整个调查过程。

4) 邮寄调查

邮寄调查是通过邮寄、宣传媒体和专门场所等将调查表或问卷送至被调查者手中,由被调查者填写,然后将调查表寄回或投放到收集点的一种调查方法。这是一种标准化调查,其特点是,调查人员和受调查者没有直接的语言交流,信息的传递完全依赖于调查表。邮寄调查在统计部门进行的统计报表及市场调查机构进行的问卷调查中经常使用。

为了提高效率,通过短信群发通知调查对象、宣传调查内容;为了提高被调查者参与的积极性,往往设计抽奖环节。短信调查具有便宜、方便、准确等优势,越来越被众多调查者,特别是企事业单位所采用。

采用通信中介调查,较派员采访节省人力、物力、财力和时间,成本较低,但是由于通信中介有覆盖范围和使用人群的限制,可能造成部分本来应该成为调查对象的不可能被调查,从而造成调查单位的代表性不够,影响调查结果的准确性和完整性;调查者和被调查者不直接接触、调查的实施没有强制性,往往反馈率较低、疑难问题回答的正确性不高。

5. 报告法

报告法是被调查者按调查机构统一颁发的调查方案,按一定的报送程序、报送时间向调查者报告有关资料的调查方法。我国现行的统计报表制度就是采用报告法收集统计资料的,这种调查方法是各地区、各部门、各单位按照有关法规的规定,必须对国家履行的一种义务。报告法的特点是有统一项目、统一表格、统一要求和统一上报程序与上报时间,其资料来源于原始记录,可以同时进行大量的调查。在报告系统健全、原始记录和核算工作系统完整的情况下,报告法可以保证所提供资料的准确性和及时性。

报告法的实施必须带有强制性,即被调查者必须按规定准确及时地向有行政权力和法律手段作支持的调查者报告统计资料;报告法通常适用于政府部门,报告单位一般是

机关团体或企事业单位,而不是个人;相对于直接观察法,报告法具有经济性和及时性,能显著节省人力、物力、财力和时间。

由于报告法采用的是调查者和被调查者不直接接触,因此方案要简明准确,通俗易懂,以防误解而影响统计质量。

2.3 统计调查方案与调查问卷的设计

2.3.1 统计调查方案的设计

统计调查是一项非常强的理论与实践紧密结合的系统性工作,一项大型的调查往往需要成千上万的人协同工作,可能涉及国民经济和社会发展的方方面面,为了使整个统计调查统一认识、统一计划、统一标准、统一组织、统一步骤、有条不紊地进行,圆满完成调查任务,事先制订一个科学周密的统计调查方案是不可缺少的前提条件。

统计调查方案是指导统计调查的纲领性文件,而正确地制订统计调查方案是保证统计调查有计划、有组织进行的首要步骤,是保证统计调查顺利进行的前提,也是准确、及时取得统计资料的重要条件。

一个完整的统计调查方案应包括以下几个方面。

1. 调查目的

明确调查目的是制订统计调查方案的首要问题,调查目的决定着调查对象、调查内容和调查方法。调查目的首先确定了“为什么要进行调查”的问题,明确这次调查要了解什么问题、解决什么问题 and 要取得什么资料等根本问题。只有在此基础上,才能进一步确定向谁调查,调查什么及采用什么方法调查,才能进一步明确具体的调查任务。如果调查目的不明,就会使整个调查工作陷入盲目状态,最终酿成整个调查工作的失败。

2. 调查对象、调查单位和报告单位

统计调查的目的确定以后,就可以进一步确定调查对象、调查单位和报告单位,回答向谁调查、由谁来具体提供资料的问题。

调查对象就是根据调查目的所确定的调查研究的事物的总体,它由性质相同的许多调查单位所组成,确定了调查对象也就明确了调查范围。调查单位是构成调查对象的各个单位,是调查资料的承担者,是在调查的对象中所要调查的具体单位,确定了调查单位,则使我们知道从哪里取得有关标志的情况和资料,即解决了向谁调查和由谁来提供统计资料的问题。调查单位可以是一个企业、一个事业单位,也可以是一个人,甚至是一件产品。报告单位也称填报单位,是负责向上报送调查内容、提交统计资料的单位。例如,要对中国大学教育状况进行调查,那么中国所有大学就构成调查对象,即调查总体,而中国的每一所大学,都是调查单位。若调查目的是了解全国大学生的学习和生活情况,则全国所有大学的大学生就是调查对象,而每一个大学生就是调查单位。

调查执行时,应注意理解调查单位和报告单位的关系。报告单位一般是在行政上、经济上具有一定独立性的单位,而调查单位可以是人、企事业单位,也可以是物。根据调查目的,调查单位与报告单位有时一致,有时不一致。如经济普查,每个经济单位既是调查单位又是报告单位;而企业设备普查,调查单位是企业的每台生产设备,而报告单位是每个企业。而在科技人员调查中,调查单位和填报单位就不一致,前者是每一个科技人员,而后者则是科技人员所在的工作单位。

3. 调查项目和调查表

调查项目即调查的内容,就是调查中所要登记的调查单位的特征名称,由调查单位的一系列品质标志和数量标志构成。确定调查项目所要解决的问题是:向调查单位调查什么?反映调查单位特征的标志是多种多样的,如要反映国有企业的生产经营情况,就需收集每个被调查企业的职工人数、职工工资、工业总产值、工业增加值、销售额、利润、成本、劳动生产率、税收等方面的相关资料。

调查中选择哪些调查项目可依据事先设计好的统计指标体系,根据调查目的和调查单位的特点来确定。确定调查项目时应注意以下几点。

(1) 项目的出发点,从事物和现象之间的相互联系,从现象的过去、现在和发展等方面出发。

(2) 项目的必需性,即根据调查目的所必需的项目。

(3) 取得资料的可能性,即只提出能够取得确切资料的项目。

(4) 项目之间的衔接性,即项目之间尽可能做到互相联系,便于核对。

(5) 项目解释的一致性,即调查的项目含义明确,不要模棱两可和引起歧义,项目动态可比性。

调查表就是承载调查项目的统计表格,是将确定的调查项目按一定顺序列示出来的表格,它是统计工作中收集统计资料的基本工具。调查表是调查方案的核心部分,调查项目应科学地分类,按照一定的逻辑顺序和技术要求在调查表中合理地排列。一份调查表能将调查资料清晰、有条理地表示出来,不仅便于收集和登记资料,而且便于调查后对统计资料的汇总、整理和分析。

调查表一般由表头、表体和表脚三部分内容组成:表头在调查表的上方居中,用来标明调查表的名称;左上填写调查单位的名称、性质、隶属关系等项;右上注明表号、制表单位、批准备案文号等。表体是调查表的主要部分,包括事物和现象的调查项目的具体内容及数字、计量单位等具体表现;表脚包括调查者和单位负责人的签名与调查日期等内容,以便明确责任,发现问题时便于查询。

调查表的形式一般有两种:单一表和一览表。单一表是在一张表上只登记一个调查单位的调查资料(表2-1),它可以容纳较多的调查项目,便于整理、分类,但内容烦琐,适用于调查项目多、调查单位少且较详细的统计调查。一览表是在一张表上登记若干个调查单位的调查资料(表2-2),它的调查项目不宜过多,一览表的使用节省人力、物力,而且一目了然,便于合计和核对差错,它适用于调查项目较少、调查单位较多时。

表 2-1 企业职工状况调查表

企业名称：企业地址：

企业性质：1. 国有 2. 集体 3. 私有 4. 外资 5. 港澳台经济 6. 其他					
企业规模：1. 大型 2. 中型 3. 小型 4. 微型					
企业盈亏状况：1. 盈 2. 亏			上年盈亏状况：1. 盈 2. 亏		

职工状况

职工总数__人	
其中：男职工__人	女职工__人

年龄构成：

20 岁以下的__人
20~30 岁的__人
30~40 岁的__人
40~50 岁的__人
50 岁以上的__人

受教育程度：

小学及以下__人
中 学__人
专 科__人
本 科__人
研究生及以上__人

企业负责人签字：填表人签字：填表日期：

表 2-2 ××市 20××年工业企业情况调查表

企业名称	企业代码	职工人数	资产总额	销售收入	利润	税金

单位负责人：填表人：填表日期：

4. 调查方式和调查方法

调查方式是指调查的组织形式，调查方法是指收集资料的具体方法的选择。调查的方式、方法包含多种，它们各自都有自己的特点和适用条件。实际工作中，在选择调查的方式、方法时，要根据统计调查的内容和调查对象的特点，结合方式方法的优缺点考虑，不论采用哪种方式组织调查，都要运用具体的调查方法进行统计资料的采集，可以只用某一种方式、方法或若干种方式、方法结合使用。我国经济普查规定，对法人单位、产业活动单位采用普查的方式，对个体经营户采用普查辅以典型调查的方式，具体收集数据一律采用访问。

5. 调查时间和调查地点

调查时间包括三个方面的含义。

（1）调查资料所属的时间。如果所调查的是时期现象，就要明确规定反映的调查对象从何年何月何日起到何年何月何日止的资料，如调查某企业工业总产值，就应规定工业总产值的起止日期，是某一年的，或是某一季度的，等等；如果所要调查的是时点现

象,就要明确规定统一的标准时点。如我国第六次人口普查规定的标准时点,即 2010 年 11 月 1 日零时(或 10 月 31 日 24 时)。

(2) 调查工作进行的时间。调查工作进行的时间即指对调查单位的标志进行登记的时间。人口普查中规定入户登记的时间为 15 天。

(3) 调查期限。调查期限即整个调查工作的期限,包括从调查准备工作开始,到收集资料,报送资料,到数据完全公布为止的整个调查工作所需的时间,2010 年全国第六次人口普查,从 2010 年 11 月开始到 2011 年 4 月 29 日发布第二号人口普查统计公报为止,历时将近 6 个月,而实际普查开始准备的时间更早。实践中,为了保证调查工作如期完成,人们还往往规定不同调查阶段的工作时间。为了保证资料时效性,对调查期限的规定,要尽可能短。

调查地点是指取得资料的合适地点。不同的场合会有不同的人群,不同的人群在不同的地点更容易得到被调查者的配合和支持,以便更好地完成调查任务,获得更加准确的资料。

6. 调查工作的组织实施计划

为了保证整个统计调查工作顺利进行,需要有一套严密细致的工作组织系统与之匹配,在调查方案中还应该有一个考虑周密的组织实施计划。其主要内容应包括以下几个方面。

(1) 成立调查领导机构和办事机构。我国历次人口普查、经济普查、工业普查、农业普查等,都成立了专门的领导机构,负责组织和协调整个统计调查工作。

(2) 配备调查人员。通常统计调查业务人员是从统计业务部门抽调的,他们是统计调查的业务指导员,对统计调查的成败有至关重要的影响作用,同时还要通过他们培训调查员。

(3) 调查宣传。调查宣传即对调查意义的宣传,其目的在于使更多的人了解和认识开展调查的目的、任务和意义,取得调查单位的理解和支持,对统计调查来说是很重要的。

(4) 相关人员的培训。一项大规模的统计调查单靠统计部门是难以完成的,需要动员社会各界人士的参与。例如,2010 年我国第六次人口普查仅普查员就 700 多万人,这些业务人员的素质直接影响统计源头的信息质量,因此,对调查员的培训是必不可少的。

(5) 各种文件及调查表格的印刷、调查经费的筹集和管理、调查方案的传达布置、试点及其他工作、调查结果公布时间等。

2.3.2 统计调查问卷的设计

问卷是指为了达到调查的目的和收集必要数据事先设计好的询问提纲或一系列问题表,是最常用的一种调查工具。调查问卷是指根据调查的目的和任务,采用提问的形式系统地记载调查内容,并由被调查者回答的一种问题清单或表格,又称调查表或问卷表。它既是一种收集数据的结构化技术,又是实施各种调查方法在收集资料时的一种必备的工具,可以是表格式、卡片式或簿记式等形式。用它可以反映人们的行为、行为结

果和对某项事物的认知程度、态度、基本意向等,从而为定量分析提供依据。完美的问卷必须具备两个功能,即能将问题传达给被问的人和使被问者乐于回答。要完成这两个功能,在问卷设计时应当遵循一定的原则和程序,运用一定的技巧。

1. 调查问卷的主要类型

调查问卷的主要类型有以下几种。

1) 从调查问卷的内容看,调查问卷可分为甄别问卷、主体问卷和背景问卷

甄别问卷是为过滤调查对象而专门设计的问卷。甄别问卷主要用来将不符合项目访问要求的被调查者剔除,找出真正符合项目要求的、合格的被调查者。项目不同,甄别的内容也有所不同,可根据实际要求进行取舍,通常包括被调查者所在行业的要求、年龄、收入、职务、决策权等要求方面的内容。通常甄别问卷会在一个完整问卷的开头部分,对甄别不合格的被调查者将在这一部分终止访问。如果没有经过甄别而直接开始问卷调查,那么很有可能得出的结果是毫无意义的。

主体问卷是指问卷中所设计的问题必须能表达调查主要内容的问卷;主体问卷是整个问卷的核心,其内容因项目不同而各异,需要根据具体要求来设计这部分问题。一般来讲,先是一般性问题,是被调查者容易回答和较为关心的问题,以引起被调查者的兴趣;中间的问题是需要被调查者思考的、开放性的问题;最后是专业性强、敏感的、较难回答的问题。

背景问卷主要用来了解被调查者个人或其家庭的一些基本情况,以帮助后期能够分析不同类型被调查者的情况有何不同。背景问卷通常包括被调查者的职业、年龄、受教育程度、职位、家庭成员数量、个人及家庭收入等方面内容。

2) 从调查问卷的作答者看,调查问卷可分为自填式问卷和访问式问卷

自填式问卷是由调查者发给(或邮寄或网上提供)被调查者,由被调查者自行填写、作答的问卷。访问式问卷是由调查者按照统一设计的问卷,向被调查者当面或电话提出问题,然后再由调查者根据被调查者的口头回答来填写的问卷。

3) 从调查问卷的结构看,调查问卷可分为结构式问卷和无结构式问卷

结构式问卷又称标准式问卷,是按照调查目的和内容精心设计的具有结构的问卷。问卷中的问题是按一定的提问方式和顺序进行安排的,对问卷中的问题和顺序,调查者是不能随意变动的。结构式问卷根据答案形式分为封闭式问卷、开放式问卷和半封闭式问卷三种。

(1) 封闭式问卷。封闭式问卷是指对提出的每一个问题都给出了明确的答案,被调查者只能从已给的备选答案中进行选择的问卷。封闭式调查问卷的优点是便于综合,缺点是有时答案可能不全。

(2) 开放式问卷。开放式问卷又称自由问卷,是指只提出问题,不提供任何可能答案,由被调查者自由回答的问卷。第一种开放式问卷不限制被调查者如何回答,被调查者可以在问题范围内说出结果;第二种开放式问卷是在前一问题的基础上,加了个尾巴“为什么?”,调查者希望获得被调查者作出前述回答的原因。使用开放式问卷的优点在于可以使调查得到比较符合被调查者实际的答案,缺点是有时意见比较分散,难以综合。

开放式问卷对探索性研究有许多好处：如提供行为的方向、问题的焦点、主要价值观念；有的问题还不是十分肯定，需先做探索。

(3) 半封闭式问卷。半封闭式问卷也称半开放式问卷，是指封闭式与开放式相结合的问卷。其有两种形式：一种是在一个问题中，除给出一定的备选答案供选择外，还相应地列出一个或若干个开放式问题以便回答；另一种是问卷的一部分问题采用封闭式，另一部分采用开放式。

无结构式问卷是指问卷中的问题没有在组织中加以严格设计和安排，只是围绕研究目的提出问题，调查者在实施调查时，可根据实际情况适当变动问题和顺序。无结构式问卷的特点在于被调查者可以自由回答，毫无限制；调查者在不变更内容与方向的前提下，可以有相当程度的自由运用。如访问某组织的成员时可以随情况而改变问法，问甲时可以说：目的是什么？问乙时可以说：当初为了什么才组成这个组织？无结构式问卷多半用在深入访问的场合，适合被调查者人数较少、不必将资料量化、必须向有关人士问差不多相同的问题；无结构问卷比较适合于小样本，但对大样本研究也能产生辅助作用，把研究中的几个主要问题提出来讨论，然后与大样本问卷数据相互印证；对于集体研究，临时发现了重大的问题，值得进一步追究。

2. 调查问卷的设计原则

在问卷调查之前，一项重要的工作内容就是设计调查问卷，它是问卷调查中最基本的工作，问卷调查的成功与失败，能否达到预期目的，在很大程度上取决于调查问卷的质量。调查问卷的设计要遵循以下几个原则。

1) 目的原则

目的原则是调查问卷设计最重要的、首先遵守的原则。问卷的主要目的是提供需要管理的信息。调查问卷设计服从于调查研究的目的和要求，所有项目必须与调查主题紧密相关，与调查主题无关的、可要可不要的项目都不要列入调查问卷，不设容易引起被调查者反感或疑惑而不能获得答案的项目；必须透彻了解调查项目的主题，能拟出可从被调查者得到最充分资料的项目，以满足调查机构对资料的需求，做到既不遗漏一个项目以致需要的信息资料残缺不全，也不浪费一个项目去取得不需要的信息资料。

2) 接受原则

接受原则就是被调查者能够接受调查问卷。这是设计调查问卷必须考虑的因素，是获得被调查者支持的关键。要求做到调查问卷提问自然、用词准确、语气亲切、通俗易懂、适合被调查者身份、易为被调查者接受与合作，对敏感性问题采取一定的技巧调查，使问卷具有合理性和可答性，避免主观性和暗示性，以免资料信息失真。

3) 简明原则

简明原则就是整体设计问卷、问题要简单明了。它是评价问卷质量的重要因素，包括内容精简扼要；语言简明、准确，而不要模棱两可，以避免理解和答问的错误，影响资料可信度；等等。应该控制问卷的长度。回答问卷的时间应尽量控制在 20 分钟以内，过长的问卷会影响被调查者的参与积极性和回答问题的准确性，最终会降低调查结果的可信度。

4) 适当原则

适当原则包括问题适当、调查对象适当等。问题适当就是选择的问题要与研究假设相符,即所选择的问题是针对研究假设的,是研究假设合理的内涵和外延;所选问题在数量上要适当,这可根据一般经验和预试的结果来确定;问题中的调查指标要求适当,调查指标是能够累加和便于累加的、指标的累计与相对数的计算是有意义的、能够通过数据清楚地说明所要调查的问题,这样就能够做到调查结果的容易得出和调查结果的说服力,也便于调查后的整理与分析工作。调查对象适当是指选择的对象要符合对研究假设的推论;问卷的结构和形式、问题的形式及用语都要考虑调查对象的适当性。

5) 排序原则

排序原则是指调查问卷的问题必须按照合理的顺序排序。问题排序是否恰当是问卷能否激发被调查者兴趣的关键。问题排序根据项目之间的联系,考虑内容、类别、时间、逻辑等方面的要求,合理设置各问题间的次序和关系。符合应答者的思维程序和逻辑,一般是先易后难、先简后繁、先具体后抽象,确保所取得的信息资料正确无误。

6) 完整性原则

完整性原则要求确保问卷题目的完整性和选项的完整性。调查问卷中题目提供的选择答案应在逻辑上是排他的,在可能性上又是穷尽的,既不浪费一个问句,也不遗漏一个问句,确保准确全面地反映研究的问题。不同的调查问卷的问题内容差异很大,一般的调查问卷的主体内容通常包括以下三个方面:一是人们的事实与行为,包括对被调查者本人的事实与行为或通过被调查者了解他人的事实与行为。如对消费者的消费行为进行调查,就要调查消费者的具体消费行为。二是人们的事实与行为后果。被调查者本人的事实与行为或通过被调查者了解他人的事实与行为的后果及产生后果的原因。三是人们的态度、意见、动机、感觉、偏好等。

3. 调查问卷的内容

一份完整的调查问卷,应包括以下几个方面的内容。

1) 问卷标题

问卷标题必须能够概括调查的主题,以使被调查者明确调查的主要内容和调查目的,应简明扼要,但又必须点明调查对象或调查主题。国家确定的调查问卷,还应在表头的左上方列出调查对象,右上方标明“表号、制表机关和文号”字样。

2) 问卷前言

问卷的卷首通常有一个简要的前言,问卷前言又称引言,一般是以信函的形式向被调查者表示感谢,说明本次调查的目的、意义、主要内容和承诺有关保证及调查的组织单位、调查结果的使用者等有关事项,其目的在于引起被调查者对填答问卷的重视和兴趣,使其对调查给予积极支持和合作。如涉及需为被调查者保密的内容,必须指明予以保密,不对外提供,等等。

3) 甄别内容

甄别内容是指通过设计一些问题先对被调查者进行过滤,筛选掉不符合条件的被调查者,然后得到满足条件的调查对象。通过甄别或过滤,一方面可以筛选掉与调查事项

有直接关系的人,以达到避嫌的目的;另一方面可以确定哪些人是合格的被调查者,哪些人不是合格的被调查者。

4) 被调查者信息

被调查者信息主要是指被调查者的一些主要标志,这些是分类分析的基本控制变量。如果被调查者是个人,其基本情况信息可能包括被调查者的性别、民族、职业、收入、教育程度、婚姻状况、家庭人口等;如果被调查者是企业,其基本情况是指企业名称、单位代码、行政区划代码、企业地址、企业规模、企业所在国民经济行业、企业登记注册类型、固定资产原值、职工人数、营业收入等情况。具体列入多少项目,应根据调查目的、调查要求及调查对象的特点而定,并非多多益善。设置这些项目,一是满足对调查资料进行分组研究的需要;二是进一步了解被调查者情况的需要;三是查询的需要。有时也可以考虑将被调查者基本情况放在问卷说明后面或者问卷的末尾。

5) 问卷主体

调查问卷的主体内容是问题和答案,是调查者所要调查的基本内容和调查问卷中最重要的部分。调查问卷的主体内容设计是否准确、科学和易懂,直接影响整个调查的价值和质量。调查问卷的主体内容应主要是提出调查的问题和可供选择的答案,根据调查目的,通过问题的逐一展开,以获取有关资料,但内容不宜过多、过繁,应根据需要而确定。

6) 问卷编码

编码是调查问卷中的一个组成部分,是对问卷中的问题和答案用数字或字符统一设计代码的工作过程。要对问卷调查所得的资料进行定量分析,就必须把调查问卷加以编码,把问题和答案变成计算机能够接受与处理的语言,然后输入计算机,再用一定的程序进行统计分析。一般情况都是用数字代号系统,并在问卷的最右侧留出“统计编码”位置。

7) 问卷说明

一份完整的调查问卷应包括必要的问卷说明或注释,问卷说明可设在问卷的最后,也可放在问卷前言之后。它是为帮助被调查者正确回答问题而作出的补充说明。问卷说明通常包括指标解释、调查须知及其他事项说明等。这部分通常还可以增加简短的几句话对被调查者的合作表示真诚的感谢;也可以稍长一些,顺便征询一下对问卷设计和问卷调查本身有何感受。

8) 作业记载

作业记载是指在调查问卷的最后记录调查人员的姓名、访问日期、对被调查者回答的评价及调查实施情况等的记录。调查实施情况记录一般记录调查中可供参考的重要情况、调查的效果和需要复查、校正的问题等。如有必要,作业证明的记载中还需注明被调查者的姓名、单位或家庭住址、联系方式等,以便于审核和进一步追踪调查。对于涉及被调查者隐私的问卷,则视情况是否不列入上述内容。

9) 结束语

结束语一般采用以下三种表达方式。

(1) 周密式结束语。对被访者的合作再次表示感谢,以及关于复核与不要漏填少填

的请求。这种表达方式既显示访问者首尾一贯的礼貌,又督促被访者填好未回答的问题和改正有差错的答案。如可以这么表述:“对于你所提供的协助,我们表示诚挚的感谢!为了保证资料的完整与翔实,请你再花一分钟,翻一下自己填过的问卷,看看是否有填错、填漏的地方。谢谢!”

(2) 响应式结束语。提出关于本次调研的形式与内容的感受或意见等方面的问题,征询被访者的意见,问题形式可用封闭式,也可用开放式。

(3) 开放式结束语。提出本次调查研究中的一个重要问题,在结尾安排一个开放式的问题,以了解被访者在标准问题上无法回答的想法。例如,“你对××××方面有何建议?”

(4) 封闭式结束语。如你填完问卷后对我们的这次调查有什么感想?

4. 调查问卷的问句设计

调查问卷的问句设计之前,需要构建一个调查问卷的总体框架,作为指导设计问卷的提问语句和对资料进行分析的一种总体思路的逻辑架构图。总体框架紧紧围绕所研究的主题,明确问卷中应设置哪些问句,不设置哪些问句,使得问卷中的每一个问题都不可或缺,不可替代;还使得各问题之间建立起一种内在的逻辑联系,为以后的统计分析提供依据。同时为保证所有可能情况都已包括在内,需为过滤性问题设立一个流程图。

在总体框架下精心设计出每一个问句,问句可以选择设计为开放式或封闭式。一般来说,问句应以封闭式问题为主,使得回答者填写时较方便、容易,能省时省力,还便于进行数据处理和定量分析。开放式问题由于回答者可以充分自由地发挥、取得的资料较生动、可对问题进行深入的研究,但要花费较多的时间和精力,且要求回答者具有较高的知识水准和文字表达能力,所以适用于探索性问题的调查。

调查问卷的问句设计需要着重考虑措辞,要把问题的内容和结构转化成清晰易懂的语句,在问题的措辞上要注意以下事项。

1) 表述的语言

问题内容的语言表述按照以下要求。

(1) 语言应浅显、易懂,不要超过被调查者的领悟能力,应该用普通用语,而不用特殊专有名词。

(2) 用语要尽量简单,不能太复杂,让社会各个阶层的被调查者理解其含义。

(3) 字句的意义力求清楚明白,措辞不要含混或暧昧。

(4) 问题不能具有双重(多重)含义、两个(多个)事实或观念,以免造成被调查者不知如何选择。

(5) 用准确的语言叙述或描述,把范围固定。

(6) 避免太普遍化的问话,以免使被调查者不知真正目的。

2) 表述的情绪

问题表述的情绪按照以下要求。

(1) 避免主观及情绪化。避免主观及情绪化字句,以免引起被调查者的自我保卫性

回答。

(2) 避免诱导回答及暗示性回答。诱导包括前言式诱导、问题的陈述偏袒和结构性的诱导。

(3) 问题不能带有倾向性。以免诱导被调查者选择某一答案,影响调查问卷的可信度。

(4) 避免涉及隐私的问题。避免不受欢迎或涉及隐私的问题,以免得到不真实的回答。

(5) 不直接询问敏感性问题。不要直接询问敏感性问题,通过间接、婉转、富有技巧性的方式方法提问敏感性问题。

(6) 问题原则上涉及社会禁忌与爱好。

3) 内容的理解

问题内容的理解按照以下要求。

(1) 提出的问题要在被调查者的知识和能力范围内能懂得。

(2) 不要引起误解或争论,力争得到真实的回答。

(3) 不要用假设或猜测语句。

(4) 不要问被调查者不知道的问题。

(5) 不要用否定句提问,避免造成被调查者的不适应。

4) 问题的原则

问题内容坚持的基本原则如下。

(1) 问题不应采用“一般”问题,以“特殊”问题为主。特殊性问题容易得到准确答案,而一般性问题难以得到可靠答案,因为一般性问题包含多个特殊性问题,这些特殊性问题有可能包含被调查者的截然相反的看法,难以统一成一个答案。

(2) 一般先问集体性问题,再问个体性问题。

(3) 只要被调查者能够明白问题的确切含义,问题的陈述就要采用尽可能简短的语言。

(4) 敏感性的问题尽可能采用间接提问,而其他问题最好直接提问。

5. 问卷的答案设计

封闭式问卷是以问和答的形式来取得资料的,若答案设计不好,即使问题提得再好,也势必影响到问卷的质量。

1) 答案设计的要求

问卷答案的设计需要满足以下要求。

(1) 答案与问题相对应。使答案与问题相对应,避免答非所问。

(2) 答案要具有穷尽性。穷尽性就是说问卷应当努力设计完备的答案,应该涵盖所有可能的回答,不能有遗漏。

(3) 答案要具有互斥性。应该努力使每个答案彼此独立,彼此不会重叠或相互包含,从而避免被调查者选择困难或者往往在只能选一项时却选择了相关的多项。

(4) 选择合适的测量尺度。鉴于答案实际上是测量尺度,所以,在能够使用高层次尺度时,应当尽量使用高层次尺度,这样在分析数据时,可以根据需要将其转换为较低层次的尺度。

(5) 答案的设计要简洁明了。答案的设计,既要避免每个答案的表述太长、太啰唆,又要避免答案数量太多。

2) 答案类型

问卷答案有开放式答案和封闭式答案两种。开放式答案可以让被调查者充分地表达自己的看法和理由,并且比较深入,有时还可获得研究者始料未及的答案,但是可能收集到的资料中无用信息较多,难以统计分析,面访时调查员的记录直接影响到调查结果,并且由于回答费事,可能遭到拒答。封闭式答案是标准化的,对答案进行编码和分析都比较容易;回答者易于作答,有利于提高问卷的回收率;问题的含义比较清楚,其提供的答案有助于理解题意,避免被调查者因不理解题意而拒绝回答。也有可能被调查者对题目理解不正确,难以觉察出来,或者可能产生“顺序偏差”或“位置偏差”,即被调查者选择答案可能与该答案的排列位置有关。

(1) 开放式问题的答案类型。开放式问题的答案类型主要包括:自由回答法是问卷中的设问项目没有指定可供选择答案,被调查者自由回答,也叫无限制回答法。空位填答法是被调查者在设问问题中空出答案的位置填答内容。词语联想法是指调研人员将一些词语给被调查者,并要求他立即回答所想到的是什么。在立即反应下,可以获得与“刺激词汇”相对应的联想。其他还有文章完成法、图画完成法、角色扮演法等不常见的方法。

(2) 封闭式问题的答案类型。封闭式问题的答案类型主要包括以下几种。

二项选择法。二项选择法对问卷中的问题给出非此即彼的两个答案,强制被调查者只选择其中一项即可。

多项选择法。多项选择法对问卷中的设问项目同时给出多种答案,由答问者从中选择一项或多项。答案应当包括需要了解的所有可能的情况;各种可能答案必须互相排斥,而不能互相交叉,更不能相互包容;答案的选择项应表达明确,使问题与答案不出现矛盾。

赋值评价法。赋值评价法也称等级评价法,是用打分数或确定等级来评价事物的好坏或优劣的方法。对两个以上分成等级的答案进行选择,只能选择一项,可分100分制或10分制,可分5级或10级。常用于满意度或者一些程度调查。

分配法。分配法根据实际情况分配调查对象的某项目各构成部分所占比重,要求各构成部分比重之和为1。

顺序选择法。顺序选择法的问题答案也有多个,要求被调查者在回答时,对所选的答案按要求的顺序或重要程度加以排列。其中,对所选的答案数量可以进行一定的限制,也可以不进行限制。

量表法。量表法又称评定尺度法,是问题答案由表示不同等级的形容词组成,并按照一定的程度排序,由被调查者依次选择的方法(表2-3)。量表是一种工具,旨在将一

些主观的、抽象的概念定量化。根据所测量的变量有四种级别,量表也从低级到高级有四种水平,分别是“定类量表”“定序量表”“定距量表”和“定比量表”,每个量表都有各自不同的统计分析方法。适用于若干个有相同答案形式的问题,可以将这些问题集中在一起构成一个问题的表达方式。

表 2-3 李克特量表

项目	很好	较好	一般	较差	很差
项目 1					
项目 2					
项目 3					
项目 4					
...					

双向列联法。双向列联法是将两类不同问题综合到一起,通常用表格来表现,表的横向是一类问题,纵向是另一类问题。这种问题结构可以反映两方面因素的综合作用,提供单一类型问题无法提供的信息,同时也可以节省问卷的篇幅。

配对比较法。配对比较法是设计一组具有两个不同选项的问题,要求被调查者从每个问题的一对选项中选择一项作为答案,由于问题的选项是成对的,且每对选项彼此相异,分布平衡,可以消除单项设计所难免的偏向。使用配对比较法的结果与顺序尺度法相当,但提供了较多的两两比较的详细内容,可以更好地以判断次序、优劣、喜好等。

连线配合法。连线配合法将问卷的调查项目与答案排成两列,请被调查者在调查项目与答案之间画线连接作答。

3) 答案设计的形式

根据所提问题的变量类型不同,答案设计可有如下几种形式。

(1) 定类问题的答案设计。这种设计需要对所研究的对象进行分类,类分好了,答案也就设计出来了。对于有习惯或者规定的较简单定类问题,直接根据习惯或者规定的种类进行分类,如性别、文化程度等。对于没有规定或者习惯分类的复杂问题,特别是较抽象的问题,为了达到穷尽性的要求,对于较复杂的问题,在分类时往往这样进行:先将定类问题分成几大类,然后在大类下分小类,最后将各小类进行综合,删除重复部分,就为所有可能答案;为了达到互斥性的要求,可以按同一抽象层次或标志分类,然后在某一确定的层次或标志上列出答案,为了满足变量层次的要求,根据调查者的需要来确定变量需要的层次。

(2) 定序问题的答案设计。问卷中大量的问题都是对人们态度的提问,而这些问题大多属于定序问题。对这类问题的答案往往采取的设计样式是:“①非常同意 ②同意 ③无所谓 ④不同意 ⑤很不同意”五等级,或者“①同意 ②无所谓 ③不同意”三等级的定序形式,以五等级定序答案最为常见。

(3) 定距问题的答案设计。定距问题的答案是用数字来表示的。例如,“您的月工资是:① 2 000 元以下 ② 2 000~3 000 元 ③ 3 000~5 000 元 ④ 5 000~9000 元

⑤ 9 000 元以上”。设计这类答案时,划分的档次不宜太多,每档的距离可以不等,但不宜太大,应注意每档的同质性。如在对工资收入划分时,应区别高、中、低的差别。

6. 调查问卷的问题顺序

在设计好全部问题语句和限制性答案之后,要将问题按一定的顺序进行排列,形成完整的问卷。调查问卷的问题编排的先后次序会影响到被调查者对答案的选择和对问卷的兴趣及答卷质量,是设计调查问卷的一项非常重要的工作。为了形成合理的结构,通常要注意两个方面:第一,要方便被调查者顺利地回答问题;第二,要便于调查后的资料整理和分析。

调查问卷的问题排列顺序方式常见的有以下几种。

1) 时间性顺序

时间性顺序就是问题按时间顺序来安排。一般来说,应根据历史的线索,由过去到现在,再到将来,这样就可以使逻辑和历史统一起来。若问卷有多个与时间有关的参考分框架,最好先处理完一个时间分框架,再处理下一个时间分框架,直到完整处理。

2) 类别性顺序

类别性顺序就是把同类性质的问题尽量安排在一起,而不要让不同性质或类别的问题互相混杂。这样就便于被调查者按照问题的顺序,回答完一类问题后再回答另一类问题,而不至于使他回答问题的思路经常中断和来回跳动。最常见的类别顺序形式,是先设计有关个人特征资料的问题,然后设计事实性的问题,再设计态度性的问题。

3) 内容性顺序

内容性顺序就是把问题按其复杂和困难程度来排列。

(1) 一般来说,应该先易后难、先熟后生、由浅入深,先一般性质问题、后特殊性质问题。问卷的前面的题目容易回答能够提高被调查者的积极性,会产生预热效应,有利于问卷答完。

(2) 把能引起被调查者兴趣的问题放在前面,更能激励被调查者快速高效地完成调查问卷。

(3) 对于敏感性强、威胁性大的问题,更应该放在各类问题的后面。这样,有利于创造一种宽松、随和、融洽的调查气氛,以便增强被调查者的信心,将思路逐步引向深入,而不至于一开始便面临困难或遭遇困境。

(4) 开放性问题放在最后。封闭性问题较易回答,若将较难回答的开放性问题放在前面,可能一开始就有遭到被调查者拒绝的危险。

4) 逻辑性顺序

逻辑性顺序就是研究者有意识将自变量问题放在前面,因变量问题放在后面,这样便于研究者进行资料的分析。首先是对辨别目标被调查者起过滤性作用的限制性问题;其次是最初几个问题,是易于回答、向被调查者表明调查很简单的适应性问题;再次,前 1/3 是与调研目的有关、需稍费力回答的过渡性问题,然后是中间 1/3 的难以回答及复杂的问题;最后部分是被调查者可能留下空白的问题以及分类和个人情况。

2.4 统计调查误差与控制

2.4.1 统计调查误差的概念和种类

1. 统计调查误差的概念

统计调查误差是统计调查所得的统计数量与调查对象的实际数量之间的差异，即调查所得的数量大于或小于调查对象的实际数量的差额。

2. 统计调查误差的种类

统计调查误差的种类有以下几种。

1) 工作误差和代表性误差

按误差的产生原因，统计调查误差可分为工作误差和代表性误差。

工作误差是由于调查工作的失误所造成的误差。如错误判断事实或者错误登记事实而造成的误差等，不管是全面调查或是非全面调查都会产生工作误差，从理论上讲是一种人为的、可以避免的误差。调查过程的各个环节上的工作不准确（如计量、登录、计算等方面错误）而造成的，不是指故意性行为，有意识地虚报、瞒报、拒报、迟报、伪造篡改等不属于工作误差，而是违法行为。代表性误差是以部分代表总体和推断总体时必然存在的误差。代表性误差由部分与总体间存在结构和水平上的差异，通常分为：没有遵守随机原则，主观地或部分主观地从总体中选取部分个体时，如故意多选有利的或不利的单位进行调查，从而使计算出的指标值与总体实际数量产生误差；遵守随机原则的前提下，由于样本指标值是随着抽到的样本不同而变化的，使样本指标与总体指标之间存在的差异。代表性误差是不可避免的，但可以计算和控制。只有非全面调查中才有代表性误差，全面调查中不存在这类误差。

2) 偶然性误差和系统性误差

按误差的来源，统计调查误差可分为偶然性误差和系统性误差。

偶然性误差是由各种偶然因素影响而对调查结果产生的误差。偶然性误差产生的原因主要是人们的无意行为。偶然性误差是随机产生的，不会具有某种倾向性，即在数量上的方向和大小都不固定。在对大量资料进行整理时，这种忽大忽小的偶然性误差往往会互相抵消。系统性误差是由各种确定性因素影响而对调查结果产生的误差，又称系统偏差。它具有明显的倾向性，数量上有固定的方向和大小。产生系统性误差的原因有主观因素和客观因素两个方面。主观因素就是人们出于某种目的故意夸大或缩小统计数据。客观因素主要是调查环境、调查条件决定的，如测量工具不准确、对指标概念的理解不清、调查范围模糊等。系统性误差对统计结果的影响较大，应尽量将它消除，以保证统计资料的准确性。

3) 空间误差、时间误差、方法误差和人为误差

按误差的性质，调查误差可分为空间误差、时间误差、方法误差和人为误差。

空间误差是指统计调查范围所产生的误差，包括重复或遗漏调查单位、跨区域统

计、跨单位统计等。时间误差是指统计调查对象因时期或时点界定不准确所产生的误差。如企业核算时间不能满足统计部门的报表制度要求而估报所产生的误差；延长或缩短时期所产生的误差；时间错位产生的误差；等等。方法误差是因使用特定的统计调查方法所产生的误差。如抽样调查中的代表性误差是指采用抽样调查方法中的随机样本来推算总体所产生的误差的平均值，即抽样平均误差，不是绝对的统计误差。对代表性误差，根据组织方法和抽取样本的容量，一般可以计算其平均误差，而且通过扩大样本量或优化调查的组织方法来缩小。又如统计部门因人力、物力和财力等资源不足，致使报送渠道不畅通，统计调查不到位，推算方法不科学、不规范所产生的误差。人为误差是指在统计设计、调查、整理汇总和推算等过程中因人为过错产生的误差。人为误差是统计误差中产生因素最多的一类，它又分为度量性误差、知识性误差、态度性误差和干扰性误差。度量性误差是指统计指标因计量或者从生产量到价值量换算所产生的误差；知识性误差是指统计人员因统计知识不够，对统计指标的含义不理解或错误理解所产生的误差；态度性误差是指统计人员因对统计工作不负责而随意填报统计数据而产生的误差，包括乱报、漏填或不按规定的计量单位填报等；干扰性误差是指统计对象或统计部门受某种利益驱动而虚报、漏报或者捏造统计数据所形成的误差。

4) 源头误差、中间环节误差和最终误差

按误差产生的环节，统计调查误差可分为源头误差、中间环节误差和最终误差。

源头误差是指起报单位或申报者所产生的误差，有些场合也叫调查误差或登记误差。中间环节误差是指统计调查数据在逐级上报过程中所产生的误差，包括加工整理、汇总和推算等环节。最终误差是指下级各基层数据汇总数或规范的方法得到的推算数与最终使用数之间的差异值。按工作环节划分的统计误差类别是相对的，中间环节误差在不同的场合有可能是源头误差，也有可能是最终误差。

2.4.2 统计调查误差的产生原因

调查误差的产生原因是多方面的，主要包括以下几点。

(1) 因调查方案规定不妥、问卷设计不科学而产生的设计误差，如采用了有缺陷的抽样框、不适当的调查方式方法、组织实施程序设计不合理。

(2) 因调查人员在执行访问调查过程中操作不当或与被调查者合作不畅而产生的登记误差，如问答欠佳、记录不准、转抄有误和录入出错等。

(3) 因计量器具不准、计量手段的局限性、数据计算与汇总有误而产生的计量误差。

(4) 因被调查者原因而产生的立意误差，主要包括不能正确回答误差和不愿正确回答误差，以及不回答误差。例如，被调查者因统计知识不够、文化水平低，对调查指标的含义不能理解或错误理解所产生的知识性误差；由于调查指标内涵过大、调查时期跨度过长，被调查者回忆不清而产生的回忆性误差；被调查者对统计调查工作不负责而随意填报数据而产生的态度性误差，包括无回答、乱报或不按规定的计量单位填报等；被调查者或统计部门受某种利益驱动，不愿意如实反映而虚报、漏报或者捏造数据所形成的意愿性误差。

2.4.3 防止与减少统计调查误差的办法

防止与减少统计调查误差的办法有以下几种。

1. 制订科学的调查方案

要正确周密地制订统计调查方案。调查过程是一项系统工程，事先必须进行周密设计，包括明确调查对象的范围，说明调查项目的具体含义和计算方法，确定合理的调查方式方法，规定合适的时间、地点，等等，以使调查人员或填报人员有一个统一的依据。

2. 搞好统计的基础工作

加强对统计人员的业务培训，提高统计人员的素质；健全原始记录，完善统计台账和内部报表，确保资料来源可靠；建立对现场调查人员进行奖惩的制度。

3. 切实抓好调查实施工作

重视对调查人员的挑选和管理、重视现场调查工作，抓实各个环节的操作，达到操作规范化，确保各地调查执行方案统一、数据统计口径一致；选择合理的资料收集方法，做到科学抽样和选典；要建立现场登记数据质量评估标准，进行调查过程的检查与监控；要加强对调查资料的审核，发现差错及时纠正；强化对调查结果进行检验、评估；等等。

4. 依法行政，依法治统

从建立健全统计法制入手，教育统计人员严格执行统计法，坚持原则，同一切弄虚作假的行为作斗争，维护统计数字的真实性。要加大统计执法力度，严惩弄虚作假行为，维护统计工作的严肃性，逐步建立全社会的统计诚信体系。

2.5 案例：我国人口统计数据的调查与发布

2.5.1 案例背景

人口统计是统计制度中一项重要的内容，是适应社会发展和国家管理的需要而产生与发展起来的，它是人类社会出现最早的统计。公元前 4500 年，巴比伦调查过人口项目；中国在公元前 2100 年的夏禹时代就有了人口数和土地数的统计；周朝以后，历代都有人口调查制度；公元 2 年有正式的全国和分地区的人口记载，奴隶社会和封建社会进行人口调查与登记是为了适应赋税、徭役、征兵的需要，统计的方法和内容都比较简单；在汉代，武帝时期进行了人类历史上第一次人口统计；资本主义社会的人口统计有了很大发展，包含人口数、性别、年龄、民族、就业和失业、行业和职业、文化程度、宗教信仰、婚姻状况、生育情况以及其他等统计，统计内容日益丰富，人口统计制度更趋健全，并逐步建立和发展了现代人口普查制度，人口统计资料越来越系统完整。在社会主义社会，经济发展和人口发展都纳入国家计划，人口统计为编制国民经济和社会发展计划、

制定人口政策、编制和检查人口计划服务,而且为国家行政管理和人口研究工作提供资料,统计方法与数据的发布方法逐步按照国际通行的标准进行,包括指标设置、指标定义、调查方法以及数据的发布方法等统计行为。

人口统计调查是按人口统计任务提出的具体要求,有目的、有计划、有组织地对全国或某一地区、某一特定部分的人口进行调查,收集有关人口原始资料的过程,它是人口统计工作的基础,是人口统计资料整理和人口统计分析的前提,目前我国建立起以人口普查、1%人口抽样调查、人口变动抽样调查与人口统计报表制度相结合的人口调查制度。

2.5.2 案例分析

1. 人口变动抽样调查

人口抽样调查是指按照随机的原则,从被研究的人口总体中抽选一部分单位作为样本进行调查,并根据调查所得的资料推断人口总体相应指标值的一种非全面的调查。我国自20世纪80年代以来实施全国定期人口抽样调查。国家统计局在1982年第三次人口普查的基础上,每年调查一次,形成了一项调查制度。调查目的是准确及时地掌握每年人口变动情况,并为国家检查人口政策和人口计划执行情况提供可靠的调查数据。人口变动抽样调查的基本方法是:以出生率作为估计样本的依据,确定置信度为95%,允许误差为0.5%,开始实施时样本规模约为总人数的0.5%,目前抽取占全国总人口的1%左右,该项调查只对国家和省级有代表性。

人口变动抽样调查的调查表在住户登记之后,被收集集中到省级统计局,由省级统计局进行编码、数据录入、编辑审核等工作,然后,将净化后的原始数据传输到国家统计局,进行最后的汇总工作。汇总出各项主要数据并根据科学方法推算出年度的总体数据,在第二年的2月下旬发表的“国民经济与社会发展统计公报”中公布。此项调查同样规定,调查机构和调查人员必须为住户申报的家庭资料保密,不得泄露。

目前,我国全国人口变动抽样调查方案的主体内容包括以下几个方面。

(1) 调查目的。为了准确、及时地掌握全国和各省(自治区、直辖市)人口变动情况,为国家和省级人民政府制定国民经济与社会发展计划、掌握人口增长情况提供可靠的人口数据,根据国办发(1992)57号文件的要求,进行年度人口变动情况抽样调查。

(2) 统计范围和调查对象。人口变动调查对象为抽中调查小区内具有中华人民共和国国籍的人。调查以户为单位进行,既调查家庭户,也调查集体户。应在抽中调查小区内登记的人包括:①调查时点居住在本户的人;②户口在本户,调查时点未居住在本户的人;③抽中调查小区年度内死亡人登记相关项目。

(3) 调查内容。调查内容分为按户填报、按人填报和按社区填报。按户填报的项目有户编号、户别、应在本户登记的人数、年度本户出生人口和死亡人口5个基本信息;按人填报的项目:姓名、性别、出生年月、民族、与户主关系等人口基本信息,调查时点居住地、户口登记地、在本市居住时间、离开户口登记地时间、离开户口登记地原因、户口性质、土地承包权、一年前常住地、一年常住地类型等反映人口迁移流动的指

标,还有是否识字、受教育程度、学业完成情况、上周工作情况、行业、职业、工作地点、前往工作地所乘主要交通工具及所需时间、参加社会保险情况、婚姻状况和年度内生育情况等指标;死亡人口填报的项目有编号、姓名、性别、出生年月、死亡年月。

所有村和社区将填报《2015年人口调查村、居委会(社区)基本情况表》,填报项目有常住人口数、户籍人口数、外来人口、外出人口、出生人口、死亡人口、家庭户人数、集体户人数、是否有集体宿舍集中区域、农林牧渔从业人员占从业人员比例、主要饮用水来源、市政排水系统、生活垃圾处理系统,共13个。根据国家需要,每年还在调查中增加或调整个别重要项目。

(4) 抽样调查方法。人口变动调查以全国为总体,以各省(区、市)为子总体。人口变动调查以全国为总体,以各省(区、市)为子总体。采取分层、多阶段、概率比例、整群抽样等方法抽样,最终样本单位为调查小区。设计样本量约为140万人。多阶段抽样的方法为:第一阶段抽取县级单位,抽取方法为分层、概率比例抽样;第二阶段抽取村级单位,采取简单随机抽样;第三阶段抽取调查小区,调查小区以2015年全国1%人口抽样调查划分的调查小区的对应地域为准。按平均每个调查小区常住人口为250人左右计算,调查小区样本量全国5500~5800个。对2016年至2019年全国人口变动调查进行为期4年的周期样本设计,在2015年建立的样本轮换框中,根据样本按比例轮换的要求选取本次调查的样本,调查小区为最终样本单位。数据采集和汇总程序由国家统计局数管中心负责统一编制并下发。2014年及以前的调查,调查员入户登记,填写纸质调查表后,由区县统计局工作人员进行数据录入工作后上报;2016年起,调查员手持PDA(电子终端设备)进行数据采集后直接上报。

(5) 调查时间和地点。人口变动调查标准时点为当年11月1日零点,调查期限为11月1日至11月中旬。调查员入户登记,采集人口数据后直接上报。

(6) 组织实施计划。①为了保证全国调查数据的范围、分类和计算方法的统一性,各地区必须严格执行调查制度的规定,不得自行更改。②调查员、调查指导员以及各级统计机构及其工作人员都要按照《中华人民共和国统计法》的规定,对调查结果,特别是被调查户的情况保守秘密,不得向调查机构以外的任何单位和个人泄露。③数据处理。全国数据由国家统计局负责汇总,各省(自治区、直辖市)的数据要按照国家统计局统一的部署和安排进行汇总。④数据发布。按年度频率通过国家统计局外网、新闻发布会、统计公报、统计年鉴或其他统计资料等形式对公众发布。

2. 人口普查

我国已成功地进行了六次人口普查,这些普查分别于1953年、1964年、1982年、1990年、2000年、2010年进行。我国最新的《全国人口普查条例》规定,我国的人口普查每10年进行一次,在年号末尾为0的年份进行。这是联合国所建议和提倡的,是世界大多数国家进行人口普查所采用的时间,少数国家为接近0的年份。

我国第六次人口普查方案主体内容如下。

(1) 调查目的。第六次全国人口普查的目的是查清2000年以来我国人口数量、结构、分布和居住环境等方面的变化情况,为科学制订国民经济和社会发展规划,统筹安排人民的物质和文化生活,实现可持续发展战略,构建社会主义和谐社会,提供真实准

确、完整及时的人口统计信息支持。

(2) 调查对象和调查单位。人口普查对象是指普查标准时点在中华人民共和国境内的自然人以及在中华人民共和国境外但未定居的中国公民,不包括在中华人民共和国境内短期停留的境外人员。人口普查以户为单位进行登记,分为家庭户和集体户。以家庭成员关系为主、居住一处共同生活的人口,作为一个家庭户;单身居住独自生活的,也作为一个家庭户。相互之间没有家庭成员关系、集体居住共同生活的人口,作为集体户。

(3) 调查项目。人口普查表分为《第六次全国人口普查表短表》和《第六次全国人口普查表长表》,普查表长表抽取 10% 的户填报,普查表短表由其余的户填报。人口普查登记的主要内容包括姓名、性别、年龄、民族、国籍、受教育程度、行业、职业、迁移流动、社会保障、婚姻、生育、死亡、住房情况等。

普查表短表共有 18 个项目,分为两个部分,按户填报和按人填报。按户填报的项目有户编号、户别、本户应登记人数、(2009 年 11 月 1 日至 2010 年 10 月 31 日)出生人口和死亡人口、本户住房建筑面积和本户住房间数 6 个项目;按人填报的项目有姓名、与户主关系、性别、出生年月、民族、普查时点居住地、户口登记地、离开户口登记地时间、离开户口登记地原因、户口性质、是否识字及受教育程度 12 个项目。

普查表长表共有 45 个项目,按户填报的项目有户编号、户别、本户应登记人数、2009.11.1—2010.10.31 出生与死亡人口、住房用途、本户住房建筑面积、本户住房间数、建筑层数、承重类型、住房建成年代、主要炊事燃料、住房内有无管道自来水、住房内有无厨房、住房内有无厕所、住房内有无洗澡设施、住房来源、月租房费用 17 项。按人填报的项目有姓名、与户主关系、性别、出生年月、民族、普查时点居住地、户口登记地、离开户口登记地时间、离开户口登记地原因、户口登记地类型、户口性质、出生地、五年前常住地、是否识字、受教育程度、学业完成情况、工作情况、行业、职业、未工作原因、三个月内是否找过工作、能否工作、主要生活来源、婚姻状况、结婚(初婚)年月、生育子女数、2009.11.1—2010.10.31 的生育状况、身体健康状况 28 项。

2009 年 11 月 1 日至 2010 年 10 月 31 日期间有死亡人口的户,同时填报《第六次全国人口普查死亡人口调查表》。包括死亡人口的户编号、姓名、性别、出生年月、死亡年月及年龄、民族、受教育程度、婚姻状况 8 个项目。

在境内居住的港澳台和外籍人员,在现住地进行登记,填写供港澳台和外籍人员使用的普查表短表。

(4) 调查时间和地点。第六次全国人口普查的标准时点是 2010 年 11 月 1 日零时,从 11 月 1 日起普查人员正式入户登记。人口普查的登记工作,从 2010 年 11 月 1 日开始到 11 月 10 日结束。

(5) 调查方式方法。人口普查采用按现住地登记的原则。每个人必须在现住地进行登记。普查对象不在户口登记地居住的,户口登记地要登记相应信息。普查区划分以村民委员会和居民委员会所辖区域为基础。每个普查区按照一个普查员所能承担的工作量,划分成若干个普查小区。每个普查小区至少配备 1 名普查员,每个普查区至少配备 1 名普查指导员,原则上 4~5 个普查小区配备 1 名普查指导员。人口普查登记,采用普查员入户查点询问、当场填报的方式。普查员应当按照普查表列出的项目逐户逐人询问清楚,

逐项进行填写,做到不重不漏、准确无误。现役军人及军队管理的离退休人员,武装警察部队由军队领导机关、武警机关统一进行普查、汇总。依法被判处有期徒刑、劳动教养的人员,由当地公安机关和监狱、劳教机关进行普查。

(6) 组织实施计划。人口普查工作,按照“全国统一领导、部门分工协作、地方分级负责、各方共同参与”的原则组织实施。国务院和地方各级人民政府设立第六次全国人口普查领导小组及其办公室,领导和组织实施全国与本区域内的人口普查工作;村民委员会和居民委员会设立人口普查小组,做好本区域内的人口普查工作;领导小组各成员单位按照各自职能,各负其责、通力协作、密切配合。人口普查所需经费,由国务院和地方各级人民政府共同负担,并列入相应年度的财政预算,按时拨付,确保足额到位。各级宣传部门和人口普查机构应采取多种方式,积极做好人口普查的宣传 work,为人口普查工作的开展营造良好的社会氛围。人口普查实行严格的质量控制制度,地方各级人口普查机构主要负责人对本行政区域人口普查数据质量负总责,确保人口普查数据真实、准确、完整、及时。在人口普查登记、快速汇总、编码、数据处理各环节实行质量验收制度,验收不合格的必须返工,直至达到规定的质量验收标准方可转入下一工作环节。

2010年8月底前完成户口整顿工作;2010年10月底前完成摸底工作;2010年11月1日开始到11月10日进行人口普查的登记工作;2010年11月15日前完成复查工作;2010年11月底前完成事后质量抽查工作。人口普查数据由人口普查机构负责进行数据处理。录入采用光电录入的方式,数据录入、编辑、审核、汇总程序由国务院人口普查办公室统一下发。人口普查机构对普查登记的主要数据,先进行快速汇总。国家统计局和国务院人口普查办公室对数据进行审核后发布主要数据公报,各省、自治区、直辖市的主要数据应于国家公报发布之后发布。国务院人口普查办公室应于2011年12月31日前完成人口普查全部数据的汇总工作。人口普查数据处理工作结束后,原始普查表按国务院人口普查办公室的统一规定销毁。

3. 人口小普查

国务院2010年颁布的《全国人口普查条例》规定,在两次人口普查之间开展一次较大规模的人口调查,每逢年号末尾为5时,就进行1%人口抽样调查,也称“人口小普查”。“小普查”的内容、组织实施形式和数据发布时间与方法等,都与大普查相似。因此在2015年开展全国1%人口抽样调查,有利于查清2010年以来我国人口在数量、素质、结构、分布以及居住等方面的变化情况,为科学制订国民经济和社会发展规划,提供科学准确的统计信息支持。

我国2015年全国1%人口抽样调查方案如下。

(1) 调查目的。2015年全国1%人口抽样调查的目的是了解2010年以来我国人口在数量、素质、结构、分布以及居住等方面的变化情况,为制订国民经济和社会发展规划提供科学准确的统计信息支持。

(2) 调查对象和调查单位。调查以全国为总体,各地级市为子总体,调查对象为抽中调查小区内的全部人口(不包括港澳台居民和外国人),采取分层、二阶段、概率比例、

整群抽样方法,其中群即最终样本单位为调查小区。全国调查的样本量约占全国总人口的1%,在我国境内抽取约6万个调查小区,覆盖人口约1400万人。应在抽中调查小区内登记的人包括:2015年10月31日晚居住在本调查小区的人;户口在本调查小区,2015年10月31日晚未居住在本调查小区的人。中国人民解放军现役军人由军队领导机关统一进行调查。调查以户为单位进行登记,分为家庭户和集体户。2015年全国1%人口抽样调查小区规模划分原则为80个住房单元,常住人口大约250人。

(3) 调查内容。主要调查人口和住户的基本情况,共设置两张入户登记表:《2015年全国1%人口抽样调查表》《2015年全国1%人口抽样调查死亡人口调查表》。《2015年全国1%人口抽样调查表》的调查内容分为“住户项目”和“个人项目”,“住户项目”调查住户的住房详细地址,年度内出生和死亡人数,住户类别,住房的类型、面积、房间数、建筑物层数、建成年代、来源、有无厨房、有无厕所等人口和住房方面情况;“个人项目”设置个人姓名、与户主关系、性别、出生年月、民族、户口登记地址、登记时居住地址、在本市居住时间、离开户口登记地的时间和原因、是否有农村土地承包权、一年常住地、五年前常住地、是否识字、受教育程度、学业完成情况、上周工作情况、行业、职业、工作地点、通勤工具、未工作原因、三个月内是否找过工作、有合适工作能否在两周内开始工作、婚姻状况、夫妇为独生子女情况、生育子女数量、过去一年生育情况、主要生活来源、参加养老保险和社会医疗保险情况、身体健康情况等调查项目。

《2015年全国1%人口抽样调查死亡人口调查表》包括死亡人口的姓名、性别、出生年月、死亡时间、民族、受教育程度、婚姻状况7个项目。

(4) 调查时间和地点。调查的标准时点为2015年11月1日零时。住户可以选择由调查员手持电子终端设备(PDA)入户登记的方式,也可以选择在互联网上填写调查表直接上报的方式。选择互联网填报的住户应于2015年11月7日前完成调查表的填写和提交。对在规定时间内没有完成的住户,调查员将再次入户使用PDA进行登记。全部登记工作应于11月15日前完成。

(5) 调查方式方法。样本的抽取由全国1%人口抽样调查办公室负责实施,全国调查的样本量约占全国总人口的1%。调查以全国为总体,各地级市为子总体,采取分层、二阶段、概率比例、整群抽样方法,其中群即最终样本单位为调查小区。二阶段抽样的方法为:第一阶段抽取村级单位,抽取方法为分层、概率比例抽样,第二阶段抽取调查小区。在划分调查小区的同时,绘制抽中村级单位内调查小区分布图,并给调查小区升序编码,绘制抽中调查小区内所有建筑物的分布图。调查采用调查员手持电子终端设备(PDA)入户登记与互联网自主填报相结合的方式。

(6) 组织实施计划。调查工作按照“统一领导、分工协作、分级负责、共同参与”的原则组织实施。国家和县以上地方各级人民政府成立2015年全国1%人口抽样调查工作领导小组及其办公室,被抽中的乡、镇和街道办事处成立1%人口抽样调查办公室,领导和组织实施全国与本地区的1%人口抽样调查工作。领导机构各成员单位要按照各自职能分工,对本行政区域的调查数据质量负责,确保调查数据真实、准确、完整、及时;经费按照分级负担原则,由中央和地方各级人民政府共同负担,并列入相应年度的财政预算,按时拨付、确保到位;各级宣传部门和调查机构应采取多种方式,积极做好

1%人口抽样调查的宣传工作。

2015年10月31日前,调查员和调查指导员完成对调查小区的人口状况摸底工作,明确调查登记的范围、绘制调查小区图、编制调查小区户主姓名底册。现场登记工作从2015年11月1日开始,采用调查员手持PDA入户询问、现场填报,或由住户通过互联网自主填报的方式进行。2015年11月25日以前完成事后质量抽查,其结果只作为评价全国调查数据质量的依据。预计2016年4月底以前调查数据的汇总、发布和管理。

我国还有人口统计报表制度,它是我国经常地、定期地按照统一规定,由各级公安机关根据户口登记资料,自下而上提供人口资料的一种制度。

习 题 2

- 2.1 简述统计设计的分类与统计设计的内容。
- 2.2 简述指标体系设计的内容,科学地建立统计指标体系,必须遵循什么原则?
- 2.3 设计企业的某一方面的评价指标体系。
- 2.4 统计调查有哪些分类?统计数据的具体收集方法有哪些?
- 2.5 什么是普查?普查有哪些主要特点和应用意义?
- 2.6 试述重点调查、典型调查、抽样调查的区别和联系。
- 2.7 随机抽样有哪些主要类型?举例说明它们的特点及适用条件。
- 2.8 完整的统计调查方案必须包括哪些内容?
- 2.9 举例说明调查单位、调查对象与填报单位及它们之间的关系。
- 2.10 某城市拟对该市专业技术人员进行调查,想要通过调查来研究下列问题:通过描述专业技术人员队伍的学历结构来反映队伍的整体质量;研究专业技术人员总体的职称结构比例是否合理;描述专业技术人员总体的年龄分布状况;研究专业技术人员完成的科研成果数是否与其最后学历有关。请回答:
 - (1) 该项调查研究的调查对象和调查单位是什么?
 - (2) 该项调查研究需要用到的基本统计分析方法是什么?
 - (3) 为完成该项调查研究任务,对每一个调查单位应询问哪些调查项目?
- 2.11 某化妆品生产商欲了解近期推出的一种新产品的使用效果,准备做市场调查,请问:
 - (1) 本调查的对象是什么?
 - (2) 调查单位是什么?
 - (3) 报告单位是什么?
 - (4) 采取什么调查方式为好?
- 2.12 某服装生产厂家想通过市场调查了解以下问题:企业产品的知名度;产品的市场占有率;用户对产品质量的评价及满意程度。回答以下问题:
 - (1) 设计出一份调查方案。
 - (2) 你认为这项调查采取哪种调查方法比较合适?
 - (3) 设计出一份调查问卷。
- 2.13 怎样减少和防止统计调查误差?试举例说明。

第 3 章

统计整理

引例

第三次全国农业普查主要数据公报（第五号）

每年开始的几个月内，从每个企业发布上一年度的年报、国家和各级政府的统计部门发布所属区域的上一年度国民经济与社会发展统计公报，每次人口普查、经济普查、农业普查都会发布公报，这些报表图文并茂、简明准确。例如，我国第三次全国农业普查第五号公报显示：2016 年，我国农业生产经营人员 31 422 万人，其中女性 14 927 万人。在农业生产经营人员中，年龄 35 岁及以下的 6 023 万人，年龄在 36~54 岁的 14 848 万人，年龄 55 岁及以上的 10 551 万人，具体数据如表 3-1 所示。

表 3-1 农业生产经营人员数量和结构

万人、%

项目	全国	东部地区	中部地区	西部地区	东北地区
人员总数/万人	31 422	8 746	9 809	10 734	2 133
性别构成					
男性	52.5	52.4	52.6	52.1	54.3
女性	47.5	47.6	47.4	47.9	45.7
年龄构成					
35 岁及以下	19.2	17.6	18.0	21.9	17.6
36~54 岁	47.3	44.5	47.7	48.6	49.8
55 岁及以上	33.6	37.9	34.4	29.5	32.6
受教育程度构成					
未上过学	6.4	5.3	5.7	8.7	1.9
小学	37.0	32.5	32.7	44.7	36.1
初中	48.4	52.5	52.6	39.9	55.0
高中或中专	7.1	8.5	7.9	5.4	5.6
大专及以上	1.2	1.2	1.1	1.2	1.4
从事农业行业构成					
种植业	92.9	93.3	94.4	91.8	90.1
林业	2.2	2.0	1.8	2.8	2.0
畜牧业	3.5	2.4	2.6	4.6	6.4
渔业	0.8	1.6	0.6	0.3	0.5
农林牧渔服务业	0.6	0.7	0.6	0.5	1.0

调查获得分散、零碎、杂乱的数据资料之后,在审核、筛选、排序等预处理的基础上,如何选择分组标志、确定分组参数等进行统计分组,运用哪些方式方法进行统计汇总,将调查获得的数据变成能反映丰富信息的统计表格,统计表格包括哪些内容、有哪些种类、有什么绘制规则,如何使这些统计表变成各种美观生动、丰富多样的统计图形,等等,这些内容都会在本章学习中涉及。

3.1 统计整理概述

3.1.1 统计整理的概念与意义

统计整理就是根据统计研究的目的和任务的要求,对收集的资料进行科学的加工整理,使之系统化、条理化,成为能够说明事物或现象的总体数量特征的资料的过程。广义的统计整理包括对原始资料的整理和对次级资料的整理两个方面,其中主要是对原始资料的整理。

统计整理在整个统计工作过程中具有不可替代的作用,其重要意义具体体现在以下几个方面。

1. 统计整理是获得能揭示总体特征的资料的途径

统计调查取得的第一手资料,通常只是反映总体各单位特征的资料,是分散、零碎、表面的具体情况,不能体现总体的特征和规律性。需要对各单位的资料进行加工整理,去伪存真、去粗取精,使之系统化、条理化,使它能说明总体情况,揭示出总体的内在特征,以便通过综合指标对总体作出概括性的说明。

2. 统计整理是整个统计工作和研究过程的中间环节,起着承前启后的作用

统计整理是统计调查的继续和发展、统计分析的前提和基础,是统计工作和研究过程中的承前启后的连接点。统计调查收集的资料只有通过科学地预处理、分组、汇总、编制图表等,才能使人们的认识实现由个别到全体、由特殊到一般、由现象到本质、由感性到理性的转化,提供准确、系统、条理清晰、能在一定程度上说明总体特征的综合资料,使统计调查得到的资料充分发挥作用,以保障统计分析顺利进行。

3. 统计整理是积累历史资料的必要手段

统计调查的资料分散、零碎、量大,特别是大规模的统计调查取得的统计资料数量庞大,不便于保存,只有整理后的系统化、条理化的资料才能更有价值和便于保存。统计工作和研究过程中通常需要跟踪研究与动态分析,以便更好地揭示事物和现象的特征与规律,这就要求有长期的资料累积。为了积累资料,要求对已有的统计资料进行审核、筛选、分类汇总或者按历史的口径对现有的统计资料重新调整、分类和汇总等,也就是必须通过统计整理来实现。

3.1.2 统计整理的程序

统计整理的全过程包括统计整理方案设计、资料预处理、统计分组、资料汇总与图表编制五个环节，具体步骤如下。

1. 统计整理方案设计

统计整理方案设计是统计整理的首要步骤和依据，是保证统计整理有计划、有组织地进行的前提和基础，是统计设计在统计整理阶段的具体化。整理方案与调查方案应紧密衔接，统计调查获得了大量统计数据，尤其是人口普查和经济普查等大规模的统计调查获得数量庞大的统计资料，同时需要投入大量的人力、物力、资金和时间。为保证统计资料整理的质量，必须对统计整理各个阶段方方面面的工作进行全面统筹安排，设计科学合理的统计资料整理方案，其内容包括：①整理工作和程序安排；②资料审核、筛选、排序的要求；③资料的分组方法；④资料汇总方式与内容；⑤整理结果的表达内容与方式；等等，要保证整个整理方案内容及其实施的统一性，整理方案是否科学，对于统计整理乃至统计分析的质量都是至关重要的。

2. 资料预处理

资料预处理包括审核、筛选、排序等方面的处理。全面审核收集的资料，以确保统计资料符合统计研究的目的和要求，做到准确、及时、完整；然后筛选出符合条件的资料，剔除不符合要求的或有明显错误的资料；再按一定顺序排列资料，以体现事物和现象的基本特征、趋势或规律。

3. 统计分组

统计分组是统计整理的关键内容和统计分析的基础。根据研究的目的和统计分析的需要，选择合适的标志对资料进行划类分组，以有利于整理出有价值的综合指标，便于进一步揭示现象的本质与规律。

4. 资料汇总

在分组的基础上，将各项资料进行汇总，计算出各组总数和总体合计数，显现总体数量分布特征；得到反映各组 and 总体数量特征的各种指标。对整理的统计资料再次审核，及时纠正汇总过程中的各种差错。

5. 图表编制

统计资料汇总的结果通过统计表和统计图的形式展示出来，使得整理的资料简捷明了、系统有序，有利于统计数据资料的积累和应用。

3.1.3 统计整理的原则

对统计资料的加工整理需要遵循以下原则。

1. 区分事物的数量和品质特征

任何事物和现象都有数量与品质两个方面的特征。数量特征体现为事物及其特性在数量上的大小多少以及增加减少,同一事物可以有不同的量,是一种不显著、渐进的、不断发生的特征;品质特征是事物和现象的根本与属性,任何事物和现象都具有自己特殊的质的规定性。在统计整理时,要根据研究的目的和任务的要求,结合调查对象的特点,区分和把握事物的质的方面与量的方面的特征及其差别程度,存在品质差异的优先区分品质的差异。

2. 显示事物的整体特征

事物和现象具有多方面的特征且它们之间存在多方面的相互联系。把握事物和现象的每一方面的特征对于全面了解这一事物有重要的作用,不能只顾一方面而忽视另一方面。统计整理时,要尽可能全面显示事物各方面的特征,研究其全貌,描绘事物的整个发展过程,有利于揭示事物和现象的总体特征与规律性。

3. 体现事物的本质特征

事物和现象具有诸多的特征与属性,其中有一个方面或几个方面的特征是基本的、主要的、关键性的,是事物的本质特征;而其余的特征可能只有辅助的、次要的、补充的作用。统计整理必须在对事物和现象进行深刻研究的基础上,抓住最基本的、最能说明问题的本质特征资料进行加工整理。

3.2 资料预处理

资料预处理是对所收集资料进行分类或分组之前的审核、筛选、排序等必要的处理。

3.2.1 资料审核

统计整理中,为了保证统计资料的质量,首先要对资料进行审核,为进一步的统计整理与统计分析奠定基础。资料审核是为了发现资料中的错误并订正。从不同渠道取得的资料的内容与类型不同,不同类型的资料在审核内容和方法上存在差异。

1. 原始资料审核

对于原始资料主要从完整性、及时性和准确性三个方面进行审核。

(1) 完整性审核。完整性审核是对调查对象的完整性和项目的完整性的审核,审查调查单位或者填报单位是否有遗漏、调查项目或指标是否填写齐全、资料的份数是否符合规定等。

(2) 及时性审核。及时性审核是对需要整理的资料符合时间规定性的审核,如审查资料所属时间期限是否符合要求、调查工作是否按时间完成、资料报送时间是否符合规

定, 查清不报、漏报或迟报的现象等。

(3) 准确性审核。准确性审核是对填报的资料的真实性和精确性的审核, 主要审核资料是否真实地反映了客观实际情况、内容是否符合实际、资料是否有错误、计算是否正确等, 重点检查调查过程中所发生的误差。准确性审核主要有逻辑检查和计算检查两种方法。逻辑检查主要是从定性角度审核资料是否有悖理论或常识、数值是否符合逻辑、内容是否合理、各项目或数字之间有无相互矛盾的现象, 常用于对定类数据和定序数据的审核, 如某企业销售成本率显著低于同行业先进企业水平或利润总额大于同期销售收入可能不符合逻辑, 等等。计算检查主要检查各项指标的计算口径和计量单位是否符合规定, 各项数据在计算方法和计算结果上是否有错误, 常用于对定距数据和定比数据的审核, 如各分项或者总体单位数值之和是否等于相应的合计数, 各部分结构之和是否等于1或100%, 出现在不同表格上的同一指标数值是否相同, 等等; 也可以从指标间的相互关系以及指标的变动趋势来检查它的正确性; 对不能满足现在要求、缺漏或有疑问的资料, 要进行有科学根据的推算、弥补和订正。

2. 次级资料审核

对于次级资料, 在审核完整性和准确性的基础上, 要突出审核资料的适用性和时效性。审核资料的适用性是根据资料的用途, 检查资料解释说明问题的程度。由于次级资料可以来自多种渠道, 有些资料可能是为特定目的通过专门调查而取得的, 或者是已经按特定目的的需要做了加工整理, 应审核资料的可靠程度、指标含义、所属时间范围与空间范围、计算方法和分组条件与规定的要求是否一致。对于使用者来说, 具体需要弄清楚资料的来源、调查者收集资料的目的、资料的口径以及有关的背景材料, 以便判断资料的可靠程度、确定这些资料是否符合分析研究的需要, 是否需要重新加工整理, 等等, 不能盲目生搬硬套。

对资料的时效性进行审核, 对时效性较强的问题, 如果所取得的资料过于滞后, 就失去了研究的意义。一般来说, 应尽可能使用最新的统计资料。资料经过审核后, 确认适合实际需要, 才有必要做进一步的加工整理。

3. 资料订正

通过对资料的审核, 如发现有缺报、缺份和缺项等情况, 应及时催报、补报; 如有不正确之处应对资料进行订正, 应分别不同情况做如下处理。

- (1) 可肯定的一般错误, 应及时代为更正, 并通知原单位。
 - (2) 对于可疑之数或无法代为更正的错误, 应要求原单位复查更正。
 - (3) 若所发现差错在其他单位也可能发生, 应将错误情况通报所有单位, 防止类似错误。
 - (4) 对于严重错误, 应重新填报, 并查明错误原因, 对违法行为应依法严肃处理。
- 以上处理方式主要适合统计报表等通过报告法获取的资料; 若是通过采访法获取的资料, 方案尽可能介绍得详细、明确, 力争获得被调查者的全力支持, 必须在现场发现才可能订正。

3.2.2 资料筛选

资料筛选就是找出符合条件的资料或剔除不符合特定条件的资料。资料筛选在经济管理中十分重要。

对审核过程中发现的错误应尽可能予以纠正。调查结束后,当资料中发现的错误不能予以纠正,或者有些资料不符合调查的要求而又无法弥补时,就需要对资料进行筛选。资料筛选包括两方面内容:一是将某些不符合要求的资料或有明显错误的资料予以剔除;二是将符合某种特定条件的数据筛选出来,对不符合特定条件的数据予以剔除。

3.2.3 资料排序

资料排序是按一定顺序将资料进行排列。资料排序便于研究者通过阅读资料发现一些明显的特征或趋势,找到解决问题的线索;排序也有助于对资料检查纠错,为重新归类或分组等提供依据;有时排序本身就是分析的目的之一,有利于发现事物和现象在总体中所处地位与状况,便于有针对性地管理。例如,每年有多个机构发布企业排行榜,如世界 500 强企业排行榜、中国企业 500 强排行榜,根据排行榜不仅可以了解企业所处的地位、发现差距,还可了解到竞争对手的状况,从而有效制定企业管理战略与策略。

定类资料排序时,字母型资料可以选择升序或降序的排序方式,升序与字母的自然排列相同,更符合人们的习惯,升序使用得更普遍。汉字型资料排序方式很多,按汉字的首位拼音字母排列,这与字符型资料的排序完全一样;也可按笔画排序,根据笔画顺序或者笔画多少进行升序或者降序排序。交替运用不同方式排序,在汉字型资料的检查纠错过程中十分有用。数值型资料排序时,无论定距数据和定比数据的排序,都只有递增和递减两种排序方式,即按照数值大小从小到大或者从大到小排列。设一组资料为 $X_1, X_2, \dots, X_N$, 递增排序后可表示为: $X_1 < X_2 < \dots < X_N$; 递减排序可表示为: $X_1 > X_2 > \dots > X_N$ 。

排序后的资料也称顺序统计量,无论是品质资料还是数值型资料,排序均可借助计算机完成。

3.3 统计资料的整理

3.3.1 统计分组的概念与作用

1. 统计分组的概念

统计分组是统计整理的关键,统计分组就是根据事物与现象的特点和统计研究的任务,按照一个或几个标志把总体分成若干组成部分的统计方法。

总体中的各单位在某些方面具有同质性,在另外一些方面又具有变异性。统计分组的目的是把不同性质的单位分开,把性质相同的单位合在一起。因此,统计分组同时具有两方面的含义:对总体单位的“分”和对总体单位的“合”。“分”是将总体区分为性质不同的若干部分,就是把同一总体中性质不同的总体单位分配到不同的组里,突出组

间差异性;“合”是将性质相同的总体单位合在一起,就是把同一总体中性质相同的总体单位归并到同一组里,突出组内同质性。

统计分组时要体现穷尽互斥原则,分组的结果表现出组内同质性和组间差异性。也就是要使总体中的每一个单位都有组可归,同时使总体中的任何一个单位只能归属于某一个组,而不能同时或可能归属于多组。

2. 统计分组的作用

统计分组的作用包括以下几个。

1) 划分事物和现象的不同类型

统计分组的根本目的是将事物和现象区分为性质不同的若干类型,研究其特点和规律性。划分为不同类型是深入认识的前提,也是统计工作中应用最广泛、最主要的分组。如我国企业按企业所有制性质不同,可划分为公有制经济和非公有制经济,国有企业、集体企业、私有企业、外资企业、股份制经济、港澳台经济等(表3-2),分组后可以反映各类企业在国民经济中的地位和作用。统计分组时,总体单位之间的性质相同与否是相对的,是由统计研究的目的决定的。例如,研究某班学生的学习成绩,就“及格”和“不及格”而言,性质是不同的,60分以上的学生就是同质的;而就择优录用而言,在及格这一组学生中,又可以区分为优、良、中、及格四个不同性质的类型。

表 3-2 2017 年我国固定资产投资来源

产业	固定资产投资/亿元	增长速度/%
内资企业投资	604 550	7.7
港澳台商投资	13 562	-4
外商投资	11 322	2.7

2) 揭示事物和现象的内部结构

内部结构是指总体内各部分占总体的比重。将事物和现象按照某个标志分成若干组成部分,揭示总体内部的构成,表明部分与总体、部分与部分之间的关系。例如,人口统计中的各种年龄构成、国民经济中三次产业的构成等(表3-3)。

表 3-3 我国 2010—2017 年国内生产总值及其结构

年份	国内生产总值/万亿元	第一产业/%	第二产业/%	第三产业/%
2010	41.303 0	9.5	46.4	44.1
2011	48.930 1	9.4	46.4	44.2
2012	54.036 7	9.4	45.3	45.3
2013	59.524 4	9.3	44.0	46.7
2014	64.397 4	9.1	43.1	47.8
2015	68.905 2	8.8	40.9	50.2
2016	74.358 6	8.6	39.9	51.6
2017	82.712 2	7.9	40.5	51.6

3) 分析事物和现象之间的依存关系

一切事物和现象都不是孤立的, 相互之间不同程度地存在着互相联系、互相依存、互相制约的关系, 通过统计分组, 可以从数量上研究现象之间依存关系的规律性。通过统计分组, 可以揭示这种关系及其在数量上的表现, 表 3-4 所示为我国 2010—2017 年财政收入与财政支出的状况。

表 3-4 我国 2010—2017 年财政收入与财政支出的状况 亿元

年份	2010	2011	2012	2013	2014	2015	2016	2017
财政收入	83 102	103 874	117 254	129 210	140 370	152 269	159 605	172 567
财政支出	89 874	109 248	125 953	140 212	151 786	175 878	187 755	203 330

3.3.2 分组标志的种类及选择

1. 分组标志的种类

1) 简单分组、复合分组、并列分组

按分组标志的多少和形式, 分组可分为简单分组、复合分组和并列分组。

(1) 简单分组。简单分组就是对总体只按一个标志进行分组。例如, 货物量按运输方式分为铁路运输、公路运输、水路运输、航空运输与管道运输五组。

(2) 复合分组。复合分组就是对总体按两个或两个以上标志进行重叠式分组。即在按某一标志分组的基础上再按另一标志进一步分组。如学校招生数量分配, 首先按生源所在各省市自治区分配, 其次在生源地按招生专业分配, 再次按学生高中阶段文理科选择, 最后看专业是否有性别限制。

(3) 并列分组。并列分组也称平行分组, 就是对总体同时用两个或两个以上的标志分别从不同的角度进行不重叠的多种分组。其特点是两种或多种分组相互独立而不重叠, 既可从不同的方面反映事物的多种结构又不致过于烦琐。如对企业按所有制性质、地域、规模和行业等分组。

2) 品质标志分组和数量标志分组

按分组标志的性质, 分组可分为品质标志分组和数量标志分组。

(1) 品质标志分组。品质标志分组就是按事物和现象的品质特征分组。居民消费按商品类别分为食品烟酒类、衣着类、生活用品及服务类、教育文化和娱乐类、交通和通信类、医疗保健类、居住类、其他用品和服务类八大类。这种分组可以反映总体的构成和不同性质事物在总体中的地位与作用。

(2) 数量标志分组。数量标志分组就是按事物的数量特征分组。选择反映事物数量差异的数量标志作为分组标志, 并在数量标志的变异范围内划定各组界限, 将总体划分为性质不同的若干组成部分。将人按身高、体重、年龄分组, 将学生按考试成绩分组都是数量标志分组, 企业按生产能力、劳动生产率分组, 商店按商品流转额、职工人数分组, 等等。这种分组的目的在于通过事物和现象在数量上的差异来反映事物在性质上的

区别。

3) 类型分组、结构分组和分析分组

按分组标志的作用和任务不同, 分组可分为类型分组、结构分组和分析分组。

(1) 类型分组。类型分组就是把复杂的现象总体划分为若干个不同性质的部分。通常按品质标志分组, 以划分社会经济类型。

(2) 结构分组。结构分组就是在对总体分组的基础上计算出各组对总体的比重以研究总体各部分的结构。通常按数量标志分组, 以研究同类总体的结构。类型分组和结构分组往往紧密联系在一起。

(3) 分析分组。分析分组就是通过分析, 研究事物或现象之间的相互关系而进行的分组。分析分组的分组标志称为原因标志, 与原因标志相对应的标志称为结果标志, 如影响某种商品消费需求的因素有该商品的价格、消费者收入、相关商品的价格、消费者偏好以及消费者对该商品的预期等。原因标志不同, 结果标志不同; 同一原因标志由于分组的不同, 结果标志也会不同。例如, 工人的劳动生产率与产值之间、商品流通过费用率与商品销售额之间的依存关系, 都可以按分析分组法来研究。

2. 分组标志的选择

分组标志是总体分组的标准或依据, 分组标志一经选定, 必将突出总体单位在此标志下的差异, 而将总体单位在其他标志下的差异掩盖起来。因此, 分组标志选择得恰当与否, 直接关系到能否正确反映总体内部的性质特征, 在实际工作中应遵循一些原则, 以确保正确选择分组标志。

1) 根据研究的目的和任务选择分组标志

统计分组是为统计研究服务的, 统计研究的目的和任务不同, 选择的分组标志也应有所不同。如研究人口的年龄构成时, 就应该按“年龄”分组; 研究各类型的工业企业在工业生产中的地位和作用时, 就应该按“经济类型”分组。每一总体都可以按照许多标志进行分组, 具体按什么标志分组, 主要取决于统计研究的目的和任务。如同是以工业部门为研究对象, 当研究的目的是分析工业部门中各种规模的企业生产情况时, 应该选择产品数量或生产能力作为分组标志; 当研究目的在于确定工业内部比例及平衡关系时, 应该以行业为分组标志, 将工业部门划分为重工业与轻工业或冶金、电力、化工、机械、纺织、煤炭等工业行业。

2) 适应被研究对象的特征来选择分组标志

社会经济现象纷繁复杂, 研究某一问题可能涉及许多标志, 科学的统计分组则应从中选择与统计研究的目的、有关事物的性质或类型关系最密切的标志, 即将最主要或最本质特征的标志作为统计分组的依据。有时在同一研究目的下, 可能有几个标志似乎都可以达到此目的, 对比应该进行深入分析, 选择主要的、能反映问题本质的标志进行分组; 对于有些现象进行分组时, 使用一个分组标志不足以区分事物的不同性质与特点, 不能全面地认识事物的变化规律, 分组时, 除了使用一个主要分组标志以外, 还要用一个或几个辅助标志作为分组补充标志。哪些标志作为主要标志, 哪些标志作为辅助标志, 这要根据研究任务来选择与确定。如研究学生学习情况时可能专业、年级、性别等是最

常用的分组标志，而分析已就业人员可能职业、文化程度、收入水平等是最常用的分组标志。

3) 结合事物和现象所处的历史条件来选择分组标志

客观事物的特点和内部联系随着条件的变化而不同，因此选择分组标志时，要具体情况具体分析，根据事物的不同条件来选择分组标志。有时有的标志在当时能反映问题的本质，但由于社会经济的发展变化，可能时过境迁，这时就要选择新的标志进行分组。例如，同样是划分企业规模，在劳动密集型行业，可采用职工人数作为分组标志；而在技术密集型行业，则应选择固定资产价值或生产能力作为分组标志。

3.3.3 统计分组参数的确定

1. 组数的确定

组数的确定包括以下几个方面。

1) 品质分组的组数

品质分组的组数取决于统计研究的任务、事物和现象的特点两个因素。有些事物的属性就决定了总体的组数；有些事物构成较复杂，组数可多可少，其组数还需要根据统计研究的任务来定。例如，按学科大类可分为自然科学和社会科学；按学科门类可以分为哲学、经济学、法学、教育学、文学、历史学、理学、工学、农学、医学、管理学、军事学和艺术学 13 个，其中理、工、农、医 4 个门类属于自然科学；继续细分为 110 个一级学科和更多的二级学科。具体选用哪个分类、分成多少组要考虑研究的任务。由于品质只能采用定类或定序两种计量尺度，只能对事物和现象按平行分类与测度等级、顺序，不能测度只有数量标志才有的间距，所以品质分组没有组距。

2) 数量分组组数和组距的确定原则

数量分组组数和组距的确定坚持以下几个基本原则。

(1) 根据统计研究的目的和任务，使分组的结果能满足统计工作的目的和要求。例如，人口按年龄分组：劳动统计、现代教育统计、人口再生产统计等分组方法各不同。

(2) 注意决定事物数量的数量界限，尽可能使各组的数量差异反映出事物在性质上的区别。分组时要尽可能保证组内的同质性和组间的差异性，使数据大小的差异与事物性质上的差异相吻合。如考试成绩按 100 分制时，一般将 60 以下、60~70、70~80、80~90、90~100 分别确定为不及格、及格、中、良、优五个等级，体现成绩的五个不同性质的区间。企业规模根据不同行业的特点，分别选用营业收入、从业人数、资产总额等标志，确定一定的数量界限，将企业划分为大型、中型、小型、微型四类（表 3-5）。

(3) 要体现出事物分布的特点、反映事物分布的规律，便于人们对总体中的特殊部分进行单独的研究。如在正常情况下，一个班级同学的成绩应该接近正态分布，我们将成绩按优、良、中、及格、不及格分为五个数量区间后应该也呈现高分和低分的人数较少，多数人处于或接近中间水平。

表 3-5 我国部分行业大中小微企业划分标准

行业名称	指标名称	计量单位	大型	中型	小型	微型
农林牧渔业	营业收入 Y	万元	$Y \geq 20\,000$	$500 \leq Y < 20\,000$	$50 \leq Y < 500$	$Y < 50$
工业	从业人员 X	人	$Y \geq 1\,000$	$300 \leq X < 1\,000$	$20 \leq X < 300$	$X < 20$
	营业收入 Y	万元	$Y \geq 40\,000$	$2\,000 \leq Y < 40\,000$	$300 \leq Y < 2\,000$	$Y < 300$
建筑业	营业收入 Y	万元	$Y \geq 80\,000$	$6\,000 \leq Y < 80\,000$	$300 \leq Y < 6\,000$	$Y < 300$
	资产总额 Z	万元	$Z \geq 80\,000$	$5\,000 \leq Z < 80\,000$	$300 \leq Z < 5\,000$	$Z < 300$
批发业	从业人员 X	人	$X \geq 200$	$20 \leq X < 200$	$5 \leq X < 20$	$X < 5$
	营业收入 Y	万元	$Y \geq 40\,000$	$5\,000 \leq Y < 40\,000$	$1\,000 \leq Y < 5\,000$	$Y < 1\,000$
零售业	从业人员 X	人	$X \geq 300$	$50 \leq X < 300$	$10 \leq X < 50$	$X < 10$
	营业收入 Y	万元	$Y \geq 20\,000$	$500 \leq Y < 20\,000$	$100 \leq Y < 500$	$Y < 100$
住宿和餐饮业	从业人员 X	人	$X \geq 300$	$100 \leq X < 300$	$10 \leq X < 100$	$X < 10$
	营业收入 Y	万元	$Y \geq 10\,000$	$2\,000 \leq Y < 10\,000$	$100 \leq Y < 2\,000$	$Y < 100$
软件和信息 技术服务业	从业人员 X	人	$X \geq 300$	$100 \leq X < 300$	$10 \leq X < 100$	$X < 10$
	营业收入 Y	万元	$Y \geq 10\,000$	$1\,000 \leq Y < 10\,000$	$50 \leq Y < 1\,000$	$Y < 50$
房地产开发 经营	营业收入 Y	万元	$Y \geq 200\,000$	$1\,000 \leq Y < 200\,000$	$100 \leq Y < 1\,000$	$Y < 100$
	资产总额 Z	万元	$Z \geq 10\,000$	$5\,000 \leq Z < 10\,000$	$2\,000 \leq Z < 5\,000$	$Z < 2\,000$

3) 数量分组组数的确定

数量分组组数必须适中，若组数太少，数据的分布就会过于集中，而组数太多，数据的分布就会过于分散，这些都不便于观察数据分布的特征和规律。组数的多少应根据统计研究的任务、事物分布的特点来定，也与数据本身的特点及数据的多少有关。由于分组目的之一是观察数据分布的特征，因此组数的多少应适中，组数的确定应以能够显示数据的分布特征和规律为目的。在实际分组时，若有规定或惯例，可以根据或参照惯例或规定来确定组数；否则在 Sturges（斯特杰斯）提出的各总体单位标志值趋于正态分布的情况，根据总体单位数 N 来确定组数（ n ），其经验公式为

$$n=1+\frac{\lg N}{\lg 2}=1+3.322 \lg N$$

(3-1)

对结果用四舍五入的办法取整数即为组数，如表 3-6 所示。

表 3-6 总体单位数与组数对应关系

N	15~24	25~44	45~89	90~179	180~359	360~719	720~1 439
n	5	6	7	8	9	10	11

当然，这只是一个经验公式，仅供参考，不能生搬硬套，在标志值变动范围小、分布比较均匀、适合等距分组情况下可采用效果最佳。实际应用时，可根据数据的多少和

特点及分析的要求,参考这一标准灵活确定组数。

2. 组距的确定

组距的确定包括以下两点。

1) 组距的大小

组距是每组中最大变量值与最小变量值之间的距离或差数。组距的计算公式为

$$\text{组距} = \text{本组最大标志值} - \text{本组最小标志值}$$

全距一定的情况下,组距与组数呈反比关系。若组距大,则组数就少;若组距小,则组数就多。若是各组组距相等,组数可根据全总体最大值和最小值及组距来确定,表示为

$$\text{组数} = \text{全距} \div \text{组距} \quad (3-2)$$

其中,全距=最大标志值-最小标志值。

组数与组距的确定,原则上应该力求符合事物和现象的实际情况,能够将总体分布的特点反映出来。如果组距过小,组数过多,容易将同质的单位划分在不同的组内,显示不出资料类型的特征;如果组距过大,组数过少,会使不同性质的单位同处一组,掩盖质的差异,因此必须科学地确定组数和组距。

2) 组距形式的选择

组距的形式有等距和不等距(也称异距)两种。等距就是标志值在各组保持相等的组距,而异距则是标志值在各组的组距不相等。选用哪种形式,要根据统计研究的目的和资料的性质、特点而定。

组距的形式应服从分组的组内同质性和组间差异性的要求,做到性质相同的单位应合并在一个组内,性质不同的单位应当分开。事物和现象的差别取决于其本质,而不取决于数学形式,必须根据事物和现象的本质特征与统计研究的目的任务来确定分组的等距与否。

等距分组的各组的标志值变动都限于相同的范围,凡是在标志值变动比较均匀或不同总体单位标志值相对比值较小的情况下,都可采用等距分组。例如,工人的年龄、工龄、工资的分组;零件尺寸的误差、加工时间的分组;农产品单位面积产量、单位产品成本的分组;等等。等距分组有很多好处,它便于绘制统计图,也便于进行各类运算。

不等距分组在下列情况下采用比较合适:第一,标志值相对比值较大的场合。如某地区不同百货商场营业额差别是很大的,营业额从50万元至5亿元,若用等距分组,即使组距为1000万元,也得分为50组,显然是不合适的。可采取公比为10的不等距分组,组距分别为50万~500万元、500万~5000万元、5000万~50000万元。第二,标志值分布很不均匀的场合。例如,一个企业职工月工资分布范围较宽,低的只有3000多元,高的达到数万元,但是多数职工的月工资密集于6000~9000元,其他部分则分布较为稀少,在这种场合若以1000元作为组距进行等距式分组,则无法显示出分布的规律性,会使得这一密集的分段分布的信息损失过大。因此,合理的做法是,在分布比较密集的区间内使用较短的组距,在分布比较稀少的部分使用较长的组距,形成不等距分组。第三,标志值相等的量具有不同意义的场合。例如,生命的每一个月对于新生儿和对于成年人是大不一样的,若对人的不同阶段的饮食研究进行年龄分组,应采用不等距分组,即1岁以下按月分组,1~10岁按年分组,11~20岁按2~5年分组,21岁

以上按 10 年分组，等等。

不等距分组没有可以依循的固定模式，需要统计人员在实践中不断探索，关键在于对所研究的事物和现象的内在联系必须十分熟悉，才能很好地运用不等距分组来揭示事物的本质。

组距式分组的假定条件是：变量在各组内的分布都是均匀的，或者各组标志值呈线性变化。通过组距式分组，突出各组之间的主要差异，抽象各组内部各单位的次要差异，各组分配的规律性更容易显示出来。组距式分组的假设与实际分布存在差异，使分组后的资料真实性受到一定程度的损害。分组过多、组距太小则容易将属于同类的单位划分到不同组，显示不出一些类型的特点；组距太大，组数太少，会把不同性质的单位归并到同一组中，失去区分事物的界限，达不到正确反映客观事实的目的。因此，组距的大小、组数的确定应考虑研究对象的经济内容和标志值的分散程度等因素，不可强求一致。

3. 组限的确定

组限是每组区间中的极端值。每组的起点数值或最小值称为下限，每组的终点数值或最大值称为上限。如果各组的组限都齐全，称为闭口组；组限不齐全，即最小组缺下限或最大组缺上限，称为开口组。从根本上说，组限应是区分事物质的差别的数量界限。组限确定得好，就能充分体现分组的功能，分清组与组之间的差别；否则，就有可能混淆现象之间的本质区别。

1) 组限的形式

组限的形式分为重合式组限和不重合式组限，组限的形式由变量的特点决定。分组标志连续变量适合重合式组限，不重合式组限适合的分组标志是离散变量。

重合式组限就是相邻两组中前一组的上限与后一组的下限数值相重。重合式组限只是形式上相重，两组的实际数值并没有重复，数据采用组距分组时遵循“不重不漏”原则：“不重”是指一项数据只能分在其中的某一组，不能在其他组中重复出现；“不漏”是指在所分的全部组别中每项数据都能分在其中的某一组，不能遗漏。在重合式组限中，为解决“不重”的问题，对于重合的那个组限，习惯上按“含下限不含上限”或者“上限不在本组之内”的原则处理，当相邻两组的上下限重叠时，恰好等于某一组上限的变量值不算在本组内，而归入作为下限值的那一组内。对于连续性变量，其变量值有小数，组限不能肯定的，数值表示只能用前一组的上限与后一组的下限重叠的方法表示。例如，在表 3-7 所示的分组中，10 这一数值不计算在“8~10”这一组内，而计算在“10~12”组中，其余类推。对于连续变量，也有可以对一个组的上限值采用小数点的形式，小数点的位数根据所要求的精度具体确定，如某产品零件尺寸可以分组为 2~3.99, 4~5.99, 6~7.99, 等等。

不重合式组限就是相邻两组中前一组的上限与后一组的下限数值紧密相连而不重复。变量值只是在整数之间变动。例如，企业数、职工数、机器设备台数等离散变量，可采用不重合式组限。

实际中，有分组标志是连续变量，组限却根据习惯采用不重合式组限。变量值属于离散变量，可采用不重合式组限，也可采用重合式组限。如果变量值在一定范围内的

表 3-7 某产品构成零件尺寸分布

零件尺寸/厘米	零件个数/个	比率/%
2~4	2	4
4~6	5	10
6~8	8	16
8~10	15	30
10~12	9	18
12~14	7	14
14~16	4	8
合计	50	100

表现既可以是整数，也可以是小数，如产值、身高、体重等连续变量，只能采用重合式组限。

2) 组限的确定

确定组限遵循组内同质性、组间差异性，能使性质相同的单位归入同一组内，使不同性质的单位划分在不同组别。为了达到上述要求，组限的确定应注意如下几点。

(1) 连续变量和离散变量的组限确定方法不同。连续变量的相邻组组限必须重叠，在计算各组单位数或标志值时，按“上组限不在内（不含上限）”的原则；离散变量以自然单位作为计量单位，相邻组组限可以断开，若计量单位是自然单位的扩大倍数，相邻组组限可以重叠。

(2) 总体各单位标志值中出现特大变量值或特小变量值时，最低组和最高组可采用开口式组限。在组距分组中，如果数据中的最大值和最小值与其他数据相差悬殊，也就是出现了极端值，为避免出现没有变量值的空白组或个别极端值被漏掉，第一组和最后一组可以采用“××以下”及“××以上”这样的开口组形式。

(3) 组限的确定应有利于表现出总体分布的特点和规律性。组限应是决定事物性质的数量界限，应反映出事物质的变化。按学生考试成绩分组，60分必须作为组限，因为它是及格与不及格的界限；按计划完成程度分组，100%必须作为组限，因为它是完成还是未完成计划的界限。

(4) 最小组下限应低于总体各单位标志值中最小变量值，最大组上限应高于总体中最大变量值，但不应过于悬殊，一般小于最小值或大于最大值不超过半个组距。

(5) 为了方便计算组限应尽可能取整数，尾数最好是5或10。如果组限是小数，要根据情况将组限数值适当缩放，其数值最好为各单位标志值的最小计量单位或最大公因数的5倍或10倍，这样比较符合习惯。

4. 组中值

组中值是上下限之间的中点数值，用以代表各组标志值的一般水平，具有平均指标的性质。组中值不是各组标志值的平均数，各组标志值的平均数在统计分组后很难计算出来，就常以组中值近似代替，组中值的这种计算方法是假定各组标志值的变化是均匀

的,恰好与组距式分组的假定条件相同。组中值仅存在于组距式分组数列中,单项式分组中不存在组中值。

不同的组限表现形式采用的计算公式略有不同。

1) 重合式组限

对于重合式组限,组距与组中值按以下方法计算:

$$\text{组距} = \text{本组上限} - \text{本组下限} \quad (3-3)$$

$$\text{组中值} = (\text{上限} + \text{下限}) \div 2 = \text{下限} + \text{组距} \div 2 = \text{上限} - \text{组距} \div 2 \quad (3-4)$$

2) 不重合式组限

对于不重合式组限,组距与组中值按以下方法计算:

$$\text{组距} = \text{下组下限} - \text{本组下限} = \text{本组上限} - \text{上组上限} \quad (3-5)$$

$$\text{组中值} = (\text{本组下限} + \text{下组下限}) \div 2 = \text{本组下限} + \text{组距} \div 2 = \text{下组下限} - \text{组距} \div 2 \quad (3-6)$$

根据需要不重合式组限的各组上限加1,将不重合式组限转化为重合式组限,不重合式组限的组中值计算公式与重合式组限统一。

3) 开口组

对于第一组是“多少以下”,最后一组是“多少以上”的开口组,其只有一个组限,组中值的计算可参照邻组的组距来决定。即

(1) 第一组“多少以下”(有上限无下限),组中值 = 上限 - 邻组组距 $\div 2$ 。

(2) 最后一组“多少以上”(有下限无上限),组中值 = 下限 + 邻组组距 $\div 2$ 。

3.3.4 统计资料汇总

1. 统计资料汇总的概念

统计资料汇总是在统计分组的基础上,将统计资料归并到各组中去,并计算各组和本体的合计数的过程。统计资料汇总使我们能看到事物和现象的全体,进而提示出总体在多方面的数量特征。

2. 统计资料汇总的组织形式

统计资料汇总的组织形式分为逐级汇总和集中汇总。

1) 逐级汇总

逐级汇总就是按照一定的统计管理系统,由各级统计机构自下而上地逐级将调查资料汇总上报。如定期统计报表。其便于审核和订正统计调查资料,满足各级对资料的需要,但是逐级汇总层次较多,所需时间较长,产生汇总差错可能性较大。

2) 集中汇总

集中汇总是指把统计调查资料集中在组织调查的最高机构或它指定的机构进行汇总。其特点就是不经中间环节。这样可以大大缩短汇总时间,便于贯彻统计的汇总纲要,可使用现代化汇总手段来提高汇总效率和质量。但是不能及时满足地方或基层领导的需要,审核和订正资料较困难。比较适合时效性强的快速普查和对汇总质量要求很高的一些重要调查。

3. 统计资料汇总的方法

统计资料汇总的方法有手工汇总和计算机汇总两种。

1) 手工汇总

手工汇总指以算盘和小型计算器为手段,通过手工操作对统计资料进行汇总。实际中常用的方法有以下四种。

(1) 点线法。点线法也叫画记法,通过画点线来汇总各组 and 总体单位数。常用“正”“※”等点线符号。这种汇总方式简便易行,尤其当总体单位数不大时。

(2) 过录法。过录法就是先将统计资料过录到整理表上,计算各种合计数,然后再按整理表汇总。这种汇总方式工作量大,可能产生转录差错;可计算总体单位数,也可计算标志总量,适用调查单位不是很多的情况。

(3) 折叠法。折叠法就是把所有调查表的相同项目和数值折在边上,一张一张错叠起来,露出需要汇总的项目和数值,然后汇总。这种汇总方式以表汇表,避免了过录,很简便;但是若有差错难以查找。

(4) 卡片法。卡片法就是在汇总大量统计资料时,首先将每一调查单位的资料摘录到一张特制的卡片上,然后用卡片进行分组汇总。基本过程如下:一是编号。根据分组要求,按分组标志(部门、地址、主管系统、企业规模)将所有调查单位的调查表上编上各种组号。二是摘录。将调查表上注明的“组号”和“标志值”摘录到卡片上。三是分组计数。根据分组要求,将卡片按组号分组,计算每组的单位总数(卡片张数)和标志总量,并填写在表中。卡片法适用于总体单位数很多。有复合分组的情况。它兼有过录法、折叠法的优点,便于保存和查考资料。

2) 计算机汇总

统计部门正日益广泛地采用计算机进行统计资料的汇总。计算机汇总具有速度快、精确度高、能自动纠正错误的特点。计算机数据处理包括对原始数据的加工、存储、合并、分类、逻辑检查、运算以及打印出汇总表或图形等。其全部过程大体上分为五个步骤。

(1) 编程序。编程序就是使用计算机高级语言将统计汇总分成一个个步骤,编成一条条指令,把高级语言源程序编译成计算机可执行的目标程序。

(2) 编码。编码就是把表示统计信息的某种符号体系转换成便于计算机或人识别和处理的另一种符号体系的过程。编码的质量不仅影响数据录入的速度和质量,而且影响统计资料处理的最终结果。

(3) 统计资料录入。统计资料录入就是把经过编码后的统计信息由录入人员通过录入设备记载到存储介质上,如软磁盘、磁带、纸带等。

(4) 统计资料编辑。统计资料编辑就是按照事先规定的一套编辑规则对输入计算机的原始数据进行分析、比较、筛选、整理等,使编辑后的全部统计资料符合编制规则的要求。

(5) 制图表、打印。制图表、打印就是对经过编辑的统计资料,执行目标程序,形成各种形式的统计图表,并把所需的统计资料、统计图表打印出来。

3.4 统计整理结果的显示

3.4.1 分布数列

1. 分布数列的概念与构成

分布数列是在统计分组与汇总的基础上，将总体中的所有单位按一定标志分组整理，并将各组按一定顺序排列，形成总体中各个单位在各组中的分布状况的统计数列，也称次数分布、分配数列。分布数列是统计整理的一种重要形式，也是统计描述和统计分析的一种重要方法，它可以表明总体的分布特征、内部结构，并据此研究总体中某一标志的平均水平及其变动的规律性。

分布数列有各组名称和各组次数两个构成要素。各组名称就是按一定标志将总体划分出来的各组，其中数量标志分组表现为各组的变量值，用 x 表示。各组次数就是各组所对应的单位数，也称频数，用 f 表示；次数的相对数形式即各组次数占总体单位数的比重，称为频率，用 $\frac{f}{\sum f}$ 表示，说明具有某组标志值在总体中出现的频繁程度或对总体水平所起的作用程度，反映了总体的构成。频率具有如下两个性质：各组频率都是介于 0 和 1 之间的一个分数，即 $0 < \frac{f}{\sum f} < 1$ ；各组频率之和等于 1，即 $\sum \frac{f}{\sum f} = 1$ 。分布数列通常表现为次数分布表和次数分布图两种形式。

2. 分布数列的种类

1) 品质数列和变量数列

按分组标志的不同，分布数列可分为品质数列和变量数列。

(1) 品质数列。品质数列就是按品质标志分组所形成的分布数列。如人按照民族、工种、职业、专业、性别、职称等分组，表 3-8 所示就是品质数列。对品质数列，若分组标志选择得好，标准恰当，则事物性质差异表现得比较明确，总体中各组如何划分易解决。

表 3-8 2017 年我国人口性别构成

性别	人口数/万人	比重/%
男	71 137	51.2
女	67 871	48.8
合计	139 008	100.0

(2) 变量数列。变量数列就是按数量标志分组所形成的分布数列。变量数列由变量和次数两个要素组成，其组限表现为不同的数值或变量区间，如表 3-9 所示。变量数列因事物性质差异表现得不明确，决定事物性质的数量界限往往因人主观认识而异，因此按同一数量标志分组时，可能出现多个不同的变量数列。

表 3-9 我国 2017 年人口年龄构成

人口年龄分组/岁	人口数/万人	比重/%
0~15	24 719	17.8
16~59	90 199	64.9
60 及以上	24 090	17.3
合计	139 008	100.0

按分组形式不同, 变量数列分为单项式数列和组距式数列。单项式数列是指各组都由一个具体的变量值来表示的数列。单项式数列中每个组的变量值均只有一个, 即一个变量值代表一组, 一般适用于离散变量且变量值变动范围不大的场合, 或习惯用离散形式表示的连续变量。例如, 家庭按人数分组可分为 1, 2, 3, …若干组, 如表 3-10 所示。

表 3-10 我国第六次人口普查每个家庭人数分布

每个家庭人数	家庭数/户	频率/%
1	58 396 327	14.53
2	97 947 686	24.37
3	107 978 654	26.86
4	70 598 493	17.56
5	40 332 512	10.03
6	16 887 554	4.20
7 以上	9 792 970	2.45
合计	401 934 196	100.00

组距式数列是变量值按一定的变化范围或间距进行分组形成的变量数列。变量数列中每个组用一个变量值区间表示, 见表 3-9。适用于按连续变量分组或变量值的变动范围较大、变量的不同取值个数较多的离散变量分组的情况。按组距形式, 组距式数列分为等距数列和不等距数列。

2) 钟形分布数列、J 形分布数列和 U 形分布数列

按次数分布的特征不同, 分布数列可分为钟形分布数列、J 形分布数列和 U 形分布数列。

(1) 钟形分布数列。钟形分布数列的特征是“两头小、中间大”, 即靠近中间的变量值分布的次数多, 靠近两端的变量值分布的次数少, 如果将变量值与其对应的频数在直角坐标系中对应的点连接起来绘制成曲线图, 宛如一口钟, 所以又称钟形分布。如人体体重、身高, 学生的成绩, 居民货币收入, 单位面积的农产品产量, 市场价格等多数社会经济现象都属于钟形分布。钟形分布有正态分布和偏态分布两种形式, 其中正态分布(图 3-1)是一种对称的钟形分布, 它以变量值中点为对称轴, 两侧变量值分布的次数随着与中点值距离增大而渐次减少, 且减少的次数基本对等, 如农作物的单位面积产量、零件的公差、纤维强度等很多现象都服从正态分布。偏态分布(图 3-1)是一种非对称钟形分布, 按其图形偏斜方向不同分为左偏态分布和右偏态分布, 长尾拖向右侧(变量值较大的一侧)、左侧偏短的偏态分布称为右偏态分布或正偏态分布; 长尾拖向左侧(变

量值较小的一侧)、右侧偏短的偏态分布称为左偏态分布或负偏态分布。

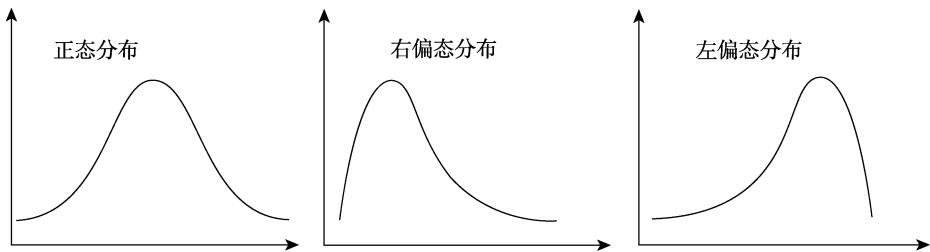


图 3-1 钟形分布形态

(2) J 形分布数列。J 形分布的特征是“一边小，一边大”，即大部分变量值集中在某一端分布（图 3-2），有正 J 形分布和反 J 形分布两种类型。正 J 形分布是次数随着变量值的增大而增多，如市场经济条件下，商品供应量随市场价格上升而增加，一般呈现正 J 形分布。反 J 形分布是次数随着变量值的增大而减小。如成年人数量按年龄大小分组，表现出年龄越高，人数越少；弹性大的商品，其需求量随市场价格上升而减少。

(3) U 形分布数列。U 形分布（图 3-2）是靠近中间的变量值分布的次数少，靠近两端的变量值分布的次数多，呈现“两头大、中间小”的 U 形分布特征。许多有寿命的客观事物或现象故障率或死亡率均呈现 U 形分布，如动植物的死亡率和机器设备故障率、失业人口按年龄的分布等。

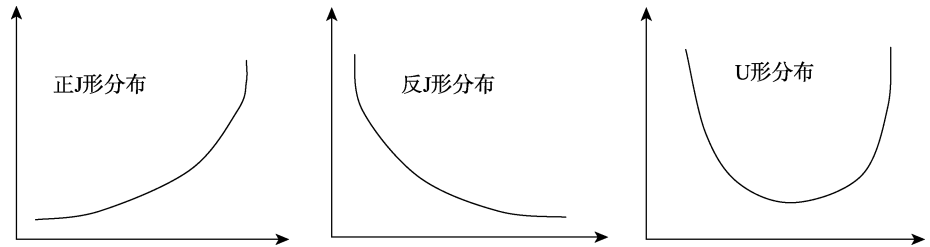


图 3-2 J 形分布和 U 形分布形态

3. 分布数列的编制

次数分布有简单次数分布和累计次数分布之分，通常用次数分布表和次数分布图作为表现形式。不同类型的资料，采取的处理方式和方法不同。品质数列主要是做分类或顺序整理，变量数列主要是做分组整理。将统计资料按计量尺度的精确程度从低到高层次顺序分为定类资料、定序资料、定距资料和定比资料，适合于低层次资料的整理和显示方法也适合于高层次的资料；但适合于高层次资料的整理和显示方法并不总能适合于低层次的资料。

1) 品质分布数列的编制

品质分布数列是按品质标志分组的数列，用来观察总体单位中不同属性的单位分布情况，如表 3-11 所示。

表 3-11 我国 2017 年货物运输周转量

亿吨·千米

运输方式	货物运输周转量	频率
铁路	26 962.2	13.75
公路	66 712.5	34.01
水运	97 455.0	49.69
民航	243.5	0.12
管道	4 757.2	2.43
合计	196 130.4	100.00

品质数列整理的基本过程：一是选择分组标志。品质数列的编制比较简单，只需将品质标志的表现一一排列出来，然后汇总出每一种标志表现出现的次数即可。最关键的是确定分组标志，要遵循分组标志选择的原则。二是列出各类别。要注意，分组时应包括分组标志的所有表现，不能有遗漏，各种表现相互独立，不得相融；特别要关注按品质标志分组时，确定各组的界限有两种情况：一种组限是自然形成的或比较明显的，如人口按性别、文化程度、党派分组等。另一种组限存在属性之间的过渡形式，使分组界限难以确定。这种比较复杂的品质标志分组，国家有关部门都制定有标准的分类目录，分组时可以依据分类目录来确定组限。例如，人口按职业分组，企业按行业分组，产品按经济用途分组，等等。三是计算各类别的频数或频率。四是编制频数分布表。五是用图形显示数据。

2) 变量数列的编制

变量数列的编制按如下步骤进行。

(1) 原始资料的预处理。先对资料进行审核，若发现资料中的错误，就进行订正；然后筛选，找出符合条件的资料，将不符合特定条件的资料予以剔除；再排序，按标志值从小到大进行排序，确定其最大值、最小值，并计算全距。

(2) 确定变量数列的形式。变量数列有单项式数列和组距式数列两种。组距式数列适用于按连续变量分组或变量值的变动范围较大、变量的不同取值个数较多的离散变量分组的情况。在组距式数列的两种形式即等距数列和不等距数列之间选择。如果选择单项式数列的形式，则直接进入第(4)步。

(3) 确定组距式数列的参数。编制组距式数列时，基本参数组数、组距、组限需要确定。一是确定组数。根据统计研究的目的和要求，优先按照规定或惯例确定组数；没有广泛认可的确定组数规定或惯例，其组数 K 的确定借用 Sturges 提出的经验公式。二是确定组距。若选用等距，组距根据全部数据的最大值和最小值及所分的组数来确定，即组距 = (最大值 - 最小值) / 组数。得到的组距数值要适当缩放，其数值最好为最小计量单位或最大公因数的 5 倍或 10 倍。三是确定组限。先选择组限的形式，再确定组限第一组下限即组限的起点确定很关键，一般可采用 $X_{1下} = \text{最小值} - \text{组距}/2$ ，然后对组限数值适当缩放，其数值最好为各单位标志值的最小计量单位或最大公因数的 5 倍或 10 倍。

(4) 统计各组的频数。在确定各组组限的基础上，遵循“不重复，不遗漏”的原则，将数据归并到各组中，计算各组 and 总体的合计数。

3) 累计次数分布表的编制

累计次数分布表是用统计表来表示频数分布，列入累计分布频数和累计频率。累计频数（或频率）可以是较小制累计频数（或频率），也可以是较大制累计频数（或频率）。

（1）较小制累计。较小制累计也称向上累计。较小制累计频数（或频率）分布，先列出各组的上限，从最小组开始，然后由标志值低的组向标志值高的组依次累计各组频数或频率。某组较小制累计频数表明该组上限以下的各组单位数之和是多少，即表示小于该组上限的频数合计数；某组较小制累计频率表明该组上限以下的各组单位数之和占总体单位数的比重，即表示小于该组上限的频率合计数。

（2）较大制累计。较大制累计也称向下累计。较大制累计频数（或频率）分布，即先列出各组的下限，从最大组开始，然后由标志值高的组向标志值低的组依次累计各组频数或频率。某组较大制累计频数表明该组下限以上的各组单位数之和是多少，表示大于该组下限的频数合计数；某组较大制累计频率表明该组下限以上的各组单位数之和占总体单位数的比重，表示大于该组下限的频率合计数。

现以 2017 年某专业 16 级 65 名同学统计学期末考试卷面成绩为例，分别进行较小制累计和较大制累计，其结果如表 3-12 所示。

表 3-12 某专业 16 级 65 名同学统计学期末考试卷面成绩累计次数分布

成绩	学生/人	频率/%	较小制累计		较大制累计	
			累计频数	累计频率/%	累计频数	累计频率/%
30~40	1	1.54	1	1.54	65	100.00
40~50	2	3.08	3	4.62	64	98.46
50~60	4	6.15	7	10.77	62	95.38
60~70	14	21.54	21	32.31	58	89.23
70~80	20	30.77	41	63.08	44	67.69
80~90	15	23.08	56	86.16	24	36.92
90~100	9	13.84	65	100.00	9	13.84
合 计	65	100.00				

较小制累计频数和频率中，统计学期末考试卷面成绩在 40 分以下的有 1 人，占总数 1.54%；在 50 分以下的有 3 人，占总数 4.62%；在 60 分以下的有 7 人，占总数 10.77%；以此类推。同样，较大制累计频数和频率中，在 90 分以上的有 9 人，占总数 13.84%；在 60 分以上的有 58 人，占总数 89.23%；等等。

累计频数和累计频率具有如下两个特点：一是第一组的累计频数和累计频率等于第一组本身的频数和频率；最后一组的累计频数等于总体单位数，最后一组的累计频率等于 1。二是某组的累计频数和累计频率等于前一组的累计频数与累计频率加上本组的频数和频率。

3.4.2 统计表

1. 统计表的观念

统计表是统计用数字说话的一种最常用的形式。从广义上看，任何反映统计资料的

表格都是统计表。也就是把收集到的数字资料,经过汇总整理后,得出一些表明社会经济现象总体单位数和一系列标志总量的资料等系统化的统计资料,将其按一定顺序填列在一定的表格内,这个表格就是统计表。

统计表有以下几方面优点:能使大量的统计资料系统化、条理化,因而能更清晰地表述统计资料的内容;利用统计表易于检查数字的完整性和正确性;利用统计表便于比较各项目(指标)之间的关系,而且也便于计算;采用统计表表述统计资料显得紧凑、简明、醒目,使人一目了然,且可以节省大量文字叙述。

2. 统计表的形式和内容

统计表的形式多种多样,根据使用者的要求和统计数据本身的特点,可以绘制形式多样的统计表,常见的统计表的形式和内容如下。

1) 统计表的形式

从外表形式看,统计表是由纵横交叉的一种表格所组成的,在这种表格上填写着反映社会经济现象的数字资料。统计表一般由五个主要部分组成,即表头、横行标题、纵栏标题、纵横格线和数字资料,必要时可以在统计表的下方增加说明或注解,以说明资料的来源,某些指标数值的计算方法、填表单位和其他需要说明的问题。

(1) 表头。表头一般应包括表号、总标题和表中数据的单位等内容。总标题也称统计表名称,一般标题内容应满足 3W 要求,需要表明统计数据的时间(when)、地点(where)以及何种数据(what),即用概括性文字简单明了地说明统计资料的时间、基本内容和范围,一般写在表的上部中端。

(2) 横行标题。横行标题说明横行数字的属性,反映统计表的主要项目,常对应各组名称,位于表格的左侧,要按时间先后或数量大小和事情的重要性等顺序排列。

(3) 纵栏标题。纵栏标题就是统计指标的名称,说明纵栏各项资料的内容或每一列中数字的属性,位于表格右部上方即右部的第一横行,一般对应指标名称,有单位的要注明单位。

(4) 纵横格线。统计表中只有纵横线,无斜线。

(5) 数字资料。数字资料也称指标数值,它是统计表的具体内容,列在各横行标题和纵栏标题的交叉处,任何一个具体数值都由横行标题和纵栏标题所限定。数字资料一律用阿拉伯数字,同一列的小数位数应尽量一致,且位次对齐。

要合理安排统计表的结构,如横行标题、纵栏标题、数字资料的位置应安排合理。当然,由于强调的问题不同,横行标题和纵栏标题可以互换,但应使统计表的横竖长度比例适当,避免出现过高或过长的表格形式。

2) 统计表的内容

从内容上看,统计表包括主词和宾词两个部分。

(1) 主词。主词就是统计表中所要说明对象或统计表中各种指标所描述的总体及其各组成部分,常列在表的左方即列于横行。它可以是各个总体单位的名称、总体各个组成部分,通常形式上表现为横行标题。

(2) 宾词。宾词是用来说明总体数量特征的各项统计指标,是说明主词的各种指标,

包括指标名称、指标数值，通常列在表的右方，即纵栏标题和指标数值所在的列。

若这样排列会使统计表的表式过分狭长或过分宽短时，可将主词、宾词全并排列或变换位置排列。

3. 统计表的种类

1) 简单表、分组表和复合表

按分组程度，统计表可分为简单表、分组表和复合表。

(1) 简单表。简单表是未经任何分组的统计表。主词由研究总体单位清单组成的一览表，如地区、国家城市等目录组成的区域表，如表 3-13 所示；宾词由时间顺序组成的编年表、指标按说明问题的主次先后顺序排列等，如表 3-14 所示。

表 3-13 2017 年我国部分省市三次产业比重和人均 GDP

省份	三次产业比重/%			地区生产 总值/亿元	人均地区生产 总值/元
	第一产业	第二产业	第三产业		
甘肃	13.9	33.4	52.7	7 677	29 238
广西	14.2	45.6	40.2	20 396	41 752
河南	9.6	47.7	42.7	44 988	47 130
湖南	10.7	40.9	48.4	34 591	50 563
吉林	9.3	45.9	44.8	15 289	56 102
广东	4.2	43.0	52.8	89 879	81 089
浙江	3.9	43.4	52.7	51 768	92 057
上海	0.3	30.7	69.0	30 134	124 606
北京	0.4	19.0	80.6	28 000	128 992
全国	7.9	40.5	51.6	827 122	59 660

表 3-14 2017 年某集团公司下属企业基本情况

下属企业	营业收入/亿元	利润总额/万元	产品销售率/%
A	7.91	3 130	95.37
B	9.74	8 580	98.98
C	6.88	4 040	97.19
合计	24.53	15 750	—

(2) 分组表。分组表是按某一标志进行分组的统计表。分组表可以从主词的角度选择某一标志分组，以揭示总体不同类型的特征、研究总体的内部构成、分析总体内事物和现象之间的依存关系等，如表 3-15 所示；分组表也可以从宾词的角度将统计指标从不同角度按某一标志分组或各种分组平行排列，如表 3-16 所示。

(3) 复合表。复合表是按两个或两个以上标志进行复合分组的统计表。在一定分析任务要求下，复合表可以把总体的标志结合起来对主词进行复合分组，更深入地分析社会经济现象的特征和规律性，如表 3-17 所示；复合表也可以是宾词的统计指标，同时有层次地按两个或两个以上标志分组，且分组重叠在一起，如表 3-18 所示。

表 3-15 某班级统计学成绩分布

考试分数	人数/人	频率/%
50~60	2	5.0
60~70	7	17.5
70~80	12	30.0
80~90	11	27.5
90~100	8	20.0
合计	40	100.0

表 3-16 2017 年某企业职工性别及文化程度情况

企业名称	职工人数/人	性别		文化程度		
		男	女	小学	中学	大学
A 分公司	660	326	334	22	333	305
B 分公司	368	156	212	14	156	198
合 计	1 028	482	546	36	489	503

表 3-17 2017 年某企业职工性别、职称情况

组别	人数/人	比重/%
男性	892	67.7
正高	294	22.3
高级	368	27.9
中级	178	13.5
助理级	52	3.9
女性	426	32.3
正高	103	7.8
高级	202	15.3
中级	76	5.8
助理级	45	3.4
合计	1 318	100.0

表 3-18 2017 年某企业职工性别及文化程度情况

企业名称	职工人数		小学			中学			大学		
	男	女	男	女	小计	男	女	小计	男	女	小计
A 分公司	326	334	12	10	22	152	181	333	162	143	305
B 分公司	156	212	8	6	14	58	98	156	90	108	198
合计	482	546	20	16	36	210	279	489	252	251	503

2) 调查表、整理表和分析表

按用途,统计表可分为调查表、整理表和分析表,分别应用于统计工作过程的各个环节。

(1) 调查表。调查表是把调查项目以表格的形式表示出来所形成的统计表。调查表

用于登记、收集原始资料,表中的数字可以说是未经综合的各个单位的标志值。

(2) 整理表。整理表是用于登记统计资料整理结果的统计表。表中数字是经过汇总后的总量指标,也称汇总表。

(3) 分析表。分析表是表述统计分析资料的统计表。它是在在统计分析中用于对整理所得统计资料进行统计定量分析的表格,往往与整理表结合在一起,成为整理表的延续,表中的数字是在总量指标基础上计算的多种相对指标和平均指标。

4. 统计表的编制规则

统计表的编制,无论主词的内容或宾词指标的配置都要目的明确、内容鲜明,能从表中获得事物和现象的具体内容。在制表时,首先要根据目的要求,做到简明、紧凑,避免过分烦琐;要重点突出,使人看过后能明白表格所要表达的主要内容,不要包罗万象;要层次分明,避免层次过多或结构混乱。

必须注意的规则有以下几个方面。

(1) 统计表的各种标题,特别是总标题的表达应该十分简明确切,概括地反映表的基本内容。总标题应标明资料所属时间、空间和表的基本内容。

(2) 表中主词各行和宾词各栏,一般应按先局部后整体的原则排列,即先列各个项目,后列总计。当没有必要列出所有项目时,可先列总计,后列其中一部分重要项目。主词与宾词位置可互换,各栏排列次序应以时间先后、数量大小、空间位置等自然顺序编排。

(3) 若统计表栏数多,常加以编号。在主词和计量单位等栏用(甲)、(乙)、(丙)等文字标明,宾词指标各栏用(1)、(2)、(3)等数字编号。

(4) 表中数字应填写整齐,对准位数。统计表中不应有空格,当数字太小可略而不计或数字不详时,用符号“……”表示;当表中不应有内容或可免填或无意义时用符号“—”表示。

(5) 统计表中必须注明数字资料的计量单位。当全表只有一种计量单位时,可把计量单位写在表头右上方;当表中需分别注明不同单位时,纵栏计量单位可专设一栏;纵栏的计量单位,要与纵标写在一起,用小字标写。

(6) 统计表一般是“开口式”的,即统计表一般为横长方形,表的左右两端不画纵线,上、下基线要画粗线。

(7) 必要时,统计表应加说明或注解。个别需要说明的指标或数据。例如,某些指标有特殊的计算口径,某些资料只包括一部分地区,某些数字是由估算来插补的,都要加以说明。此外还要注明统计资料的来源,以便查考。说明或注解不应写在表中,在数字上角用“*”号标出,文字叙述写在表的下方。

说明或注解一般写在表的下端,制表人、制表及发出日期和主管部门负责人盖章生效,以示对提交的数字负责,便于收表单位查询联系。

3.4.3 统计图

为了使统计资料的表达直观生动、通俗易懂、便于分析比较等,可以利用统计图进

行研究,分析事物和现象之间的数量关系及变化发展情况。

1. 统计图的概念

统计图是利用点、线、面、体等绘制成几何图形或具体事物的形象和地图等形式,用以表现事物和现象的数量特征、数量关系和数量变动的图形。

统计图表示统计资料目的明确,准确反映统计资料,鲜明、形象、生动,便于比较,信息量大,较统计表能更集中反映事物运动趋势、美学效果好、容易吸引人、形象性佳。具有形象具体、简明生动、通俗易懂、一目了然的优点。统计图是表现统计数字大小和变动的各种图形的总称,主要用于表示现象间的对比关系;揭露总体结构;检查计划的执行情况;揭示现象间的依存关系,反映总体单位的分配情况;说明现象在空间上的分布情况。

统计图一般采用直角坐标系,横坐标常用来表示事物的组别或自变量 x ,纵坐标常用来表示事物出现的次数或因变量 y ;或采用角度坐标(如圆形图)、地理坐标(如地形图)等。

统计图一般由图题、图号、图目、图尺、图线、图形和图注等几个部分组成。

1) 图题、图号

图题也称图名,是指统计图的标题或名称,它反映和标明统计图的内容,通常包括内容、时间和地点;图号是统计图的编号。

2) 图目

图目是图中的标题,也称标目,是指在横轴的下面和纵轴的侧面所标注的表明事物的类型、地点、时间、单位、指标等的文字或数字,说明横、纵轴所代表的事项及其单位;也就是说,纵横两轴应有纵图目和横图目,并注明度量衡单位。

3) 图尺

图尺是统计图中的坐标单位,也称尺度。就是测定指标数值大小的标尺,包括尺度线、尺度点、尺度数和尺度单位。纵横两轴都应有尺度。

4) 图线

图线是构成统计图的各种线,一般有基线(基准线或零点线)、图示线(表现各种几何图形的线)、指导线(网格线,有纵、横之分)、轮廓线(边框线)、断裂线(折叠线或省略线)等。

5) 图形

图形即图式,是根据统计资料用较粗的图示线绘成的图形,它是统计图的主体部分,可以绘成曲线、条形、平面、立体图形等。主要通过它来表明社会经济现象的数字资料。

6) 图注

图注即统计图的注释和说明部分,包括图例、说明、资料来源等。图例是截取图形的一部分用以说明图形内容的样本。

2. 绘制统计图的一般要求

在绘制统计图时要遵循以下基本要求。

(1) 根据统计资料的性质和分析研究的目的正确选择统计图的类型。

(2) 统计图的内容应具有鲜明性。既要突出统计的要求,又要强调客观的真实性。

(3) 统计图的设计要符合科学性原则。统计图的形式和排列要有艺术性,为使图形美观并便于比较,统计图的长宽比例一般为7:5,有时为了说明问题也可加以变动。

(4) 每个图都应有标题,标题要简明扼要、切合统计图的内容;标题位置在图域之外,一般放在图域的下面。

(5) 纵、横轴都应有图目,并注明统计资料的计量单位,计量单位应放在尺度线的顶端或外侧。尺度数的位数不宜过多,如果过多,应扩大其计量单位,以减少位数。

(6) 在同一统计图内比较几种不同的事物时,须用不同的线条或颜色表示,并附图例说明。图例的形状、颜色、线纹图案等都应与图形本身相一致。

(7) 统计图横轴尺度自左至右,纵轴尺度自下而上,数值一律由小而大;尺度间隔要宽松;用算术尺度时,等长的距离应代表相等的数量,图中如果省略图尺,在图形上应标注指标数值。

3. 统计图的绘制

统计图的类型丰富多样,下面介绍几种常见的。

1) 直方图

直方图是用直方形的宽度和高度来表示频数分布的图形(图3-3)。横轴表示各组名称(或各组组限),纵轴表示频数(一般标在左方)或频率(一般标在右方)。如没有频率只有频数,则只保留左侧频数;如没有频数只有频率,则只保留左侧频率。直方图能够显示各组频数分布的情况,易于比较各组之间频数的差别。

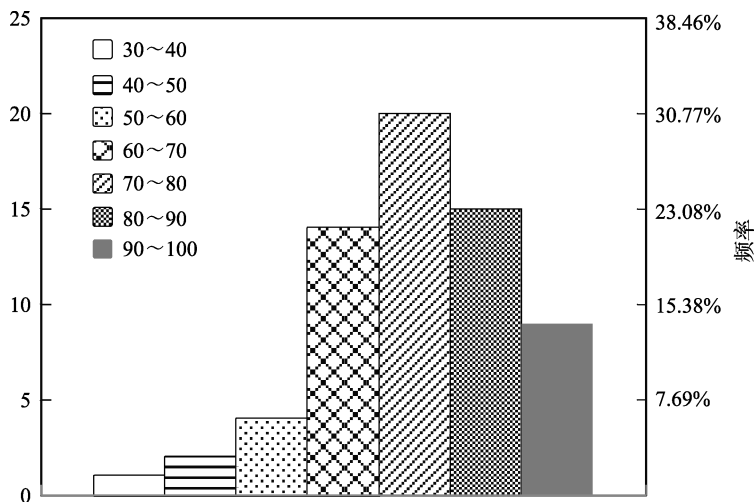


图3-3 某专业16级65名同学统计学卷面成绩分布直方图

直方图绘制先根据各组的频数及频率确定其在纵轴上的坐标,并依据各组组距的宽度与频数的高度绘成直方形。绘制时注意以下几个方面。

(1) 坐标轴的横轴代表变量值,要用相等的距离表示相等的数量,而纵轴坐标要从

0 开始。

(2) 在分组时, 各组之间范围的数值是连续的, 各矩形间不留空隙。

(3) 对于组距相等的资料可以直接作图; 组距不等的资料先进行换算, 全部转化为组距相等的频数, 用转化后的频数作图, 或按实际组距为宽、以频数密度为高绘制直方图, 其中频数密度等于频数除以组距, 表示单位组距的个体频数或次数, 也称分配密度。

2) 折线图

折线图是用折线表示数量变化特征和规律的统计图, 分为一般折线图和累计折线图。常用折线图来描绘统计事物和现象总体指标的动态、研究对象间的依存关系、总体中各部分的分配情况和事物与现象发展变化的规律和趋势等。折线图与直方图都用高度表示频数或频率, 宽度表示组距的长度, 均是较好地显示频率分布图示方法; 不同点在于直方图面积的变化能同时反映组距和频数(或频率)两个因素的变化, 而折线图较简单, 且对资料所描绘的轮廓更加清晰, 易于显示数据的变化趋势。

绘制折线图时注意以下几个方面。

(1) 一般情况下, 横轴表示各组名称或各组组限, 纵轴表示各组频数或频率。

(2) 折线图的长宽比例要适当, 一般应绘制成横轴略大于纵轴的长方形, 其长宽比例大致为 10:7。

(3) 一般情况下, 纵轴数据下端应从 0 开始, 以便于比较。如果数据与 0 间距过大, 可以在纵轴 0 值附近采用断裂线或折叠线表示省略或折叠了部分数据, 对于横轴可作类似的处理。

(4) 若实际需要, 可以在一个坐标系中画两条或两条以上的折线, 来表示不同组的数据变化趋势, 但也应注明图例说明。

(5) 保证折线图所围的面积与直方图的面积相等, 从而使两种图形表示的频数分布一致。折线图的起点或终点要与横轴相交, 将第一个矩形的顶部中点通过竖边中点(该组频数一半的位置)连接到横轴, 或将最后一个矩形顶部中点与其竖边中点连接到横轴。

① 一般折线图。一般折线图可以在直方图的基础上将直方图每个直方形顶端中点用折线连点而成; 若不绘直方图, 可用各组组中值与对应的频数(或频率)确定的坐标点连接而成, 如图 3-4 所示。起点是在距左边最低组半个组距处的横轴上, 终点是在距右边最高组半个组距处的横轴上, 它表示频数分布的图形, 可根据图表资料绘制折线图。

② 累计折线图。累计折线图分为较小制累计折线图与较大制累计折线图, 较小制累计折线图或较大制累计折线图均以各组名称为横轴, 以累计频数(或频率)为纵轴。但两者的绘制方法有所不同。绘制较小制累计折线图时, 是从最小组的下限开始, 依次用折线连接各组上限与该组累计频数确定的坐标点而成(图 3-5); 绘制较大制累计折线图时, 是从最大组的上限开始, 依次用折线连接各组下限与该组累计频数确定的坐标点而成(图 3-6)。

3) 曲线图

曲线图是以曲线的升降来表示事物和现象动态的图形。曲线图与折线图的绘制方法基本相同, 当对数据所分的组数很多时, 组距会越来越小, 这时所绘制的折线图就会越来越光滑, 逐渐近似地表现为一条平滑的曲线, 这种图形称为曲线图。运用曲线图可以

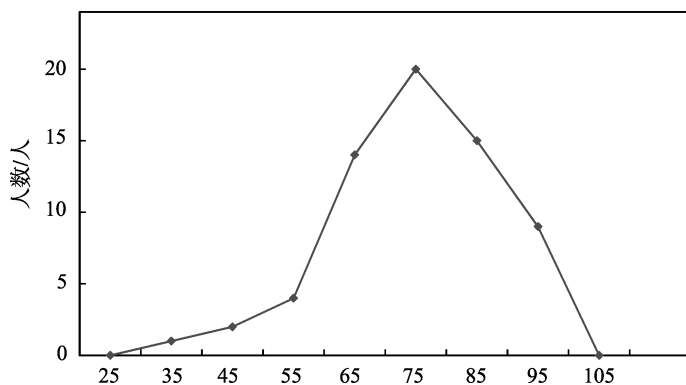


图 3-4 某专业 16 级 65 名同学统计学卷面成绩分布折线图

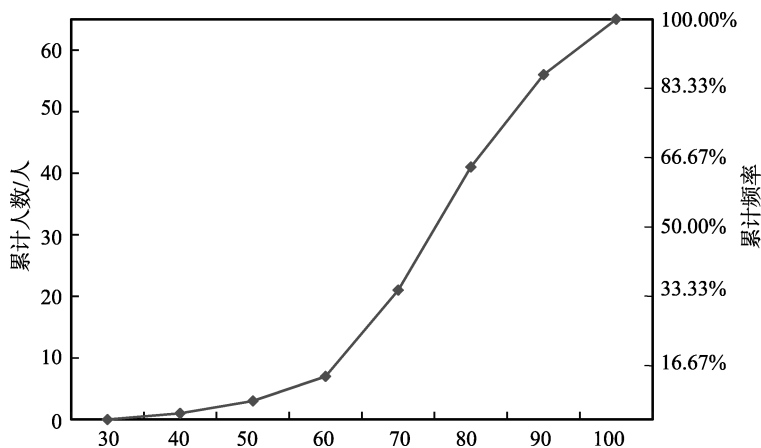


图 3-5 某专业 16 级 65 名同学统计学卷面成绩分布较小制累计折线图

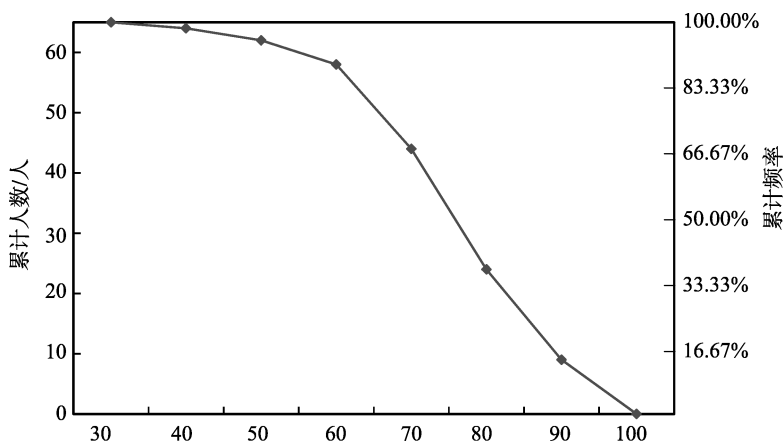


图 3-6 某专业 16 级 65 名同学统计学卷面成绩分布较大制累计折线图

表明事物的发展规律、总体单位的分配情况、揭示事物间的依存关系或表明计划执行的

进度。适用于变量较多、组数趋向于无限时的情形，其实质是一种关于连续变量的次数或比率分布的函数关系图。常见的曲线图有次数分布曲线图和动态曲线图等，次数分布曲线图有前面介绍的钟形分布图、J 形分布图、U 形分布图等；动态曲线图有简单动态曲线图、带纹曲线图和依存关系曲线图等。

4) 条形图

条形图是以条形的长短、高低来比较相关统计指标数量大小、多少的图形。条形图的各组条形之间有间隔，其可以是不同指标之间的比较，如总量指标、相对指标或平均指标；可以是不同时间、不同区域、不同单位之间的比较；也可是实际与计划的比较。适用于相互独立的资料（资料有明确分组，不连续）的比较。条形图的特点在于能够显示每组中的具体数据，易于比较数据之间的差别；但不能明确显示部分与整体的对比。条形图与直方图的区别：一是条形图是用条形的高度表示频数的大小，而直方图实际上是用长方形的面积表示频数，当长方形的宽相等的时候可以用矩形的高表示频数；二是条形图中，横轴上的数据是孤立的，是一个具体的数据，而直方图中，横轴上的数据是连续的，是一个范围；三是条形图中，各长方形之间有空隙，而直方图中，各长方形是连接在一起的。

条形图的形式可以纵排或横排。纵排的条形图称为柱形图，用横轴做基线表示统计指标，用纵轴做尺度线表示指标数值；横排的条形图称为带形图，用纵轴做基线表示统计指标，用横轴做尺度线表示指标数值。

条形图绘制的几个要点：一是坐标轴的横轴为观察项目，纵轴为数值，纵轴坐标一定要从 0 开始，取一个单位长度表示数量的多少要根据具体情况而定。二是各直条的宽度应相等，间距也相等，间距宽度和直条相等或为其一半。复式直条图在同一观察项目的各组之间无间距。三是排列顺序可以根据数值从大到小，从小到大，或按时间顺序排列。四是复合条形图有几种不同的形式，图中表示不同项目的条形，要用不同的线纹或颜色区别开，并注明图例说明。

(1) 单条形图。单条形图用来比较同类指标的条形图（图 3-7）。纵排形式条形图：先绘制直角坐标，横轴表示基线，纵轴表示水平，基线上每隔相同距离截取适当线段排列各条形，纵轴以零为起点，标出适当尺度。

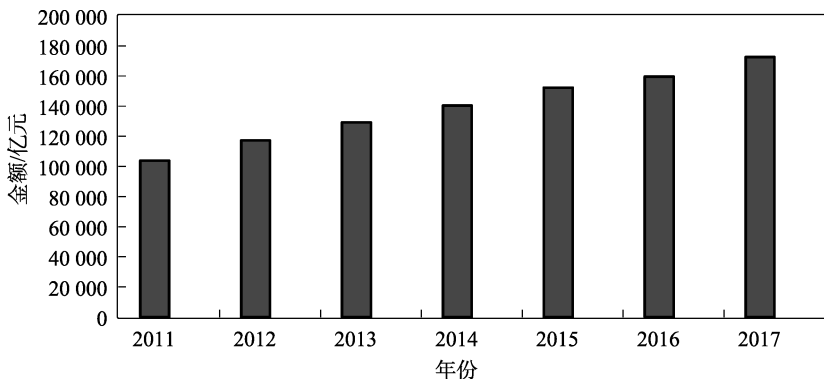


图 3-7 我国 2011—2017 年财政收入单条形图

(2) 复式条形图。复式条形图是用来比较两种以上统计指标的条形图，它常常用来表现分组资料（图 3-8）。复式条形图不是一个一个地排列，而是以两个以上的条形为一组，一组一组地排列来进行比较的一种图形，它既可以进行组与组之间的比较，又可以进行组内的比较。

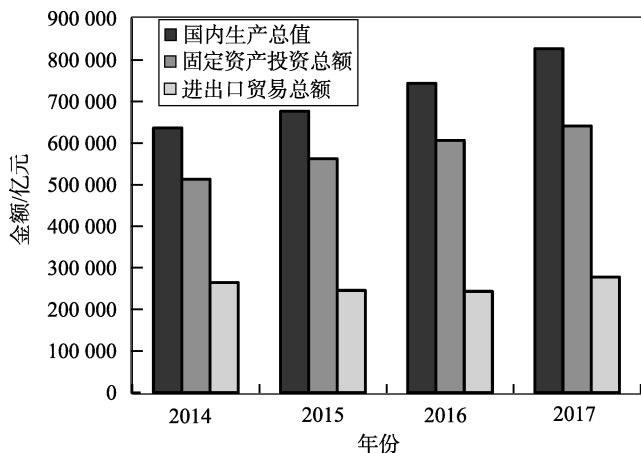


图 3-8 我国 2014—2017 年国内生产总值、固定资产投资和进出口贸易总额复式条形图

(3) 分段条形图。分段条形图也称结构条形图，用来表明事物总体的结构及其比较。可以采用纯百分数（图 3-9）和绝对数与百分数相结合（图 3-10）两种形式。以一个独立的条形或几个条形的全部长度代表被说明现象的总体，并把条形分割为几个小段，用来表示构成这一总体的各个组成部分。它既可以比较现象的各部分在总体中所占比重的大小，又可以说明现象在不同时期的构成资料。这些资料可以是绝对数，也可以是百分数。

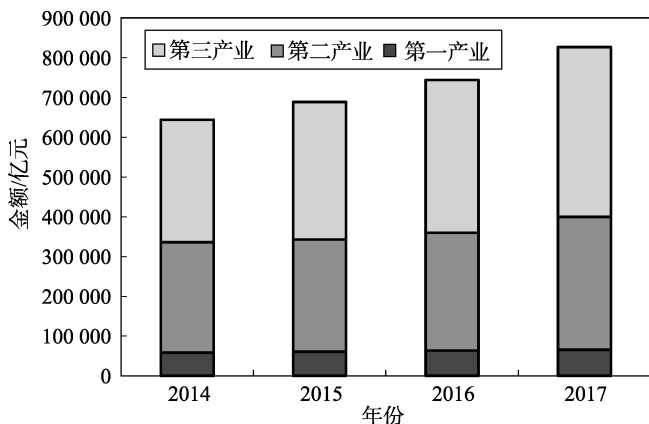


图 3-9 我国 2014—2017 年国内生产总值分段条形图

分段条形图绘制时要注意：一是标尺，一定要有标尺，位于图的上方或下方，起始的位置、总长度和百分条图一致，并和百分条图平行。全长为 100%，分成 10 格，每格

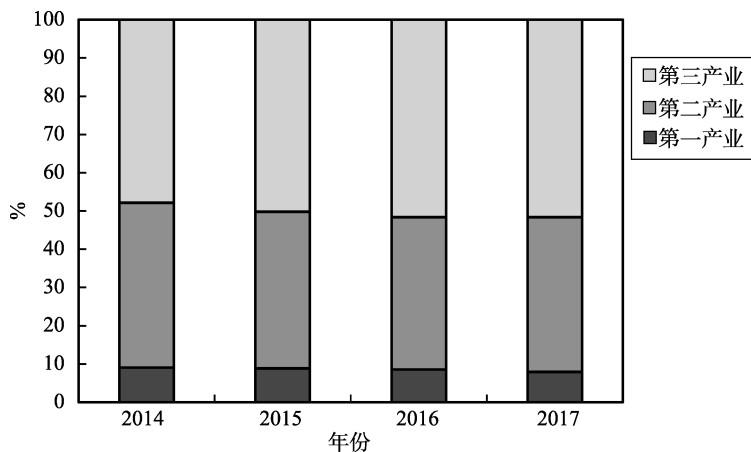


图 3-10 我国 2014—2017 年国内生产总值百分数分段条形图

10%。二是分段，按各部分所占百分比的大小排列，在图上标出百分比。三是多组比较。若要比的事物不止一个，可以画几个平行的百分条图，以示比较。四是在图外要附图例说明。各条图的排列顺序相同，图例相同。

5) 平面图

平面图就是以平面面积来表明统计指标数量的图形。平面图主要有以下几种。

(1) 正方形图。正方形图是以正方形的面积代表指标数值大小的一种图形。正方形的面积由其边长决定，常用比例方式来显示和比较事物或现象。在绘图时，只需把指标数值开方求得边长，再按放大或缩小比例绘出相应的若干个正方形图进行比较。

(2) 长方形图。长方形图是以长方形的面积代表指标数值大小的一种图形。长方形的面积由其长和宽决定，因此可以用来比较统计指标，说明内部结构及变化，显示和比较两个因素构成的综合现象或事物。

(3) 圆形图。圆形图也称饼图，它是用圆形及圆内扇形的面积来表示数值大小的图形。主要用于表示总体中各组成部分所占的比例。圆形图包括圆形面积图、圆形结构图和圆环图。

圆形面积图的面积由其半径决定，常用比例方式来显示和比较事物与现象指标数值的大小、多少。圆形结构图就是将一个圆按比例分为若干部分，每个部分代表一种事物和现象指标数值占总体数量的百分比。圆形结构图以一个圆面积为 100%，用圆内各扇形面积所占的百分比来表示各部分所占的构成比例，比较适用于定类和定序数据的显示，如图 3-11 所示。绘制要点：一是绘制圆形，按照各组所占的比重乘以 360° 所得到的角度数将圆形划分为若干不同的部分；二是从相当于时钟 12 点或 9 点的位置开始顺时针方向绘图；三是每部分用不同线条或颜色表示，并在图上标出百分比，下附图例说明；四是当比较不同资料的构成时，可以画两个相等大小的圆，在每个圆的下面写明标题，并用相同的图例表示同一个构成部分。

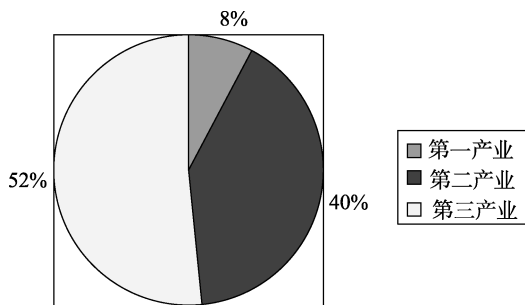


图 3-11 2017 年我国国内生产总值圆形结构图

圆环图是用圆环来代表统计总体，使用环中的每一小段来显示统计总体特征的图形。在圆环中显示数据，表示其占整体的百分比，其中每个圆环代表一个数据系列（图 3-12）。像圆形结构图一样，圆环图显示各个部分与整体之间的关系，但是它可以包含多个数据系列，而数据系列源自统计表的行或列；图表中的每个数据系列具有唯一的颜色或图案，并且在图表的图例中表示；可以在图表中绘制一个或多个数据系列，而圆形结构图只有一个数据系列。圆环图的绘制步骤如下：首先绘制圆环，有几个总体就绘制几个圆环；然后分别计算每个总体中各组所占的比重；最后按照各组所占的比重乘以 360° 所得到的角度数将圆环划分为若干不同的小段。

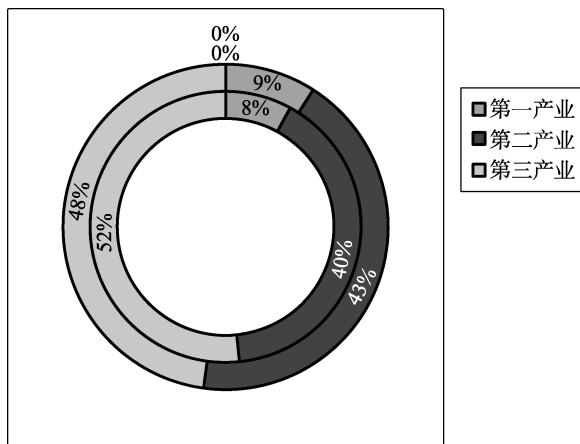


图 3-12 2017 年我国国内生产总值圆环图

圆环图有两种类型：一般圆环图和分离型圆环图。分离型圆环图除了将扇区从圆环图中心移开一段距离外，均与一般圆环图相同。这导致在圆环图扇区之间出现空隙。

6) 象形图

象形图是运用事物和现象本身的形象的长度、大小、多少来显示其统计数据的图形。用象形图表示数据，每一象形图均代表一个固定的数量，如果少于一个图形所代表的数量，按比例绘画部分图形；要注明每个图形所代表的数量与计量单位。采用这种表示方式具有直观、生动、形象的优点，但是不甚准确。

象形图主要有单位象形图、长度象形图和面积象形图三种。

(1) 单位象形图。单位象形图就是利用一个简化的事物或现象图形代表一定数量某种或某类事物或现象。如用一个人图像代表 100 名职工，用一棵小树代表 100 万亩绿化面积，用一穗稻子代表 100 万吨粮食产量，等等。

(2) 长度象形图。长度象形图采用单位长度的象形图代表一定距离或空间长度的某种或某类事物和现象。如用 1 厘米长的铁路图案表示 100 千米的铁路、用 1 厘米长的河流图形表示 50 千米江河里程等。

(3) 面积象形图。用事物和现象单位面积的图案代表一定面积的某种或某类事物或现象。如用一平方厘米面积的房屋或建筑物图案表示一万平方米的房屋或建筑物面积等。

7) 散点图

散点图是以点的密集程度、分布形态与趋势来反映事物和现象之间的相关关系的图形，又称散点分布图（图 3-13）。散点图能直观表现出影响因素和预测对象之间的总体关系趋势，或因变量随自变量而变化的大致趋势；能通过直观醒目的图形方式反映变量间关系的变化形态，以便决定用何种数学表达方式模拟变量之间的关系。散点图不仅可传递变量间关系类型的信息，也能反映变量间关系的明确程度；通过对数据相关性的直观观察，不但可以得到定性的结论，而且可以通过观察剔除异常数据，提高估算相关程度的准确性。

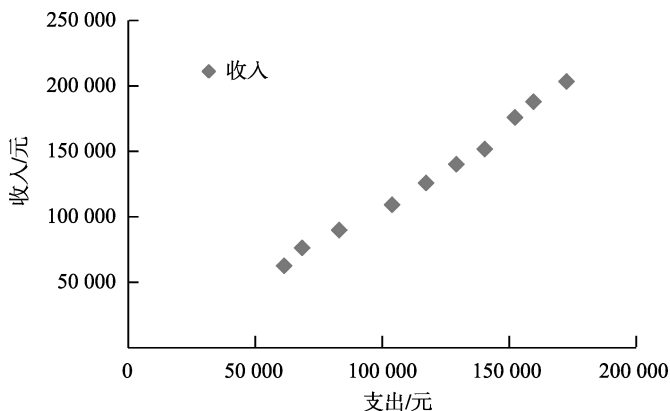


图 3-13 2008—2017 年我国财政收支关系散点图

散点图通常用于显示和比较数值，特别是比较跨类别的聚合数据，类别由图表中的不同标记表示，当要在不考虑时间的情况下比较大量数据点时，最好使用散点图；散点图尤其适用于双变量资料，反映两个事物、现象之间的相关关系和进行相关回归分析。散点图中包含的数据越多，比较的效果就越好；对于处理值的分布和数据点的分簇，散点图就很理想。绘制要点有以下几个。

(1) 横轴代表一个变量，纵轴代表另一个变量。

(2) 纵横轴的坐标起点不一定要从 0 开始，根据资料的情况而定。

(3) 默认情况下，散点图以圆圈显示数据点。如果在散点图中有多个序列，就需要

考虑将每个点的标记形状更改为方形、三角形、菱形或其他形状。

8) 茎叶图

茎叶图是将一组数据按位数进行比较,分为“茎”和“叶”以统计数据的图形,又称“枝叶图”。茎叶图图形是由数字组成的,用于显示未分组的原始数据的分布,包括“茎”和“叶”两部分(图 3-14)。它的思路是将一组数据按位数进行比较,将数的大小基本不变或变化不大的位作为一个茎(主干),将变化大的位的数作为叶(分枝),叶上只保留一位数字,列在茎的后面;“茎”和“叶”之间用竖线隔开,竖线左边是“茎”、右边是“叶”,这样就可以清楚地看到每个茎后面的几个数,每个数具体是多少。对于 $n(20 \leq n \leq 300)$ 个数据,茎叶图最大行数不超过 $L = [10 \times \lg n]$ 。

数据个数	树茎	树叶
1	3	2
2	4	5 8
4	5	0 3 4 7
14	6	0 0 1 1 2 3 4 4 5 6 7 7 8 8
20	7	0 0 1 1 2 3 3 4 4 5 5 6 6 7 7 7 8 8 9
15	8	0 0 1 2 2 3 4 5 5 6 6 7 8 8 9
9	9	0 0 1 1 3 4 6 8 9

图 3-14 某专业 16 级 65 名同学统计学卷面成绩分布茎叶图

茎叶图有三列数:左边的一列为统计数,它是上(或下)向中心累计值,中心数表示最多数组的个数;中间的一列表示“茎”,也就是变化不大的位数;右边的是数组中的变化位,即“叶”,是按照一定间隔将数组中的每个变化的数一一列出来,像一条枝上抽出的叶子一样,人们形象地称为茎叶图。茎叶图作图过程是先作“茎”后填“叶”,将分组标志的组距视为茎,按数的大小从上到下(或从下到上)排列。将每一个观察值视为一个树叶,每一个树叶按照树茎的要求长在应长的树茎上。若数值较多,树叶较长,可将高位数重复两次,个位数分为长度相等的两枝。为了便于分析,可将 1/4、3/4 分位数及中位数用符号标出。

茎叶图是一个与直方图相类似的特殊工具,茎叶图类似于横置的直方图,将茎叶图的“茎”和“叶”逆时针方向旋转 90° ,实际上就是一个直方图,可以从中统计出次数,计算出各数据段的频率或百分比,从而可以看出分布是否与正态分布或偏态分布逼近;茎叶图与直方图不同在于:茎叶图既能给出数据的分布状况,又能给出每一个原始数值,保留了原始数据的信息;直方图只可观察一组数据的分布状况,但没有给出具体的数值,失去原始资料的信息。茎叶图在质量管理上运用较广,特别是作为更细致分析阶段使用。

用茎叶图表示数据有两个优点:一是从统计图上没有原始数据信息的损失,所有数据信息都可以从茎叶图中得到;二是茎叶图中的数据可以随时记录,随时添加,方便记录与表示。茎叶图只便于表示两位有效数字的数据,而且茎叶图只方便记录两组数据,

两组以上的数据虽然能够记录,但是没有表示两组记录那么直观、清晰。

9) 箱线图

箱线图是一种运用五个特征值来显示一组数据分散情况、形状如箱子的统计图形,又称箱形图、箱须图或盒状图。箱线图用于显示未分组的原始数据的分布,由一组数据的最小值、上四分位数、中位数、下四分位数与最大值五个统计量绘制而成,由一个箱子和两条线段组成(图 3-15)。其运用常用的统计量,能粗略地判断数据是否对称、提供有关数据位置和分散程度的关键信息、直观明了地识别一组数据中的异常值,尤其在比较不同的总体数据或几个样本时更能体现其差异。箱线图不能提供关于数据分布偏态和尾重程度的精确度量;对于批量较大的一组数据,箱线图反映的形状信息更加模糊;用中位数代表总体平均水平有一定的局限性;等等。箱线图最好结合其他描述统计工具如均值、标准差、偏度、分布函数等来描述数据批的分布形状。

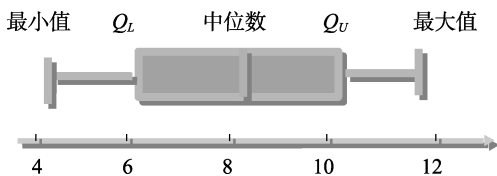


图 3-15 简单箱线图

箱线图的绘制方法是:首先找出一组数据的五个特征值,即最大值、最小值、中位数 Me 和两个四分位数(下四分位数 Q_L 和上四分位数 Q_U),连接两个四分(位)数画箱子,再将两个极值点与箱子相连接。

箱线图的绘制步骤如下。

(1) 画数轴,度量单位大小和该组数据的单位一致,起点比最小值稍小,长度比该组数据的全距稍长。

(2) 画一个矩形箱,两端边的位置分别对应该组数据的上下四分位数(Q_U 和 Q_L)。在矩形箱内部中位数位置画一条线段为中位线。

(3) 在 $Q_U + 1.5IQ_R$ (四分位差) 和 $Q_L - 1.5IQ_R$ 处画两条与中位线一样的线段,这两条线段为异常值截断点,称其为内限;在 $F + 3IQ_R$ 和 $F - 3IQ_R$ 处画两条线段,称其为外限。处于内限以外位置的点表示的数据都是异常值,其中在内限与外限之间的异常值为温和的异常值,在外限以外的为极端的异常值。

(4) 从矩形箱两端边向外各画一条线段,直到不是异常值的最远点,表示该组数据正常值的分布区间。

(5) 用“○”标出温和的异常值,用“*”标出极端的异常值。相同值的数据并列标在同一数据线上,不同值的数据标在不同数据线上。至此一组数据的箱线图便绘出了。

经过排序以后的某专业 16 级 65 名同学统计学卷面成绩分布如下: 32 45 48 50 53 54 57 60 60 61 61 62 63 64 64 65 66 67 67 68 68 70 70 71 71 72 73 73 74 74 75 75 76 76 77 77 77 78 78 79 79 80

80 81 82 82 83 84 85 85 86 86 87 88 88 89 90 90 91 91 93
94 96 98 99

求得 5 个统计量分别为 32, 66, 75, 85, 99, 如图 3-16 所示。

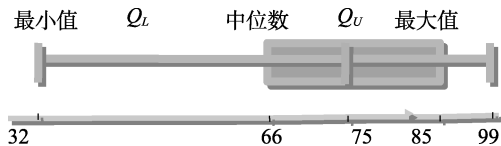


图 3-16 某专业 16 级 65 名同学统计学卷面成绩分布箱线图

说明：这 65 名学生成绩存在异常值，说明他们的生产状况处于正常范围内。

10) 线图

线图是以线段的上升或下降来表示一种事物或现象随另一种事物或现象变化或事物和现象在时间上的发展变化情况的图形(图 3-17)。可以用来显示某一总体的多个指标或者多个总体的某一指标的发展变化的特征与趋势。通常用于时间序列的连续性资料，反映事物和现象连续的动态变化规律。

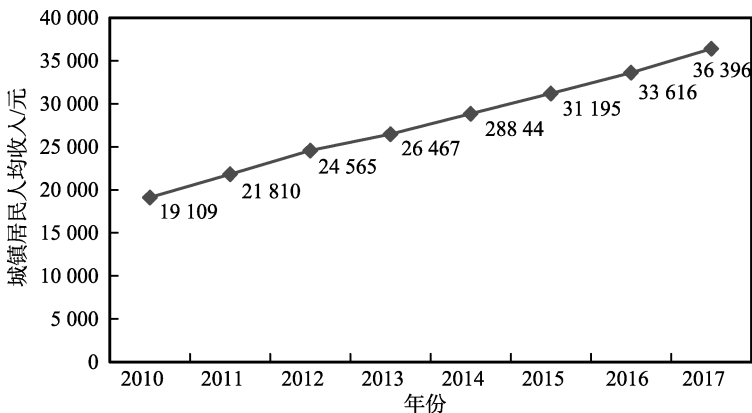


图 3-17 我国 2010—2017 年城镇居民家庭人均收入线图

线图绘制要点有以下几个。

- (1) 横轴表示时间或组段，纵轴表示频数或频率。
- (2) 数据点画在组段中间位置，相邻的点用直线连接。无数据的组段用虚线连接。直线不能任意外延。
- (3) 同一张线图上不要画太多条曲线。当有两条或两条以上曲线在同一张线图上时，须用不同颜色或不同的图形形式加以区分，并附图例加以说明。
- (4) 图形的长宽比例要适当，大致为 10 : 7。
- (5) 一般情况下，纵轴数据下端应从“0”开始，以便于比较。数据与“0”之间的间距过大时，可以采取折断的符号将纵轴折断。

11) 雷达图

雷达图是一种用由若干条从圆心出发的夹角相等的半径组成的坐标轴来显示多个变量的图示方法,其反映的指标的分布组合在一起形状像雷达而得名,又可称为戴布拉图、蜘蛛网图。用于研究多个样本之间的相似程度,通常为若干个总体的某一指标或某一总体的若干个指标的分布特征,特别在显示或对比各变量的数值总和时十分有用。

假定各变量的取值具有相同的正负号,总的绝对值与图形所围成的区域成正比,设有 n 组样本 $S_1, S_2, \dots, S_n$, 每个样本测得 P 个变量 $X_1, X_2, \dots, X_p$, 要绘制这 P 个变量的雷达图,其具体做法是:先作一个圆,然后将圆 P 等分,得到 P 个点,令这 P 个点分别对应 P 个变量,再将这 P 个点与圆心连线,得到 P 个辐射状的半径,这 P 个半径分别作为 P 个变量的坐标轴,每个变量值的大小由半径上的点到圆心的距离表示;再将所有标志值描绘在相应的坐标轴上,将同一样本的值在 P 个坐标上的点连线。这样, n 个样本形成的 n 个多边形就是一个雷达图。

2017 年某企业收益性、生产性、流动性、安全性、成长性财务“五性”指标分值如表 3-19 所示,编制指标雷达图如图 3-18 所示。

表 3-19 2017 年某企业财务“五性”指标分值表

分值项目	收益性	生产性	流动性	安全性	成长性
行业最佳分值	100	100	100	100	100
本企业分值	70	78	65	98	80
行业最差分值	10	40	18	20	12

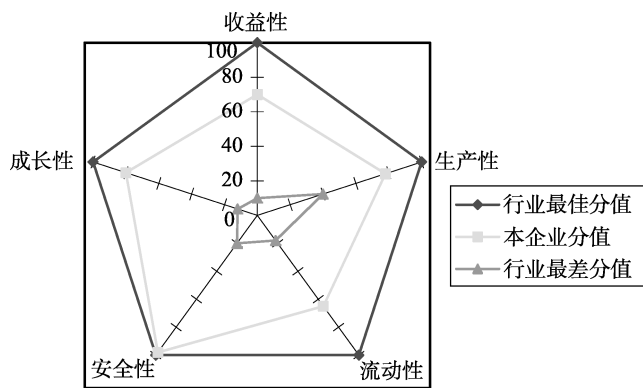


图 3-18 2017 年某企业财务“五性”指标雷达图

3.5 案例：变量数列的编制

3.5.1 案例背景

某企业有 135 名生产工人,上月生产 A 产品产量为

114	120	144	132	124	100	147	134	128	108	138	126	146	104	154
158	145	94	138	114	140	109	124	130	124	96	142	130	116	128
107	148	114	124	130	110	97	120	146	132	126	116	124	142	136
158	118	104	132	116	96	100	128	118	104	154	116	148	94	108
126	134	101	124	152	120	132	140	150	124	114	106	126	136	146
106	91	130	149	146	118	138	104	158	128	148	106	110	126	122
134	114	155	96	128	114	134	120	158	126	126	136	141	116	148
124	134	114	108	128	104	150	128	116	122	99	104	134	108	114
126	118	146	126	138	106	142	136	157	96	126	128	146	103	152

对其进行变量数列编制和绘制统计图。

3.5.2 案例分析

变量数列的编制按如下步骤进行。

1. 原始资料的预处理

先对资料进行审核,若发现资料中的错误,就进行订正;然后筛选,找出符合条件的资料或剔除不符合特定条件的资料;再排序,按标志值从小到大进行排序,确定其最大值、最小值,并计算全距。

该企业 135 名生产工人上月生产 A 产品产量,按照从小到大顺序排列为

91	94	96	96	97	99	100	101	103	104	104	104	106	106	107
107	108	108	108	108	109	110	110	112	112	112	113	113	114	114
114	114	114	114	115	116	116	116	117	117	117	117	118	118	118
118	118	119	119	120	120	120	120	121	122	122	122	122	122	123
123	123	123	123	124	124	124	124	124	124	124	124	125	125	126
126	126	126	126	126	126	127	127	127	128	128	128	128	128	128
129	129	130	130	130	131	132	132	133	133	133	134	134	134	134
134	135	136	136	136	137	138	138	139	139	140	141	142	142	144
145	146	146	146	147	148	148	149	150	152	154	155	157	158	158

从上可得,生产 A 产品产量最大值为 158,最小值为 91,全距=最大值-最小值=158-91=67。

2. 确定变量数列的形式

由于变量数列有单项式数列和组距式数列两种。组距式数列适用于按连续变量分组或变量值的变动范围较大、变量的不同取值个数较多的离散变量分组的情况。本案例中 A 产品产量是离散变量,有 56 个不同数值,适合采用组距式数列。

组距式数列有等距数列和不等距数列两种形式可供选择。等距数列适合标志值变动比较均匀或不同总体单位标志值相对比值较小的情况,本案例中分布比较均匀、不同数值的相对比值较小,仅为 $158/91 \approx 1.74$,符合等距数列的条件和要求。

3. 确定组距式数列的参数

编制组距式数列时,基本参数组数、组距和组限需要确定。

(1) 确定组数。由于本例题中生产工人日产量没有广泛认可的确定组数的规定或惯例,其组数 K 的确定借用 Sturges 提出的经验公式: $K=1+\frac{\lg n}{\lg 2}=1+\frac{\lg 135}{\lg 2} \approx 8$, 可以将该企业 135 名生产工人按生产 A 产品产量分为 8 个组左右。

(2) 确定组距。由于案例选用等距,组距根据全部数据的最大值和最小值及所分的组数来确定,即组距=(最大值-最小值)÷组数。得到的组距数值要适当缩放,其数值最好为最小计量单位或最大公因数的 5 倍或 10 倍。本案例中,最大值为 158,最小值为 91,则组距=(158-91)÷8=8.375。为便于计算和符合人们的习惯,组距数值可以适当放大,其尾数最好为计量单位的 5 倍或 10 倍,因此本案例组距可选择 10。

(3) 确定组限。先选择组限的形式,本案例中变量值虽属于离散变量,但由于实际操作中,离散变量分布范围较宽,数值个数较多时可采用重合式组限,本案例生产工人按生产 A 产品产量分布范围较宽,不同数值的个数很多,考虑选择为重合式组限。

确定组限第一组下限即组限的起点确定很关键,第一组下限小于最小值,且距最小值不超过半个组距。一般大于可采用 $X_{1下}=\text{最小值}-\text{组距} \div 2$, 然后对组限数值适当缩放,其数值最好为各单位标志值的最小计量单位或最大公因数的 5 倍或 10 倍,本案例 $X_{1下}=\text{最小值}-\text{组距} \div 2=91-10 \div 2=85.5$, 组限数值略微放大,但小于最小变量值。第一组下限数值选择 90 较为合适,为各单位标志值的最小计量单位的 5 倍。 $X_{1上}=X_{1下}+\text{组距}$, $X_{2下}=X_{1上}$, $X_{2上}=X_{2下}+\text{组距} \cdots \cdots$ 以此类推,从而各组组限分别确定为 90~100, 100~110, 110~120, 120~130, 130~140, 140~150, 150~160。

4. 统计各组的频数

在已经分成 7 组并确定各组组限的基础上,将各个工人生产的 A 产品产量归并到各组中去,并遵循“不重复不遗漏”“上限不在本组之内”的原则。然后计算各组 and 总体的合计数,如表 3-20 所示。

表 3-20 某企业 135 名工人生产 A 产品产量情况

A 产品产量	频数/人	频率/%
90~100	6	5.00
100~110	15	13.33
110~120	28	16.67
120~130	43	26.67
130~140	23	18.33
140~150	13	11.67
150~160	7	8.33
合计	135	100.00

习 题 3

- 3.1 统计整理应遵循哪些原则？
- 3.2 统计整理的一般程序是什么？
- 3.3 数据的预处理包括哪些内容？
- 3.4 如何正确选择分组标志？关键是什么？
- 3.5 单项式分组和组距式分组分别在什么情况下运用？
- 3.6 简述组距式分组的步骤，等距分组和不等距分组的频数分布各有什么特点？
- 3.7 统计汇总的组织方式有哪些？
- 3.8 分别比较圆形图和圆环图、直方图和条形图、茎叶图与直方图各有什么区别与联系？
- 3.9 统计表由哪几个主要部分组成？制作统计表应注意哪几个问题？
- 3.10 定类数据和定序数据的整理与图示方法各有哪些？
- 3.11 某企业按工人日产量分组，工人人数分布资料如表 3-21 所示。

表 3-21 工人人数分布资料

日产量分组/件	工人人数/人
50~60	6
60~70	12
70~80	18
80~90	10
90~100	7
合计	53

根据表 3-21 指出：①表中变量数列属于哪一种变量数列；②表中的变量、变量值、上限、下限、次数（频数）；③计算组距、组中值、频率。

- 3.12 某专业二年级 80 名学生统计学考试成绩分别为

68 89 88 84 86 87 75 73 72 68 75 82 97 58 81 54
79 76 95 76 71 60 90 65 76 72 76 85 89 92 64 57
83 81 78 77 72 61 70 81 57 89 49 84 86 87 75 73
72 68 75 82 97 81 67 81 54 79 87 95 76 75 60 90
65 76 72 70 86 85 89 89 64 57 83 81 78 87 72 61

学校规定：60 分以下为不及格，60~70 分为及格，70~80 分为中，80~90 分为良，90~100 分为优。

要求：①将该班学生分为不及格、及格、中、良、优五组，编制一张次数分布表，绘制次数分布，并计算累积频数和累积频率；②指出分组标志及类型；分组方法的类型；分析本班学生考试情况；③绘制直方图、折线图。

- 3.13 下面为 40 个人的血型资料，试绘制其次数分配表，并绘制圆形图、条形图。

O O A B A O A A O B O B O O A O O A A
A A AB A B A A O O A O O A A A O A O O AB

3.14 某地区 35 个工业企业工人人数如下：

322 674 339 357 346 295 465 355 332 316 453 442 417 587
 369 545 323 430 560 528 333 311 410 604 281 461 432 421
 484 473 392 354 499 338 375

试根据上述资料：①分别编制等距及不等距的分布数列；②根据等距数列编制较小制累计和较大制累计的频数和频率数列，并绘制直方图、折线图。

3.15 2007 年我国几个主要城市各月份的平均相对湿度（%）数据如表 3-22 所示，试绘制箱线图，并分析各城市平均相对湿度的分布特征。

表 3-22 2007 年我国几个主要城市各月份的平均相对湿度

%

月份	北京	长春	南京	郑州	武汉	广州	成都	昆明	兰州	西安
1	49	70	76	57	77	72	79	65	51	67
2	41	68	71	57	75	80	83	65	41	67
3	47	50	77	68	81	80	81	58	49	74
4	50	39	72	67	75	84	79	61	46	70
5	55	56	68	63	71	83	75	58	41	58
6	57	54	73	57	74	87	82	72	43	42
7	69	70	82	74	81	86	84	84	58	62
8	74	79	82	71	73	84	78	74	57	55
9	68	66	71	67	71	81	75	77	55	65
10	47	59	75	53	72	80	78	76	45	65
11	66	59	82	77	78	72	78	71	53	73
12	56	57	82	65	82	75	82	71	52	72

3.16 已知某地区 2001—2008 年的地区生产总值数据，如表 3-23 所示。

表 3-23 某地区 2001—2008 年的地区生产总值

年份	地区生产总值/亿元
2001	216.18
2002	266.38
2003	346.34
2004	467.59
2005	584.78
2006	678.85
2007	747.72
2008	795.53

其中，在 2008 年的地区生产总值中，第一产业为 142.99 亿元，第二产业为 391.50

亿元，第三产业为 261.04 亿元。要求：①根据 2001—2008 年的地区生产总值数据绘制线图和条形图；②根据 2008 年的地区生产总值及其构成数据，绘制圆形图和圆环图。

3.17 2017 年 A、B 两市农村居民家庭平均每人全年消费性支出数据如表 3-24 所示，试绘制雷达图。

表 3-24 2017 年 A、B 两市农村居民家庭平均每人全年消费性支出 元

项目	A 市	B 市
食品	2 959	3 712
烟酒及用品	258	262
衣着	731	551
家庭设备用品及服务	749	893
医疗保健及个人用品	513	346
交通和通信	468	527
娱乐教育文化用品及服务	1 242	1 135
居住	578	820
合计	7 498	8 246