调研工具箱丛书之九

(美) 阿琳·芬克 (Arlene Fink) 著
杨晓泉 译

如何管理、分析和解读调查数据

How to Manage, Analyze, and Interpret Survey Data

(第二版)

中国劳动社会保障出版社

调研工具箱丛书之九 (美) 阿琳·芬克 (Arlene Fink) 著
杨晓泉 译

如何管理、分析和解读调查数据

How to Manage, Analyze, and Interpret Survey Data

(第二版)

中国劳动社会保障出版社

图书在版编目(CIP)数据

如何管理、分析和解读调查数据/(美)芬克(Fink, A.)著；杨晓泉译. —北京：中国劳动社会保障出版社，2004
调研工具箱丛书
书名原文：How to Manage, Analyze and Interpret Survey Data

ISBN 7-5045-4450-7

Ⅰ. 如… Ⅱ. ①芬… ②杨… Ⅲ. ①调查研究-数据管理 ②调查研究-数据-分析 Ⅳ. C37

中国版本图书馆 CIP 数据核字(2004)第 017339 号

中国劳动社会保障出版社出版发行
(北京市惠新东街 1 号 邮政编码：100029)
出 版 人：张梦欣
*
新华书店经销
北京印刷二厂印刷 北京密云青云装订厂装订
850 毫米×1168 毫米 32 开本 4.625 印张 85 千字
2004 年 7 月第 1 版 2004 年 7 月第 1 次印刷
定价：11.00 元
读者服务部电话：010-64929211
发行部电话：010-64911190
出版社网址：http://www.class.com.cn

目　录

学习目标

调查是一个收集信息的系统，这些信息来自人群或者与人群有关，用以描述、比较或者解释人们的知识、态度和行为。调查系统包含为完成收集有效数据而设计的多种活动，如确定信息收集的目标；设计研究方案；准备可靠而且有效的调查手段；实施调查；管理和分析调查数据；报告调查结果。

调查研究人员可以通过询问人们问题而直接收集信息；或者通过检查有关人们思想和行为的书面、口头以及视觉记录，间接地收集信息。调查者还可以通过观察在自然状态或者实验状态中的人群而获取信息。

本书有两个主要目的：教你如何管理调查数据；帮助你更好地使用统计和定性调查的信息。本书介绍了数据管理和统计学的基本词汇，以及选择和解读通常用于分析调查数据的统计和定性方法背后的原则和逻辑。本书不能教你成为一位调查统计学家。要想成为调查统计学家，你需要进行正规的学习。如果在选择或者应用统计方法方面你

需要帮助，你必须得到统计学家的建议。如果此书能够如愿以偿，你将能够确切地告诉统计学顾问或者定性研究人员你的需要，并且你将能够解读提供给你的数据。

这些特定目标能够使你：

✓组织和管理用于分析的数据

- 起草分析计划
- 定义和格式化数据文件
- 创建编码手册
- 确定编码的可靠性
- 确认处理不完整数据或者缺失数据和异常值，以及重新编码的技术

✓确定把数据准确录入电子数据表、数据库管理程序和统计程序的方法

✓学会使用以下分析术语：

- 分布
- 临界值
- 偏度
- 转换
- 集中趋势度量
- 离散趋势
- 变异
- 统计显著性
- 实际显著性

- p 值
- α 值
- 直线性
- 曲线性
- 散点图
- 无效假设

✓列出选择合适分析方法的步骤

✓辨别名义等级、有序等级、数值等级和数据等概念以便做到：

- 分清自变量和因变量
- 正确使用均值、中位数和众数
- 正确使用极差、标准差、百分位数和四分位数间距
- 理解相关和回归的逻辑和使用
- 学会操作和解读假设检验的步骤
- 比较和对照假设检验与置信区间的使用
- 计算比值比和风险比值
- 理解卡方分布和卡方检验的逻辑和使用
- 理解 t 检验的逻辑和使用
- 理解方差的逻辑和使用
- 阅读和解释计算机输出结果

✓理解定性数据内容分析的基本步骤

✓进行开放式问题的分析，这些开放式问题询问被访者最喜欢什么和最不喜欢什么

第1章　数据管理

数据管理包括调查人员组织调查数据使之便于分析的活动。数据管理以制定分析计划为起点，以数据分析的开始为终点。数据管理行为包括：

- 起草分析计划
- 创建编码手册
- 确定可靠的编码
- 检查调查数据的缺失值
- 录入数据并且证实数据录入的准确性
- 清理数据

如果你计划做复杂的分析，你将会发现自己在不断地管理和修正更新数据，以便确信数据是干净的、完整的和适合于你要做的分析的。如果你的调查相对简短，涉及大约50人以及5个变量以下，你仍然必须确保数据的干净和完整，只是用比较少的时间管理文档。

有些调查者估计：数据管理所用时间相当于典型的数

据分析过程时间的 20% ~ 50%。对于你将用于数据管理的时间抱有现实的态度，并且确保你拥有资源（人员、时间、财力）对你来说是重要的。下面，将依次讨论每一种数据管理行为。

一、起草分析计划

分析计划描述你计划进行分析的主要调查目的、假设或研究问题。例 1.1 展示了一个分析计划样本的一部分。

例 1.1　儿童受暴力伤害调查的分析计划的一部分

• 调查目的 1：比较男孩和女孩中至少一星期一次目击和没有目击暴力事件的人数

• 假设：目击暴力事件的女孩人数比男孩多

• 变量：性别（女孩和男孩）；目击暴力行为（是或否——每星期一次或更多）

预计的分析：卡方检验，用来检验至少一星期一次目击和没有目击暴力事件的男孩和女孩的数量差异。

• 调查目的 2：比较男孩和女孩恐惧量度的得分

• 研究问题：男孩和女孩恐惧量度的得分是否有区别

• 变量：性别（女孩和男孩）；平均得分（连续从1～50）

• 预计的分析：t 检验，用来检验男孩和女孩恐惧量度平均得分的差异

调查人员通常在最终选定调查项目和确认调查方法之前草拟初步的分析计划。决定你想分析什么甚至会预示出你的调查内容。基金机构（如国家卫生研究院）和论文委员会要求你列出你的调查假设或者研究问题，描述你将如何分析数据来检验假设釉及回答问题。

不管你的分析计划如何好，抽样和数据收集过程中遇到的实际问题也许会迫使你修改你的计划。假设你是调查者，你的分析计划如例1.1中所示（调查目的2）。你决定除了检验“恐惧得分”的平均分差异，还要比较“恐惧”得分高于25分及低于24分的男孩和女孩的数量。因而你只好修改你的计划，使之既包括计划中的 t 检验又包括卡方检验。你应该预计到最初的计划不得不做修改，特别是在涉及很多数据的大规模的调查中。

二、创建编码手册

编码是计算机程序用来识别变量的单位或符号。假设

1 000个男孩和女孩完成了关于学校中暴力的调查问卷。你所关心的问题之一是有多少男孩报告在学校中经常受到威胁。要找出关于这个主题的内容，你必须使计算机程序知道要寻找哪些变量。在此案例中，变量是性别和暴力威胁。

看一下例1.2中所引用的有关学校中暴力的调查。小方框右边用来记录被访者答案的数字就是编码。你可以要求统计软件程序告诉你有多少人第1题回答1并且第8题回答4。要做到这一点，你必须告诉程序你所感兴趣的变量的名称（问题1为SEX，问题8为SCHLHURT）以及它们的值（1=男孩，4=几乎总是）。借助于许多统计程序，你还能够指定在数据中何处能找到变量。参见下面有关录入数据的部分。

例1.2 学校暴力调查摘录

1. 你是男孩还是女孩？[SEX]

   男孩 □1

   女孩 □2

8. 去年在学校里有人告诉你他们将伤害你的频率是多少？[SCHLHURT]

   从来没有 □1

   有时有 □2

很多次 □3

几乎总是 □4

在例1.2中，方括号中的SEX和SCHLHURT对应于问题所代表的变量。所有的编码手册包含对问题的描述、编码和与调查相关的变量。一个好的编码手册包含充足的信息，使得将来的研究人员能够复制这个调查、研究方法和研究成果。例1.3展示了一个编码手册的目录。

例1.3 编码手册目录

Ⅰ. 调查团队描述

这一部分包括对负责实施调查的机构及个人的特点和经验的描述。

Ⅱ. 方法

A. 抽样

1. 抽样设计（包括合格入选标准，如12岁或以下、上个月读过至少3本书）

2. 抽样方法（如分层随机抽样、方便抽样）

3. 样本量

4. 登记

5．抽样统计（包括加权和抽样误差计算）

B．研究对象：知情同意

C．研究设计——被访者如何分组；调查实施的次数及进度安排

Ⅲ．调查

A．调查的副本以及每个回答是如何编码的；与每一个问题相关的变量名

B．培训数据收集人员；质量控制

C．可靠性和有效性的信息

Ⅳ．数据文件描述

变量名（如 EDUC），变量标签（教育），变量值及变量值标签（1 = 高于 12 年级，2 = 低于 11 年级，9 = 无数据）

Ⅴ．调查手段

问题及评分方案

你可以在调查结束后对编码手册进行整理。此时，你

还应该整理用于分析的数据文件（有些人视之与编码手册等同）。参见例 1.4 数据文件和编码手册。所有的变量被分解成不连续的单位，也称为“值”，“值”对应于变量的编码。例如，在学校、邻里及其他地方受暴力威胁的频率有 4 个值：0 = 从来没有，1 = 有时，2 = 很多次，3 = 几乎天天。编码是 0、1、2、3。一个一位数的编码被用来代表没有提供信息，此例中为 9（你也可以用其他的一位数的编码）。同样，13 个值或编码代表被访者出生地的 13 个国家，两位数字的编码被用来代表缺失值（99）。通常，编码手册按照变量在调查中出现的顺序列出变量。数据文件可以照此顺序也可以不照此顺序。

尽管统计软件程序在术语上有所变化，它们都要求变量名用大写字母（如 POOL 或 EDUC）并且避免使用特殊字符，如逗号或分号。有些程序限制一个变量名中字符的个数（通常最多约为 8 个）。变量标签是变量的实际的名字（如“生活质量感知”是变量名为 POOL 的变量标签）。为了理解你的数据，统计程序需要知道你的每一个变量的变量名（如 COMM）、变量标签（社区）、变量值标签和变量值（1 = 都市，2 = 乡村）。请注意，你的程序可能会用稍微不同的术语，但是概念完全一样。

在大规模调查项目中，编码手册是项目的正式记录。它包含调查手段（如果有关，包括评分系统）；变量名、变量标签、变量值及编码；编码在数据文件中的位置；调