卜凯农户调查数据汇编(1929～1933)丛书

卜凯农户调查数据汇编(1929～1933)(江苏篇)

胡 浩　钟甫宁　周应恒/编著

科学出版社

卜凯农户调查数据汇编(1929～1933)丛书

卜凯农户调查数据汇编
(1929～1933)(江苏篇)

胡 浩　钟甫宁　周应恒/编著

科学出版社
北 京

内 容 简 介

本书是基于留存于南京农业大学经济管理学院的卜凯调查中间统计表，还原整理出的农户层面的数据。主要包括农户家庭规模、农场劳动力利用、农作物生产、储蓄借贷、土地租佃等内容。由于数据量庞大，我们将其中主要内容分为江苏、浙江、山东等13册出版。卜凯基于这次调查，形成了《中国土地利用》这本巨著。本书虽然未能完全重现当年的全部调查内容（留存于南京农业大学的只有22个表头的中间统计表，其他部分不知所踪），但也填补了20世纪30年代统计的空白，因为除了满铁调查资料以外，目前未发现系统的中国农户调查数据。

本书能够为相关研究提供民国时期的农业生产、农民生活、农村社会等基础资料，可供经济学、社会学、农学、历史学等学者参考借鉴。

图书在版编目（CIP）数据

卜凯农户调查数据汇编. 江苏篇: 1929～1933 / 胡浩, 钟甫宁, 周应恒编著. —北京：科学出版社，2017.3

卜凯农户调查数据汇编(1929～1933)丛书

ISBN 978-7-03-046731-7

Ⅰ. ①卜… Ⅱ. ①胡… ②钟… ③周… Ⅲ. ①农户经济-农村调查-统计数据-江苏省-1929～1933 Ⅳ. ①F32-66

中国版本图书馆CIP数据核字(2015)第303772号

责任编辑：魏如萍 王丹妮 / 责任校对：李 影

责任印制：霍 兵 / 封面设计：无极书装

科学出版社出版

北京东黄城根北街16号

邮政编码：100717

http://www.sciencep.com

北京通州皇家印刷厂印刷

科学出版社发行 各地新华书店经销

*

2017年3月第 一 版 开本：787×1092 1/16

2017年3月第一次印刷 印张：27 3/4

字数：658 000

定价：158.00元

（如有印装质量问题，我社负责调换）

序　言

卜凯先生是一个传奇人物。他在金陵大学农业经济系工作二十多年，开创了中国农业经济学科的正规大学本科和研究生教育，主持了最早、规模最大的农户调查，是利用现代统计学方法研究中国农户经济的先驱者。卜凯先生根据 1929～1933 年农村调查的数据出版的经典著作《中国农家经济》和《中国土地利用》在国内外学术界的影响历久不衰，近年来更引起学者重新深入分析的浓厚兴趣。由于卜凯先生的两部著作只包含对汇总数据的分析，而现代的数量分析更看重农户层面的微观数据，这就有必要寻找和发掘当年农户调查的原始资料。南京农业大学经济管理学院是当年金陵大学农业经济学的继承者，也是当年调查数据最可能的保存之地，因而担负着挖掘这一宝库的历史责任。经过十多年的艰辛工作，总算可以把初步成果的一部分呈现在读者面前了。

卜凯先生全名约翰·洛辛·卜凯[①]（John Lossing Buck），1890 年出生于美国纽约州德彻斯县快乐谷的一个德裔农户。19 岁那年卜凯进入康奈尔大学农学院，同学中有胡适、赵元任、过探先、邹秉文、吕彦直等中国学生，1914 年卜凯毕业，次年受美国长老会海外传教协会委派到中国安徽宿州传教并从事农业推广工作。卜凯本来就对古老的中华文化十分入迷，宿州的工作给他提供了亲身了解中国普通农民生活的机会，与赛珍珠的婚姻（1917 年）更为他走访农家增添了非常难得的翻译，使其可以更准确地了解农家生活并开展初步的调查工作。

1920 年卜凯先生应康乃尔大学校友、金陵大学农林科主任芮思娄（J. H. Reisner）的邀请，到金陵大学创立中国第一个农业经济系，也为卜凯先生之后组织大规模农户调查奠定了基础。从 1922 年夏天开始，卜凯要求选修“农场管理学”的学生必须回家乡调查 100 户以上的农家经济情况，到 1930 年农家调查共完成 7 个省份 17 个地区 2866 户，卜凯将所有资料汇总写成《中国农家经济》一书。费正清主编的《剑桥中国晚清史》和《剑桥中华民国史》的近现代农业和农村史部分，主要资料即来源于此。

1928 年亚太地区非政府组织太平洋国际学会委托金陵大学农业经济系主持实施一项名为“了解农村社会现实而为农业改进提供依据”的土地调查，每年提供 10 000 美元的经费。卜凯将全国分为十几个区，每区设一名调查主任，下有调查员数百人，深入农户实地调查；然后把实地调查取得的数据和资料计算汇总并进行分析，形成了名为《中国土地利用》的巨著。1936 年《中国土地利用》英文版由商务印书馆出版，1937 年日本出版了两种日译本，1938 年又在美国芝加哥大学出版社出版，1941 年由卜凯的弟子乔启明、邵德馨、黄席群、孙文郁、杨铭崇等教授译成中文，并在成都成城出版社出版。

现在呈现在读者面前的就是对当年调查数据重新整理所得到的初步成果。如上所述，卜

① 卜凯先生的中文译名很多。根据卜凯先生当年使用的名片复制品和第二次婚姻的结婚证书影印件（均由卜凯先生儿子提供），可以确定他本人使用的正式译名为卜凯。

凯先生的两本经典著作只包含汇总数据，而现代的统计分析更注重农户层面的微观数据，因此最理想的情况是找到原始调查表格。很可惜，我们只找到分类统计的中间表格。这些表格不是当年直接入户调查时使用的原始登记表，而是回来后把调查内容分类列表，如土地利用状况、产量、产品使用量等，以类别为划分标准，每张表上只列出本类内容的少数项目供分析使用。好在这些表格同时分地区列出每一个农户的数据，所以还可以尝试恢复以农户为对象的完整微观数据。从 2000 年开始，我们开始这项恢复农户数据的繁重工作。最初的工作是与日本东京国际大学几位统计学权威教授合作，试图先拍照再利用计算机软件把图形转换成文本数据；但现有软件无法辨识手写数字，只能采取手工输入的办法建立数据库。如果直接依据原始表格人工输入，已经保存了近 80 年的纸张很可能损毁，所以，只能先扫描原始表格，然后根据扫描件把一个个数字手工录入电脑，最后再设计相应的程序,将数万张表格中的分类数据恢复成农户层面微观数据库。

这十多年工作的艰难一言难尽：辨识已经褪色的手写数字，校对由大量空行、空格带来的录入错误，不同地区不同度量衡和货币单位的统一，等等，实在难为了十多位担任录入和整理工作的研究生。现在总算看到初步成果了，经过整理的农户资料将陆续出版，数据库将投入使用，数据的挖掘分析也将次第展开，可以预见今后将产生出一系列重要的研究成果。

卜凯先生的学术研究和贡献不仅仅限于农户调查。半个多世纪以前他就指出农民收入问题的根源不在于农业生产力本身；任何一个生产者的收入取决于能利用的资源，因而农民收入取决于能够利用的资源是否足以产生必要的收入。他的研究也不局限于农业经济学领域。抗日战争时期他担任美国财政部驻华代表期间曾经给时任美国总统的罗斯福提出警告，指出美国白银收购法案导致中国白银大量外流，由于中国实行银本位制，大量白银外流导致的通货膨胀将无法控制，必然会动摇中国的社会和政治稳定。20 年后，著名的美国经济学家、诺贝尔经济学奖获得者弗里德曼提出同样的看法，即美国的白银收购法案是导致国民党政府垮台的重要因素之一。我们的工作也是对卜凯先生的纪念。

在这里，我们要感谢日本东京国际大学的松田教授及其领导的团队在工作前期做出的努力和提供的经费资助，感谢美国康乃尔大学的 Calum G. Turvey 教授在整理和分析农户数据过程中提供的帮助与合作，也感谢南京农业大学人文社科处后期提供的大量经费，保证我们能顺利完成现在的工作。当然，我们也要感谢南京农业大学经济管理学院为这一看上去枯燥无味的工作付出大量时间和心血的研究生，特别是郑微微、于敏捷、虞祎等，没有他们的贡献，不可能获得现在的成果。

钟甫宁

2017 年 2 月

目　录

卜凯调查原始数据的挖掘与整合

胡　浩　郑微微　于敏捷　虞　祎　钟甫宁　Calum G. Turvey
（南京农业大学　康奈尔大学）

南京农业大学经济管理学院保存着卜凯（John Lossing Buck，1890—1975）教授1929～1933年农村调查的中间统计数据，卜凯教授正是基于这些数据书就了农业经济学的经典著作《中国农家经济》和《中国土地利用》（分三册，分别为论文集、地图集和统计资料），“不特材料丰富，持论亦复公允，盖一切论断完全根据于调查所得之数字，故其准确程度，远非一般仅能代表个人观感之著作所能同日而语也”[①]。对于卜凯的农村调查，国内外很多学者给予了很高的评价，梁方仲曾言：“它是该领域的第一个研究，第一次试图如此全面、系统地研究这样一个深广的课题。”这两部著作问世以来，一直被西方及中国香港、台湾学术界誉为了解中国农村的经典著作而被广泛引用，如周锡瑞、黄宗智、马若孟、Mark Elvin、Linda Gail Arrigo、Roll、张五常等都不同程度地利用卜凯调查在自己的研究领域取得了不菲的成绩，但凡是研究民国时期农村社会、农家经济的研究几乎都将此作为最重要的史料加以利用。

前人对20世纪前后中国农业的研究基本是基于上述两本著作，因为目前尚没有出现比这更全面的农户调查数据。事实上，对这一历史时间中国农业的研究一直是经济史学界的研究热点，因为该历史时期是中国历史上承前启后的关键时期，向前追溯是对中国传统经济的再评价，在一定程度上解释整个中国经济历史；向后推导涉及如何评价西方对中国经济的冲击，以及整个中国革命的效果。然而由于数据原因，对这一历史时期的研究一直存在分歧。如果能够在保存于南京农业大学的中间统计表的基础上，还原整理出农户层面的数据，将填补民国统计的空白，为相关研究提供民国时期的农业生产、农民生活等基础资料，以便验证研究分歧，加深对该历史时期的深刻理解。本书将着重介绍对这些原始数据的挖掘整理过程，在数据录入、整合、校对的基础上，借助数理分析和统计学工具，对数据重新进行分类，通过与《中国农家经济》和《中国土地利用》中的图表比对，解决单位不统一等问题，最终建立农户数据库，实现使用者与数据的对接。

一、工作基础

2000年前后，南京农业大学发现了卜凯农村调查的原始资料（household micro data），

① 出自谢家声、章之汶为《中国农家经济》作的序。

保存状况总体良好，但部分资料出现残缺、字迹模糊等情况，如果不及时整理，将失去对这部分资料进行整理的机会。2002 年 11 月，南京农业大学经济管理学院与日本东京国际大学合作开发卜凯大规模农村调查数据，但因为数据量庞大，最终仅利用了其中部分数据对特别支出和土地生产率进行了分析。2003～2006 年，南京农业大学继续和日本东京国际大学进行国际合作研究，2007 年又获得东洋文库的资金帮助，重新进行卜凯调查原始资料的电子化活动，至 2008 年年末，以数码照片扫描的形式完成了全部资料的电子化（图 0-1）。然而，因为资金和人力不足，原始资料仅以扫描件的形式保存，未进行进一步的整理与利用。2011 年，南京农业大学人文社会科学重大项目启动，使该项整理工作得以继续实施。

卜凯的两本著作中分别记载了卜凯的两次农村调查范围，第一次是 1921～1925 年，调查了 7 个省份 17 个县 2866 个农户；第二次是 1929～1933 年，调查了 22 个省份 168 个县 16 786 个农户，调查范围史无前例，除了吉林、黑龙江、海南、西藏、新疆，其他省份都覆盖。

南京农业大学现存的原始资料共有 118 包装盒、24 956 页。原中间统计表的标题共有 189 个，去掉重复，统计表标题为 86 个，经整理后确认原始调查资料覆盖 172 个样本县，其中 4 个县为卜凯第一次调查的样本县，剩余 168 个县为卜凯第二次调查的样本县，与《中国土地利用》对应。

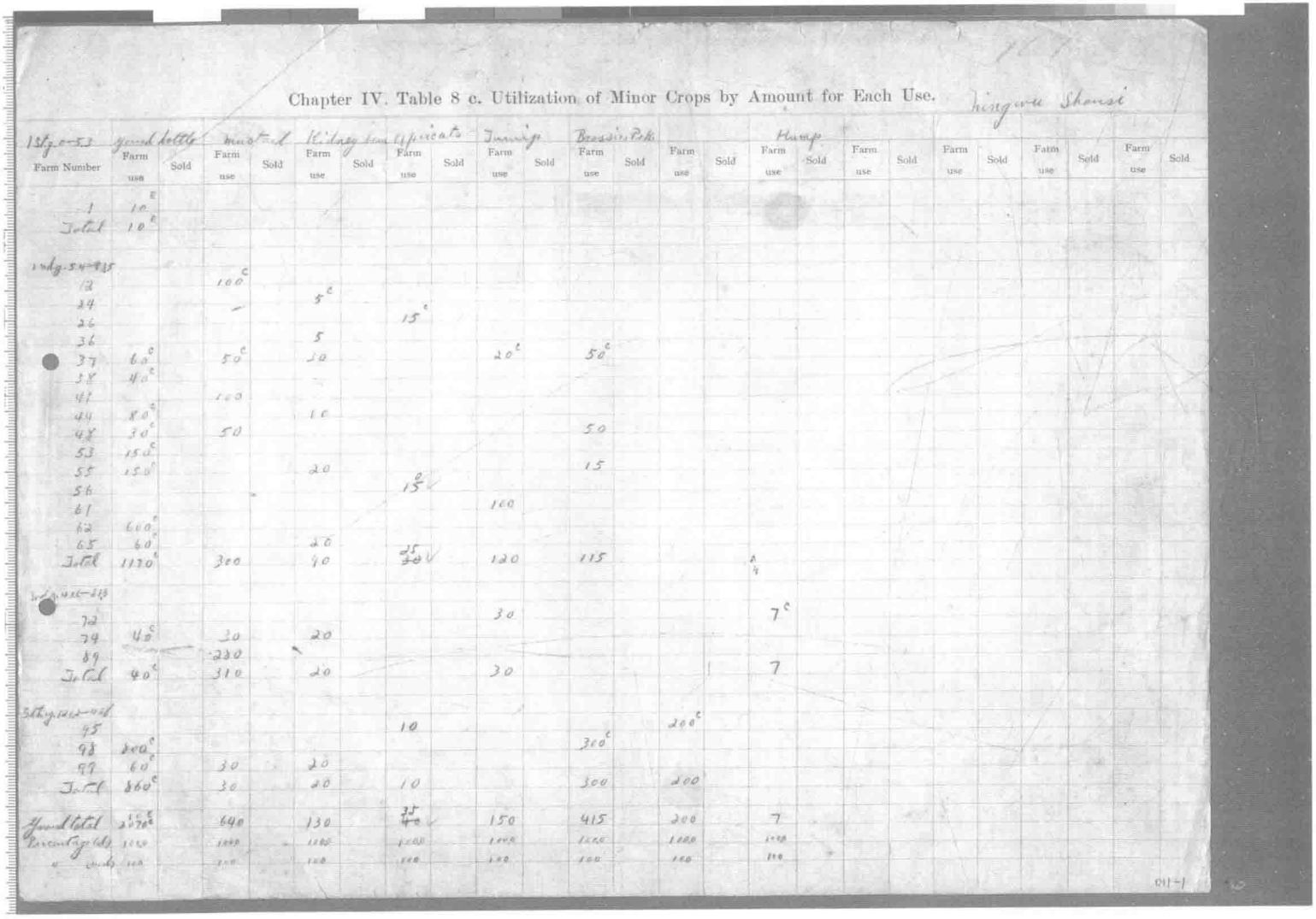
现存的 83 个中间统计表中，各个标题（调查内容）的保存状况不同。而且，中间统计表不是每个农户的问卷，而是该调查项目下农户的调查数据及合计等，如图 0-1 所示。依据中间统计表，课题组可以获知农户的原始数据。

二、主要工作内容

（一）基础数据录入与核对

基于原始资料的电子扫描图片，课题组组织了 80 多名二年级以上农林经济管理专业的本科生及 20 多名研究生进行基础数据录入。录入过程要求将每张电子图片中的所有信息（包括标题、地区、变量、数据等），按照电子图片固定的数据表格格式录入 Excel，一张电子图片对应一页 Excel，每页 Excel 以对应的电子图片编号命名，一盒电子图片文件对应一个 Excel 文件。同时，课题组对已经录入的基础数据进行两遍以上的逐字核对，尽可能减少由人工录入导致的低级错误。在此基础之上，课题组还将所有英文标题（包括标题、地区名、变量等）翻译成中文，最终完成英文版与中英文版两种版本的基础数据电子化工作。

梳理录入 Excel 的基础数据过程中，课题组发现中间统计表的数据内容可分为两大类，一类是标有农户编号的农户统计表，另一类是标有地区信息的县乡统计表。其中，22 个中间统计表标题属于农户信息表，剩余 61 个中间统计表标题属于县乡统计表。而县乡统计表中的数据与卜凯 1937 年出版的《中国土地利用（统计资料）》中的数据基本一致，因此，课题组暂时将针对 22 个中间统计表标题下的农户信息表进行电子化与数据库建设（表 0-1）。



Chapter IV. Table 8 c. Utilization of Minor Crops by Amount for Each Use. Jingyuan Shanxi

Farm Number	ground kaoliang Farm use	Sold	mustard Farm use	Sold	Kidney bean (Phaseolus) Farm use	Sold	Turnip Farm use	Sold	Brassica Pekinensis Farm use	Sold	Hemp Farm use	Sold	Farm use	Sold	Farm use	Sold	Farm use	Sold	Farm use	Sold	Farm use	Sold
1	10 E																					
Total	10 E																					
Vlg. 54-835			100 C																			
12			—		5 C																	
24							15 C															
26					5																	
36																						
37	60 C		50 C		10				20 C		50 C											
38	40 C																					
47			100																			
44	80 C				10																	
48	30 C		50								50											
53	150 C																					
55	150 C				20						15											
56							15 C															
61									100													
62	600 C																					
65	60 C				20																	
Total	1130 C		300		40		25/30 V		120		115				1/4							
Vlg. 44-63																						
72									30						7 C							
74	40 C		30		20																	
89			230																			
Total	40 C		310		20				30						7							
3 Vlg. 44-63																						
45							10						200 C									
93	300 C										300 C											
97	60 C		30		20																	
Total	360 C		30		20		10				300		200									
Grand total	2070 E		640		130		35/40 V		150		415		200		7							
Percentage (%)	100		100		100		100		100		100		100		100							
" Cards	100		100		100		100		100		100		100		100							

011-1

图 0-1　卜凯农村调查原始资料电子化图例

表 0-1　农户信息表标题

序号	标题
1	Size of Family　农户家庭规模
2	Able-Bodied Men　健全男子在一年内的工作情况（15 岁以上、60 岁以下）
3	Proportion of All Farm and Subsidiary Works Performed by Family and Hired Labor, by Men, Women and Children　农场家庭劳动力和雇佣劳动力中男工、女工和童工分别从事农场工作和副业的工作时间
4	Amount and Distribution of Live Stock　各种家畜的数量
5	Farm Area Devoted to Different Uses Grouped by Size of Farm　农场中不同用途的土地面积
6	Number and Area of Graves in Farms　农场内坟墓的数量及其所占的面积
7	Number, Distance and Size (Crop Area in Local Units) of Plots and Fields　田地块数与丘数之大小距离及数量（地方单位）
8	Proportion of Farm Area Rented　农场租用面积的百分比
9	Number of Mow of Crop Area Devoted to Various Crops　不同农作物的耕作面积
10	Amount of Fertility Produced on the Farm　农场肥料生产数量
11	Amount and Kind of Fertilizers Applied per Mow　每亩肥料的使用种类和数量
12	Changes in the Use of Fertilizers　农场施用肥料的变迁
13	Changes in the Use of Fertilizers　施用肥料种类的变迁
14	Yields per Mow of All Crop (in ton &catties)　农作物的亩产量（吨、斤）
15	Most Frequent Yield per Mow of the Byproduct of Important Crops(in catties)　各种主要农作物的副产物每亩的通常产量（斤）
16	Most Frequent Yield of Important Crops by Soil Types and Irrigations　按照土壤类型和灌溉地类型分组的主要作物通常产量
17	Utilization of Crops by Amount for Each Use　各种作物各项用途的数量
18	Utilization of Crops by Amount for Each Use　各种作物各项用途的数量（农副产品）
19	Utilization of Minor Crops by Amount for Each Use　各种次要作物各项用途的数量
20	Savings　储蓄
21	Credit and Indebtedness　借贷和债务
22	Special Expenditures　特殊支出

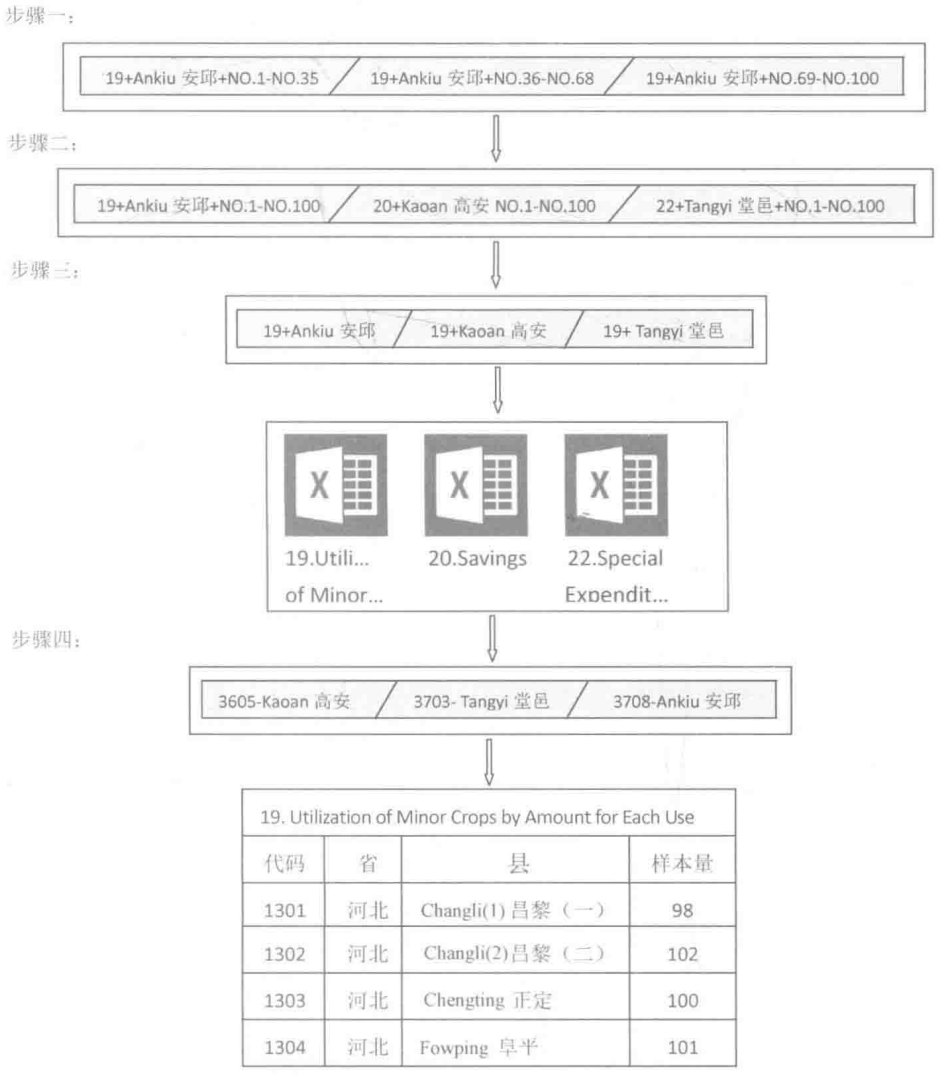
注：本书只摘录了部分卜凯调查原始数据，主要内容为与农业生产相关且完整度比较高的数据。表 12 和表 13 为农场肥料施用变化信息表，因表内肥料变化原因等信息严重缺失，本书中未摘录；表 15、表 18 和表 19 为农副产品及次要作物信息表，受篇幅限制，书中也并未摘录。此外，关于调查地区，本书只摘录了调查内容比较全面的地区，一些调查内容较少的地区在书中也并未摘录。表 0-2 同此

进一步整理 22 个中间统计表标题下的农户信息表发现，相同地区同一标题下的农户信息几乎都不是记录在同一张统计表中，而是分布于好几张中，有些甚至分布于 20 多张统计表中，并且同一标题下的所有农户信息也并不都是分布于同一个文件夹里，有些甚至分布于 20 多个文件夹里。因此，为完整地体现农户调查数据，课题组对同一标题内容下各个地区的农户信息进行整合。

（二）基础数据整合

根据录入 Excel 的基础数据，整合步骤如下：步骤一，将每张 Excel 的农户信息进行确认并命名，命名格式为“标题编码+地区名+农户编号”，如“19+Ankiu 安邱+NO.1-NO.35”；步骤二，基于步骤一，将分布于不同 Excel 中相同标题下相同地区的所有不同编号农户信息整合到同一页 Excel 中，并同样以“标题编码+地区名+农户编号”格式命名，（例如，将 19+Ankiu 安邱+NO.1-NO.35、19+Ankiu 安邱+NO.36-NO.68 和 19+Ankiu 安邱+NO.69-NO.100 合并为 19+Ankiu 安邱+NO.1-NO.100）；步骤三，基于步骤二，继续将分布于不同 Excel 中的相同标题下所有地区农户信息整合到同一个 Excel 中，并以“标题编码”格式命名（例如，将 19+Ankiu 安邱、19+Kaoan 高安和 19+Tangyi 堂邑等合并到同一个 Excel 中，并命

名 19.Utilization of Minor Crops by Amount for Each Use；步骤四，基于步骤三，最后将各标题对应的 Excel 中所有地区赋予统一的地区代码，按照代码进行排序，并于每个 Excel 的第一页制作该 Excel 包含标题内容、调查地区、调查样本总量等信息的目录，为后期的数据库建设奠定基础。图 0-2 为详细的步骤图。



19. Utilization of Minor Crops by Amount for Each Use

代码	省	县	样本量
1301	河北	Changli(1) 昌黎（一）	98
1302	河北	Changli(2)昌黎（二）	102
1303	河北	Chengting 正定	100
1304	河北	Fowping 阜平	101

图 0-2　基础数据整合步骤展示图

以上整合工作使零散的农户基础数据得到了良好的规整，农户调查数据库建设初现雏形，但仍然存在一些由原始手稿保存不善受损、手稿本来存在笔误、手稿字体辨识不清、遗漏等带来的问题，需要做进一步的甄别与修复。

（三）数据甄别与修复

需要进一步甄别与修复的数据大概可分为以下三类。

1. 数值

数值问题主要表现为：手稿字迹辨识不清、手稿修改痕迹反复、遗漏等。较为典型类型如下。

例 1 手稿中多数小于 1 的小数都是用“.数字”来表示，数据录入过程中，有些会因“.”的笔迹模糊或淡化而无法确定该值是否为小于 1 的小数（图 0-3）；手稿中还存在数值辨认不清的问题，如 1 与 7，3 与 5 等（图 0-4）。

图 0-3 手稿样本 1

图 0-4 手稿样本 2

例 2 如图 0-5 与图 0-6 所示，原始手稿修改痕迹反复，容易误导录入工作。

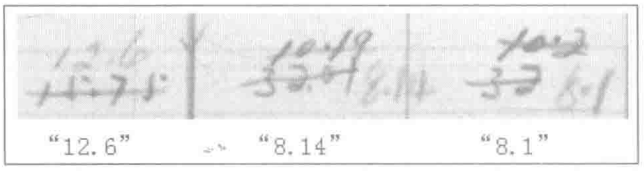


图 0-5 手稿样本 3

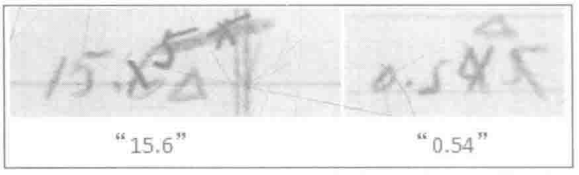


图 0-6 手稿样本 4

上述三种数值问题均可以通过原始手稿中每个地区每一列变量的 Total（总计）与 Grand total（总计）值推算相关模糊数据（图 0-7）。

图 0-7 手稿样本 5

例 3　手稿中关于“等成人劳动力”与“农场厩肥生产量”的数值，只有部分地区有记录，很多地区存在数值遗漏现象。

关于数值遗漏问题，对照卜凯的《中国土地利用》可以发现，“等成人劳动力”与“农场厩肥生产量”的数值并非调研所得的第一手资料，而是卜凯根据实地情况进行的数值换算结果。“等成人劳动力，即成年男子为 1 个劳动力，女子为 0.8 个劳动力，儿童为 0.5 个劳动力”，“农场厩肥生产量按每家畜单位年生产 7250 千克计算”。按照卜凯对数值的处理方法，可以对遗漏数值进行修复。

2．单词拼写

单词拼写问题主要表现为：手稿中由于书写习惯差异，单词书写字体不一，容易使母语非英文的录入者辨识不清，如“S”与“R”，“a”与“o”等。该类问题主要存在于作物名称拼写中，课题组的处理方法如下。

如图 0-8 所示，手稿样本 6 为一种农作物副产物的名称，根据字迹辨识出的单词拼写可能为“sope stalk”，词义并非农作物副产物。那么根据农作物副产物源自生产的农作物，课题组对照该地区生产的所有农作物，最后推测出为“Rape stalk”。以此类推，作物名称单词拼写问题均可通过对照该地区的作物生产投入表、产出表及用途表等进行推测。

图 0-8　手稿样本 6

3．中文方言

卜凯的原始手稿中存在不少以“中文方言”形式记录的信息，这些方言体表述与一般的书面语表述存在很大差异。

如图 0-9 所示，手稿样本 7 为江西省德安县的一种以中文方言记录的主要作物，针对作物名称问题，课题组仍然采用通过对照投入表、产出表及用途表进行推测，最后推测出该作物名为“Turnips（白萝卜）”。但此方法只适用于主要作物，一些次要作物的名称问题仍然未能得到准确推测，如“秋子”“土瓜”等，有待做进一步的探究。

图 0-9　手稿样本 7

（四）农户数据深度核对

基于修复后的农户数据，课题组进一步核对不同标题内容下相同地区相同编号的农户信息是否存在一一对应关系。首先，课题组核对了各地区的农户样本总量，发现各标题下相同地区的农户样本总量是一致的；其次，课题组将不同标题下相同变量的农户数据进行抽样核对，如标题“14. Yields Per Mow of All Crop”中的农作物通常产量与标题“16.Most Frequent Yield of Important Crops by Soil Types and Irrigations”中按土壤及灌溉类型分类的农作物通常产量，发现两个标题下相同地区相同农户编号同种作物的通常产量数值是一致的；最后，课题组进一步抽样核对农户农业生产投入（标题 11）与产出（标题 14）、土地租赁（标题 8）与纳租（标题 17）等信息，发现相同地区相同农户编号的这些信息均存在一一对应关系。进而，课题组断定不同标题下相同地区相同编号的农户信息都是存在一一对应关系的，农户信息基本完整。

基于以上判断，课题组再次将农户农业生产的投入（标题 11）与产出（标题 14）、产量（标题 14）与用途（标题 17）、土地租赁（标题 8）与纳租（标题 17）等具有相关性的数据，按照相同地区相同农户编号进行一一核对，进一步检查农户数据中的错行现象，并进行校正。例如，图 0-10 为相同地区相同农户编号土地租赁与纳租信息的合并校对，1～5 号农户中只有 2 号与 4 号农户租入土地，但 3 号与 5 号农户显示有纳租信息，说明数据存在错行，在通过与不同用途的农场面积（标题 5）进行比对后确认土地租赁信息正确，故将 3 号与 5 号的农户纳租数据调整为 2 号与 4 号农户的。

Farm Number	Land Ownership					Crop Farm use
	Ower	Part-owner			Tenant	Rent
		Ower	Rented	Total		
1	11.5					
2		2.4	10			
3	13.5					200
4			15			
5	15.5					300

（a）存在错行

Farm Number	Land Ownership					Crop Farm use
	Ower	Part-owner			Tenant	Rent
		Ower	Rented	Total		
1	11.5					
2		2.4	10			200
3	13.5					
4			15			300
5	15.5					

（b）已经矫正

图 0-10　土地租赁与纳租信息对照图例

此外，课题组还进一步核对挖掘出来的农户数据是否与卜凯的《中国土地利用（统计资料）》中的统计数据一致，样本是否存在缺失。首先，课题组核对农户数据标题与《中国土地利用（统计资料）》中的统计标题，发现《中国土地利用（统计资料）》中的统计标题几乎包含所有农户数据标题，而《中国土地利用（统计资料）》中气候、营养、物价及税则、农产运销、生活程度、人口等部分的数据，在农户数据中未能找到，可能存在缺失。其次，课题组核对农户数据样本量与《中国土地利用（统计资料）》中的样本量，发现各地区的农户数据样本量与《中国土地利用（统计资料）》中记录的调查样本量一致，并且《中国土地利用（统计资料）》中记录的调查数据刚好为该地区农户调查数据的均值、中值或百分比。因此，可以推断除非某样本县关于某项内容的数据确实不存在，否则其中农户样本量完整，而调查内容（气候、营养、人口等）存在整块缺失。已经挖掘的农户数据可通过与《中国土地利用（统计资料）》中的数据进行比较来实现进一步的校对及修复。

（五）度量衡转换率推算

卜凯进行农村调查期间，民间度量衡并未统一。卜凯在原始手稿中记录的衡量单位均为非统一单位，但在对应的《中国土地利用》中均将其转化为统一的公制单位。因此，为保证数据的可比性，在整理农户数据的过程中，有必要对相应的度量衡问题进行探究。课题组对以下几个方面的度量衡转换率进行了推算。

1．面积度量单位

卜凯在原始手稿中以“Mow”来记录各地区的耕地面积，并在《中国土地利用》中记录“每地区土地测量单位不一”，即各地区“Mow”的衡量标准不一。但在《中国土地利用》中这些“Mow”均被统一为“公顷”。因此，课题组通过查阅卜凯的《中国土地利用》材料，发现《中国土地利用（统计资料）》中记录了“Mow”与“公顷”的折合率（用α表示）。最后，课题组将该折合率作为耕地面积单位的转化率。

2．产量度量单位

卜凯在原始手稿中以“T”“C”“O”“P”等来记录各地区不同作物的产量，并且在《中国农家经济》中记录“不同地区衡量质量的单位不一”，有些是重量单位，如“C”“O”，有些是体积单位，如“T”“P”，在《中国土地利用》中这些产量单位均被统一为“公斤”或“斤”。而《中国土地利用（统计资料）》中仅记录了各地区“C”与“公斤”的折合率。那么如何将这些除“C”以外的非统一单位转化成统一的公制单位？以“T”为例，课题组发现：原始农户表中有各地区各种主要作物的平均单位产量，用字母 Y 表示，单位为“T/Mow”，《中国土地利用（统计资料）》中有已经转化为统一单位的各地区各种主要作物的平均单位产量，用字母 X 表示，单位为“公担/公顷”，两者所表达的产值是一致的。因此，课题组构建“T”与“公斤”的转化公式：

$$\beta=\frac{X(\text{公担}/\text{公顷})\times 100(\text{公斤}/\text{公担})}{\alpha(\text{种植面积}/\text{公顷})\times Y(\text{T}/\text{种植面积})}\quad(\text{单位：公斤/T})$$

其他非统一单位“O”“P”等均可按照以上公式进行转化。

为检验以上转化公式是否正确，课题组选取了一个以“C”为单位的作物，按照公式进行转化，最后得到的转化率与卜凯的《中国土地利用（统计资料）》中一致。因此，以上公式可普遍应用于作物产量单位的转化。

3．距离度量单位

卜凯在原始手稿中并未标注衡量距离的单位，如变量“Distance of the farthest plots”和“Average distance”。而《中国土地利用（统计资料）》中有这两个变量的地区均值，单位为“公里”。因此，对照《中国土地利用（统计资料）》，能计算出该未标注单位与“公里”的转化率，基于此各地区的距离值均能转为统一的公制单位“公里”。

4．货币度量单位

卜凯在原始手稿中以“D”（吊）与“$”来表示货币单位，而在《中国土地利用（统计资料）》中以统一的“元”和“银元”表示。因此，对照《中国土地利用（统计资料）》，课题组也可以相应地推算出“D”或“$”与“元”和“银元”之间的关系。