

统计汇总的程序设计

— 大量数据处理法



(日) 铃木芳雄 著
崔晓光 译 高拱宸 校

中国统计出版社

统计汇总的程序设计

大量数据处理法

〔日〕铃木芳雄 著

崔晓光 译 高拱宸 校

中国统计出版社

鈴木 芳 雄
統計集計のプログラミング

1975年4月25日发行

发行所 株式会社 ノエマ

本书根据1975年版译出

2929/10

统计汇总的程序设计

[日] 鈴木芳雄 著

崔晓光 译 高拱宸 校

中国统计出版社出版

新华书店北京发行所发行

北京印刷一厂印刷

787×1092毫米 32开本 7.75印张 17万字

1985年2月第1版 1985年2月北京第1次印刷

印数：1—20,000

统一书号：4006·046 定价：1.55元

我 的 推 荐

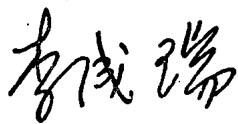
要实现我国统计工作的现代化，大家就要坐下来认真地读几本书。日本铃木芳雄先生所著的《统计汇总的程序设计》是一本好书，我愿意把它推荐给我国有志于统计工作现代化的同志们。

这本书对于统计工作者是很有用的。统计工作者虽然一般自己不搞程序设计，但也需要了解程序设计的要点，否则，你所设计的调查表和统计汇总表就不能适应计算机的要求，不能顺利地编码和录入，也不能很好地通过程序来检验和处理调查资料。这样，就不能很好地发挥计算机的优越性来提高统计的水平。这本书对于电子计算机工作者当然更是有用的。统计汇总的程序设计很简单，这里面大有学问，大有用武之地。如果不认真掌握这方面的知识，吸收这方面的实践经验，不搞好系统分析，要把大量统计资料按规定的要求汇总出来就会遇到很多困难。

这本书是在总结日本人口普查数据处理经验的基础上写出来的。在具体运用上，当然要结合我国的实际情况。在这里我提出一个愿望：在不久的将来，我国的同志们也能总结我国人口普查数据处理的经验以及其他统计数据处理的经验，写出具有我国特色的科学著作来，供大家学习、参考。

铃木芳雄先生曾经担任日本总理府统计局制表部电子计算课课长。他具有丰富的经验和学识，并对中国人民怀有深厚的

友好感情。一九八〇年他曾应邀访华。我在一九七九年和一九八四年两次访问日本时，都曾听取过他对有关情况的介绍。今天，铃木先生的著作在我国翻译出版，是他对中日两国统计学术交流的一个贡献，特在此表示感谢！同时，也感谢本书的译校者崔晓光同志和高拱宸同志。

A handwritten signature in black ink, reading '高拱宸' (Gao Gongchen) from right to left.

1985年1月15日，于北京。

前 言

很早以前，人们就想利用机器进行计算，这一愿望由法国人帕斯克爾和德国人萊浦尼特实现了。但他们制造的计算机很简单。从那时的计算机到今天的电子计算机已经历了几百年的时间。

1945 年在美国宾西法尼亚大学诞生了最早的“恩尼阿克”(ENIAC)计算机，它主要是用于实现科学计算的高速处理。据说“恩尼阿克”计算机一天进行的计算量相当于三百天的手工计算量。从“恩尼阿克”以后，计算机的发展异常惊人。过去，计算机主要用于科学计算，而现在已开始转向事务计算或数据处理方面。

统计处理的范围涉及到科学计算和事务计算两个方面，而这两个方面的计算，可以说是最适合于用计算机来处理。关于科学计算方面的统计分析，已有各种理论书和程序设计参考书出版，而关于事务计算方面的统计汇总及程序设计参考书的出版就很少，至于大量数据汇总方面的书籍则更是凤毛麟角。

汇总这一作业是由人和计算机共同进行的，如单从计算机本身来看，好象并不复杂，因而没有什么理论可言，但要掌握系统的技术，那确实不容易。也许是这个原因，对作者来说也同样，与其说是已经掌握了汇总系统的理论，倒不如说是通过

效法前辈的作法才熟悉并掌握了它。计算机出厂时只不过是一个铁块（一台裸机），只有通过用户编制的程序，它才能发挥出本来的功能。但即使是简单的汇总处理，也由于用户所处的环境不同，设计的方针当然也不同，这就要求以汇总理论作为依据，设计用户独特的系统。本书的目的就是为设计用户的独特系统提供理论上的基础知识，并以大量数据的汇总系统作为中心来表述。

大量数据的处理，要受许多因素的制约，如处理时间的限制，再运行的措施，程序数量的多寡等，而这些因素又互相制约，设计并不那么简单。如果掌握了这个处理系统，然后再来看少量数据的处理，仅从其制约因素少这一点来说，其设计自然是容易得多。基于这个理由，本书才选择了把大量数据处理作为论述汇总系统理论的题材。

本书内容共分七章，第一章叙述了汇总系统的概要，介绍了汇总系统的五大功能和两种汇总系统；第二章至第六章分别说明了每一功能程序设计的要点；第七章列举了简单统计表汇总系统的实例。以上是本书的内容。论述时使用图表较多，实际上是考虑到本书是供对程序设计有经验的人阅读的。因此，关于程序设计的入门知识略而未谈。

最后，本书完成之际，谨对总理府统计局的永山贞则先生以及宫本皓次、山下登、安野胜吾、松井博、松村西泽诸先生的协助，对承担本书出版的诺埃玛社的门山允先生的厚意深表感谢。

1975年3月 铃木芳雄

目 录

前言

第一章 汇总系统概述

- § 1-1 统计汇总和统计分析..... (1)
- § 1-2 统计调查的制表顺序..... (2)
- § 1-3 统计表的结构..... (15)
- § 1-4 统计汇总的五大功能和汇总系统..... (19)
- § 1-5 复合型功能汇总系统..... (22)
- § 1-6 分离型功能汇总系统..... (24)

第二章 检验程序的设计

- § 2-1 检验程序的基本功能..... (28)
- § 2-2 检验功能的种类..... (30)
- § 2-3 错误数据的修改..... (39)
- § 2-4 错误数据的处理形式..... (44)
- § 2-5 错误数据处理的标准化..... (48)
- § 2-6 输出数据的信息..... (59)
- § 2-7 表格检索技术..... (64)

第三章 分布程序的设计

- § 3-1 分布程序的基本概念..... (72)
- § 3-2 分布状态..... (74)
- § 3-3 表格映象的分配..... (77)
- § 3-4 合计栏的设置和删除..... (84)
- § 3-5 矩阵代码和位移..... (88)
- § 3-6 识别代码和分布项目..... (92)

§ 3-7	逐级合计	(94)
§ 3-8	记录集的输出	(96)
§ 3-9	标准程序框图	(100)
第四章 合计程序的设计		
§ 4-1	合计程序的概要	(107)
§ 4-2	合计(表外合计)前的分类	(110)
§ 4-3	合计程序的形式	(117)
§ 4-4	RID 的设置	(127)
§ 4-5	标准程序框图	(129)
第五章 预编辑程序的设计		
§ 5-1	预编辑程序的目的	(135)
§ 5-2	预编辑程序的功能	(136)
§ 5-3	预编辑程序的处理	(149)
§ 5-4	预编辑程序的构成	(155)
§ 5-5	预编辑程序和子程序	(161)
§ 5-6	预编辑程序的流程图	(172)
第六章 编辑程序的设计		
§ 6-1	编辑程序的目的	(176)
§ 6-2	编辑程序的功能	(177)
§ 6-3	编辑程序的流程图	(189)
第七章 汇总系统设计示例		
§ 7-1	应汇总的统计表	(192)
§ 7-2	检验程序的设计	(198)
§ 7-3	分布程序的设计	(210)
§ 7-4	合计程序的设计	(217)
§ 7-5	预编辑程序的设计	(223)
§ 7-6	编辑程序的设计	(228)
译后记		

第一章 汇总系统概述

§ 1-1 统计汇总和统计分析

无论是政府机关或是民间，引入电子计算机后，最先开展机械化业务的就是统计工作。事实上，从对计算机利用情况的调查来看，显然也是统计和薪金计算方面的利用居多。

统计处理，是笼统的说法，如果从处理的形态来看，它可分为统计汇总和统计分析两种。

由于统计是表示整体特性的，而为了了解某整体的特性，首先要对该整体的每个数据，如销售统计的销售传票、货运统计的发货单等原始资料进行汇总。根据汇总情况，才能掌握销售额有多少，货运量有多大。象这样通过逐个数据的汇总来掌握整体情况的过程，就叫统计汇总。

另一方面，使用汇总的数字进行整体现象的分析，叫做统计分析，例如算出比例和指数等对两个以上整体进行比较，观察随时间推移所发生变化的动态数列分析，对两种标志间相互关系的分析，等等。

如上所述，汇总和分析是表里一体的。但从计算机处理的角度来看，统计分析方面，各种计算例程序①或计算程序都

① 按适当顺序排列的能使计算机执行某种任务(如解题、检索数据或对一个系统进行控制等)的指令的集合，称为例程序——译者注

以程序库的形式准备妥当，计算程序作为一个系统已经完备，所以在使用时，就用不着很深的计算机知识了。而统计汇总方面，(计算程序)作为一个系统而准备妥当的情况较少，大多数的情况下，计算机的用户必须根据各种各样的统计表设计相应的程序。

统计汇总，由于以下种种理由而难以系统化：第一，统计表要平易简明，要使人人都能理解，编制程序时要注意制表逻辑简明易懂，这就要靠担任设计的程序员作出判断；第二，即使同样是统计汇总，因数据量、统计表种类及汇总项目多少等方面的不同，处理所需时间也有显著差别。因受处理时间的制约，难以采用整齐划一的办法，大概正是出于这个理由，才不得不个别地考虑各种不同的汇总系统。

当然，并不是说统计汇总就没有通用的程序或程序包。如果统计汇总的对象是那些数据量较少而且内容简单的统计表，那么通用程序或程序包还是很多的。但在数据量很大，统计表复杂、繁多的情况下，通用的汇总系统在表现能力和处理时间上还存在一定问题，因而有时不得不放弃应用。

由于统计分析方面的系统已很完备，以数值分析、统计分析为主的各种参考书也容易得到，这些方面的问题本书不拟再加讨论，腾出篇幅着重阐明系统方面尚不完备，或通用系统很难适应的统计汇总工作应如何对待，尤其将以大量数据汇总为核心，说明它如何处理。

§ 1-2 统计调查的制表顺序

在研究汇总系统之前，首先有必要了解一下统计调查的制

表(汇总)顺序。图 1—1 是制表顺序的简图。本书不是从统计调查规划者的立场进行阐述,而是从制表规划的立场出发进行若干解释。

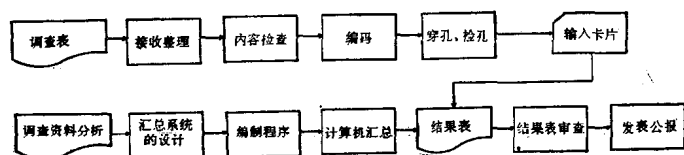


图 1—1 统计调查的制表顺序

图 1—1 的说明如下。

一、调查表(调查单)

在统计调查中,通常是使用调查表进行调查的,调查表设计的好坏,直接关系着统计调查的成败,因而设计调查表是极为重要的工作。

设计调查表时要以下列三个问题作为重点。

1. 所提的各个问题,是否已将调查目的具体化;
2. 提问的方式是否得当;
3. 是否考虑到汇总的便利。

关于第一个问题,不仅要有统计学方面的知识,而且还要有与该项调查有关的经济学、社会学、人口学方面的广泛知识,还要熟悉我们日常生活中的习惯、常识等。

关于第二个问题,要根据调查对象的文化程度而有所不同。一般说,调查对象(申报人)大多数是与统计很少接触的人,所以在设计调查表时,必须考虑到要使人们在填写调查表时不致慌乱出错。为此,提问的文字要避免使用专业词汇,要写得尽

可能简明(避免冗长),还要了解申报人对哪种方式的提问容易回答等,并注意问题的安排。这些方面对提高登记的精确度,都是不可缺少的措施。

随着调查表一起散发“填表说明”一类的登记须知也好,在调查表的空白处印上提示性说明也好,申报人不一定去看。因此不能把希望过高地寄托在申报人的文化程度上。最根本的还是要坚持以问题本身容易填写为前提,只有这样,取得的资料才是可靠的。

第三个问题,关于汇总的简便程度。不论是手工汇总还是机器汇总,为了有效地进行汇总,都要在调查表上编码。机器汇总时,还要根据调查表进行诸如穿孔之类的工作。这样,调查表就成了进行汇总的原始资料。因此,调查表的设计考虑到汇总与否,汇总的效率大不相同。为此,在设计调查表之前,最好听一听统计汇总部门的意见。

在理解了以上的设计基础之后,我们再来看一下调查表格式的设计要点。调查表格式请参看本书第239页。

(一) 调查项目的安排

调查项目的安排,要考虑到申报人登记的便利。这就是说,如果填写比较容易,就可以减轻申报人的心理负担,缩短登记的时间,减少登记的误差和遗漏。

为了使登记简便宜行,调查项目可按以下方式安排:

1. 登记项目要在调查表上集中安排,不能分散。

因为分散容易引起“漏登”;

2. 类似整理号码之类的分类项目要集中置于调查表的上方;

3. 登记项目按从上到下, 从左到右的登记顺序排列;

4. 项目的安排要便于依次登记。

(二) 登记项目名称的印刷

为了防止登记栏内的“登记差错”, 要印出登记项目的名称, 最好顺便把要登记的数据项目单位也印出来。

(三) 采用框格形式设计

全部登记项目都要在指定的框格内登记, 这种框格登记的方式称为框格形设计。由于框格形设计的采用, 展览了登记栏, 减少了“漏登”, 穿孔员的操作也变得轻松些。

(四) 不动文字的印刷

在登记项目里要从几个固定的答案中选择一个答案时, 预先在调查表上印上答案, 可以大大节省登记时间。印好的答案叫做不动文字。例如在调查表上, 给问题的答案编上号码, 申报人在确认答案之后, 便在该答案号码上划个圆圈。

(五) 穿孔项目的标记

当穿孔项目和非穿孔项目一起存在的时候, 要考虑把穿孔项目集中在调查表的某一部分并在周围划上粗线, 或者在非穿孔栏内“布网”(即标上网状记号)。这样就使得穿孔员易于操作, 并减少穿孔错误。

图 1-2 就是布网示例。这里, 登记是以一千日元为单位的, 为了防止登记时发生单位的错误, 设计时规定了小到一元的码位格, 并事先印上零。由于穿孔的项目要以千日元为单位, 因此就在百位以下的三格里标上网状记号。

(六) 穿孔栏区分线的印刷

〔关于住宅等建筑物的购入、卖出、新建、改建和土地买卖的调查〕

1973年1年内买卖的住宅等建筑物和土地以及新建、改建的建筑物的
金额是多少？

住宅等建筑物 (购入、新建、改建价) (实价) (代码)

	千 百 十 亿 万 万 万 万 千 百 十 一	千 百 十 亿 万 万 万 万 千 百 十 一	
(1) 家庭用.....	元 0 0 0	元 0 0 0	5 1
(2) 事业用.....	元 0 0 0	元 0 0 0	5 2

土 地

(买价)

(实价)

	千 百 十 亿 万 万 万 万 千 百 十 一	千 百 十 亿 万 万 万 万 千 百 十 一	
(1) 家庭用.....	元 0 0 0	元 0 0 0	5 3
(2) 事业用.....	元 0 0 0	元 0 0 0	5 4

您家有购入住宅等建筑物和土地或新建住宅的计划吗？请用“○”把合适的号码
圈上

1. 今后3年内购入或新建
2. 3年以后购入或新建
3. 尚无计划

图 1-2 布网示例

①	在这个企业就职时一个月的薪金	万	千	百	十	元
	(请登记包括超勤津贴的纳税额)					

图 1-3 金额栏设计示例

在穿孔项目各位之间印上虚线(图 1—3)，对数量、金额等项目每三位印上一条实线，这样就可以减少申报人的登记错误和穿孔员的穿孔错误。

(七) 调查表的印刷颜色

选择调查表的印刷颜色时，要尽可能避免给穿孔员的眼睛造成疲劳。

印刷颜色虽有黑、蓝、绿、紫、茶色等，但以绿色和茶色为佳。

(八) 调查表的尺寸

调查表大小的决定要考虑到穿孔时容易操作，以及保管空间和业务部门间的交接等问题。

二、接收整理

汇总第一阶段由接收和整理调查表开始。一般情况下，调查表密封后送到汇总机关，汇总机关启封时，要依照交接单对调查表是否安全送到进行核实，然后将这些调查表整理成册，以利于以后的制表工作。这个过程叫做调查表的接收和整理。

每册的单位既要根据内容检查、编码、穿孔等工作的方便和调查表的收藏、搬运的便利而定，还要考虑到调查表的大小。一般以 500 张左右为一册比较适当。当分好册后，要用较厚的纸附在调查表前后作为封面、封底，以防调查表的损坏。封面要注明调查的基本属性，如都道府县名称、市街村名称、调查区号码、册的号码、调查表的张数等。这样，即使汇总时发生问题，查找调查表也容易。

三、内容检查

内容检查就是对调查表的登记内容，如漏登、错登、与其

他调查项目是否有矛盾等进行检查，如有问题就要更正。内容检查是在收回调查表时由调查员进行的，为了尽可能彻底地发现错误，需要加倍仔细。内容检查有两种，即对每一调查项目进行个别检查和对多个项目之间的矛盾进行相关检查。在个别检查时，要检查必须回答的问题回答了没有，只能选择一种答案的是否选择了两种以上的答案，或者做了完全错误的回答。另一方面，经过个别检查而未发现错误的项目，在各相关项目结合检查时还可能出现矛盾，检查这些矛盾正是相关检查的任务。比如，三岁的煤矿工、男性助产士就是属于这类矛盾。

无论是个别检查还是相关检查，只要发现了错误就必须改正。如能从调查表登记的项目中得到正确的回答当然好，但实际上却相当困难，很多事情不询问申报人是搞不清楚的。

四、编码

无论是手工汇总还是机器汇总，原封不动地对调查表上的答案进行汇总，其效率是很低的。因此，就必须进行编码作业以便汇总。在调查项目里使用不动文字时，回答的同时已经编上了代码，所以没有必要重新编码。但对行业、职业这类项目的编码，则需要有专门的知识。对这类问题，申报人很难用代码直接回答。所以行业、职业要尽可能简明扼要地登记下来，然后通过专业人员进行编码。

在机器汇总的情况下，编码要考虑到穿孔的方便。为防止穿孔员误读，字要书写清楚，代码的位数要一致（采用两位码时，不要省略前位的零，如2要写成02）。为避免增加击键的次数，要尽可能减少代码的位数。为此就要进行编码体系的设计，并开展有关工作。