

全国统计教材编审委员会

“十五”规划教材

STATISTICS:
FROM DATA TO CONCLUSIONS
— STUDY GUIDE AND EXERCISES

吴喜之·主编

统计学:从数据到结论 —学习指导与习题



 中国统计出版社
China Statistics Press

全国统计教材编审委员会

“十五”规划教材

统计学：从数据到结论 ——学习指导与习题

Statistics: From Data to Conclusions
— Study Guide and Exercises

吴喜之 主 编

中国统计出版社
China Statistics Press



(京)新登字 041 号

图书在版编目(CIP)数据

统计学:从数据到结论——学习指导与习题/吴喜之主编.

—北京:中国统计出版社,2005.8

ISBN 7-5037-4780-3

I. 统…

II. 吴…

III. 统计学-高等学校-教学参考资料

IV. C8

中国版本图书馆 CIP 数据核字(2005)087065 号

统计学:从数据到结论——学习指导与习题

作 者/吴喜之

责任编辑/刘 畅、张 云、吕 军

装帧设计/艺编广告

出版发行/中国统计出版社

通信地址/北京市西城区月坛南街 75 号 邮政编码/100826

办公地址/北京市丰台区西三环南路甲 6 号

电 话/(010)63459084 63266600-22500(发行部)

印 刷/河北天普润印刷厂

经 销/新华书店

开 本/787×1092mm 1/18

字 数/250 千字

印 张/16.75

印 数/1-3000 册

版 别/2005 年 8 月第 1 版

版 次/2005 年 8 月第 1 次印刷

书 号/ISBN 7-5037-4780-3/C·2086

定 价/32.00 元

版权所有。未经许可,本书的任何部分不准以任何方式在世界任何地区以任何文字翻印、拷贝、仿制或转载。

中国统计版图书,如有印装错误,本社发行部负责调换。

出版说明

“十五”期间是我国加大教育改革力度，全面推进素质教育，教育体制、教育方法发生重大变革的时期。教材建设必须紧跟教育改革的步伐，建设适应社会主义市场经济和现代化建设需要的高质量教材。为了适应这种新形势的需要，全国统计教材编审委员会制定了《2001—2005年全国统计教材建设规划》（以下简称《规划》），并根据《规划》的要求，主要采取招标的方式组织全国有关院校的专家、学者编写了这批统计学“十五”规划教材。

这批教材力求以第三次全国教育工作会议作出的《中共中央、国务院关于深化改革全面推进素质教育的决定》为指导思想，在充分总结“九五”期间统计教材建设经验的基础上，认真贯彻大胆探索和创新的原则，努力使统计教材具有前瞻性和实用性。选题中不仅包含了一些国内统计研究和教材建设上的空白领域，也包含了统计研究的最新成果。为了配合教师教学、方便地使用这批教材，我们还特地编制了专供教师使用的电子课件，这些课件将在国家统计局统计教育中心网站([hzhttp://edu.stats.gov.cn](http://edu.stats.gov.cn))上挂出，以供需要的教师下载。另外，对于部分教材，我们还编辑出版与之相配套的习题集，以方便教师和学生在学习中使用的，也使这批教材在编辑出版上形成一个比较完整的体系。我们相信，这批教材的出版和发行，对于推动我国统计教育改革，加快我国统计教材体系和教材内容更新、改造的步伐，都将起到积极的促进作用，同时，对我国统计教材建设也将起到较好的示范与导向作用。

限于水平和经验，这批教材的编审、出版工作还会有缺点和不足，诚恳欢迎教材的使用单位、广大教师和同学们提出批评和建议。

全国统计教材编审委员会

2004年1月

前 言

这本作为《统计学：从数据到结论》的教学辅导书是许多长期教授统计学课程的老师丰富教学经验的结晶。每个老师都在其所完成的部分上体现了各自的风格。这对于各部分内容和角色很不一样的教科书来说是很自然和必要的。我确信，这本教学辅导书会使得许多读者感觉学习统计更加容易了；任课教师也可以参照本辅导书的例题和习题出考试题和编写自己的习题。

参与编写这本学习辅导书的老师们的分工如下：暨南大学经济信息管理系的王斌会负责第1-4章；西南财经大学统计学院王青华负责第5-6章；中南财经政法大学统计学系张学毅负责第7-8章；山东工商学院统计学院吴风庆和薛伟负责第9章；首都经济贸易大学统计学系刘娟负责第10-14章；对外经济贸易大学国际工商管理学院统计与数量分析学系丁岚负责第15章；山东工商学院统计学院李丽和杨海山负责第16章；中国人民大学统计学院阎洁负责第17章；中国人民大学统计学院的吴喜之负责全书的协调、组织和订正。特别要提到的是国家统计局教育中心的孙洪娟同志对本辅导书的所有例题和习题进行了逐一试算，为保证质量起了关键的作用。这本教学辅导书没有包括仅仅介绍概念的第18章。

本书的例题、习题的数据都有存于光盘中的电子版（除了少数Excel文件之外，所有数据都有文本和SPSS

形式两种)。

这本书的实现是在国家统计局教育中心的组织和指导下进行的。没有教育中心和中国经济出版社的关心、组织和帮助,没有各有关院校师生的支持,这本书绝无问世的可能。我们谨在这里向所有各方面致以衷心的感谢。

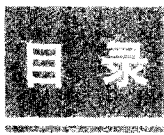
这本书还附有对于《统计学:从数据到结论》的勘误。这些勘误中有很多来自教学一线的学生和教师;他们的关心和爱护,是本书有关课程成功的关键。出错误是难免的,但出了错误不改或回避则是不负责。希望读者继续对教科书和这本教学指导书提出宝贵的意见,使得我们在再版时能做得尽量让读者满意。

吴喜之

2004年6月

前言

第一章 一些基本概念	1
一、学习指导	1
二、例题与分析	2
三、练习	7
四、参考答案	8
第二章 数据的收集	9
一、学习指导	9
二、例题与分析	10
三、练习	18
四、参考答案	19
第三章 数据的描述	20
一、学习指导	20
二、例题与分析	22
三、练习	33
四、参考答案	34
第四章 机会的度量:概率和分布	39
一、学习指导	39
二、例题与分析	40
三、练习	43
四、参考答案	44
第五章 简单统计推断:总体参数的估计	48
一、学习指导	48
二、例题与分析	55
三、练习	58
四、参考答案	62



第六章 简单统计推断:总体参数的假设检验	66
一、学习指导	66
二、例题与分析	75
三、练习	81
四、参考答案	86
第七章 相关分析和回归分析	93
一、学习指导	93
二、例题与分析	95
三、习题	99
四、参考答案	99
第八章 列联表、χ^2 检验和对数线性模型	102
一、学习指导	102
二、例题与分析	103
三、练习	107
四、参考答案	109
第九章 方差分析	112
一、学习指导	112
二、例题与分析	113
三、练习	120
四、参考答案	125
第十章 寻找多个变量的代表	
——主成分分析和因子分析	132
一、学习指导	132
二、例题与分析	133
三、练习	140
四、参考答案	140



第十一章 把对象分类:聚类分析	144
一、学习指导	144
二、例题与分析	145
三、练习	152
四、参考答案	152
第十二章 把对象归到已知的类中:判别分析	156
一、学习指导	156
二、例题与分析	157
三、练习	164
四、参考答案	164
第十三章 两组变量之间的相关:典型相关分析	170
一、学习指导	170
二、例题与分析	171
三、练习	178
四、参考答案	178
第十四章 行变量和列变量的关系:对应分析	180
一、学习指导	180
二、例题与分析	181
三、练习	185
四、参考答案	187
第十五章 时间序列分析	191
一、学习指导	191
二、例题与分析	192
三、例题与分析(季节性 ARZMA 模型)	200
四、练习	210
五、参考答案	210



第十六章 总体分布未知时的检验:非参数检验方法	217
一、学习指导	217
二、例题与分析	219
三、练习	225
四、参考答案	231
第十七章 生存分析简介	234
一、学习指导	234
二、例题与分析	235
三、练习	252
四、参考答案	252
《统计学:从数据到结论》勘误	256

第一章

一些基本概念

一、学习指导

1. 学习目的和要求

了解统计的主要思想,掌握统计的基本概念,明确统计的研究对象。

2. 内容提要

随机性和规律性是统计的两个重要概念。从随机性中寻找规律性是统计的主要思想。统计研究的对象是变量,变量分为定性变量和定量变量。利用收集到的数据对变量之间的关系进行分析是统计分析的主要内容。统计软件是进行数量分析的重要工具。

3. 重点难点

注意区分定性变量和定量变量;如何描述变量之间可能存在的关系。

SPSS 的使用;SPSS 作图,定性变量在 SPSS 如何定义,定性变量在 SPSS 如何编码,定性变量的列联表分析。

二、例题与分析

例 1.1 定量变量间的关系

1. 研究背景与目的

利税与工业总产值关系分析:为了了解某市工业总产值与利税总额的关系,以便能从工业总产值去预测下年度的税收总额,今收集了 1988 - 1999 年 12 年间的的数据如表 1.1。

表 1.1 1988 - 1999 年 12 年间某市工业总产值与利税总额 (单位:万元)

年份	工业总产值 x	利税总额 y
1988	401.92	62.80
1989	854.08	226.08
1990	1105.28	180.86
1991	1130.40	213.52
1992	1708.16	580.27
1993	1783.52	439.60
1994	2009.60	504.91
1995	2361.28	575.25
1996	3416.32	660.66
1997	3642.40	720.94
1998	4647.20	766.16
1999	5149.60	947.02

2. 数据说明

数据来源为《2000 年中国统计年鉴》;文件名:工业总产值与利税 .sav
变量名称: x :工业总产值, y :利税总额。

3. 通过 SPSS 软件来分析研究该问题

(1) 输入数据:将数据输成 12 行 3 列,分别记每列变量名为 t 、 x 、 y 。

(2) 绘制线图,考察该市工业总产值与利税总额随时间变化的趋势。

1) 选方法:选择菜单 Graphs → Line..., 打开 Line 对话框,选择“Values of individual cases”,单击 Multiple 图形,按 **Define** 键,打开 Multiple Lines 子对话框;

2) 定变量:将变量 x 和 y 放入 Lines Represent 框;变量 t 放入 Category Labels: Variable 框。

3)得结果:按 按钮执行,结果见图 1.1

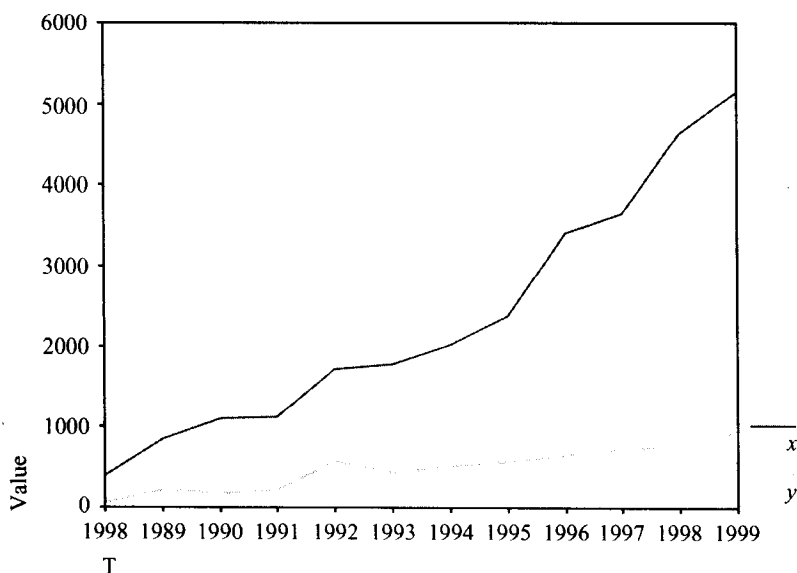


图 1.1 工业总产值 x 与利税总额 y 的曲线图

从图中可以看到,随着时间增加,工业总产值比利税总额增加快。

例 1.2 定性和定量变量间的混合关系

1. 研究背景与目的

为了了解某工厂工人的一些基本情况,现从有 2000 个工人的工厂中随机抽取 20 名工人,其资料如下:(抽样方法见例 2.1)

工人编号	性别	年龄	文化程度	技术级别
1	男	52	文盲	6
2	男	30	初中	3
3	男	19	初中	2
4	男	46	高中	4
5	女	47	小学	4
6	男	34	小学	2
7	女	22	初中	8
8	男	31	高中	5
9	男	55	初中	8

续表

工人编号	性别	年龄	文化程度	技术级别
10	男	32	初中	5
11	女	49	中专	4
12	男	34	初中	4
13	男	34	初中	4
14	男	61	技工	7
15	男	36	初中	4
16	女	40	技工	5
17	女	30	中专	4
18	男	45	中专	5
19	女	40	文盲	4
20	女	35	高中	5

在本例中,性别为定性变量,年龄为定量数据,而文化程度和技术级别可看作有序(等级)定性变量。

2. 数据说明

数据来源为“《统计学原理》学习指导”-韩兆洲、王斌会主编;

文件名:工人资料 .sav

3. SPSS 分析

在用 SPSS 等统计分析软件进行分析以前,通常要将定性或有序数据用哑元表示,例如对本例的性别和文化程度进行转化:

性别:男-1,女-2

文化程度:文盲-1,小学-2,初中-3,高中-4,中专-5,技工-6。下表是 SPSS 输入界面:(见下页)

1) 双因素分析——列联表分析法:文化程度与性别等的关系

选择菜单 Analyze→Descriptive Statistics, 打开 CrossTabs 对话框,将变量“性别”放入 Row(s)框,将变量“文化程度”放入 Column(s)框,选择 Display Clustered Bar Charts 选项,按 按钮执行,结果如下:

性别 * 文化程度 Crosstabulation

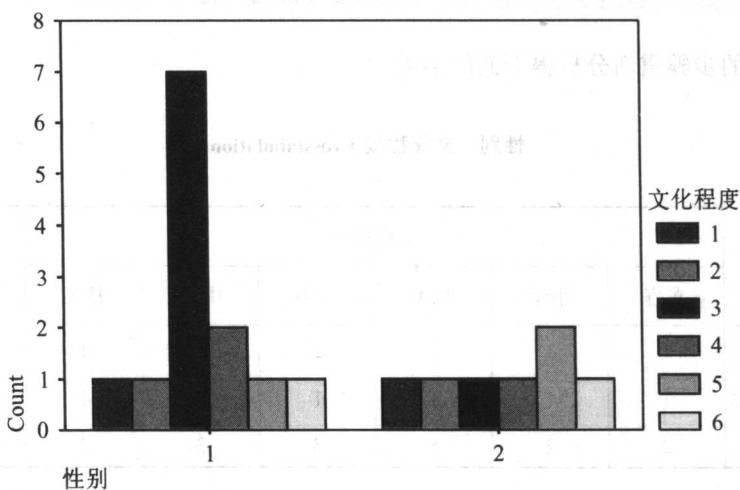
		文化程度						Total
		1	2	3	4	5	6	
性别	1	1	1	7	2	1	1	13
	2	1	1	1	1	2	1	7
Total		2	2	8	3	3	2	20

工人资料 - SPSS Data Editor

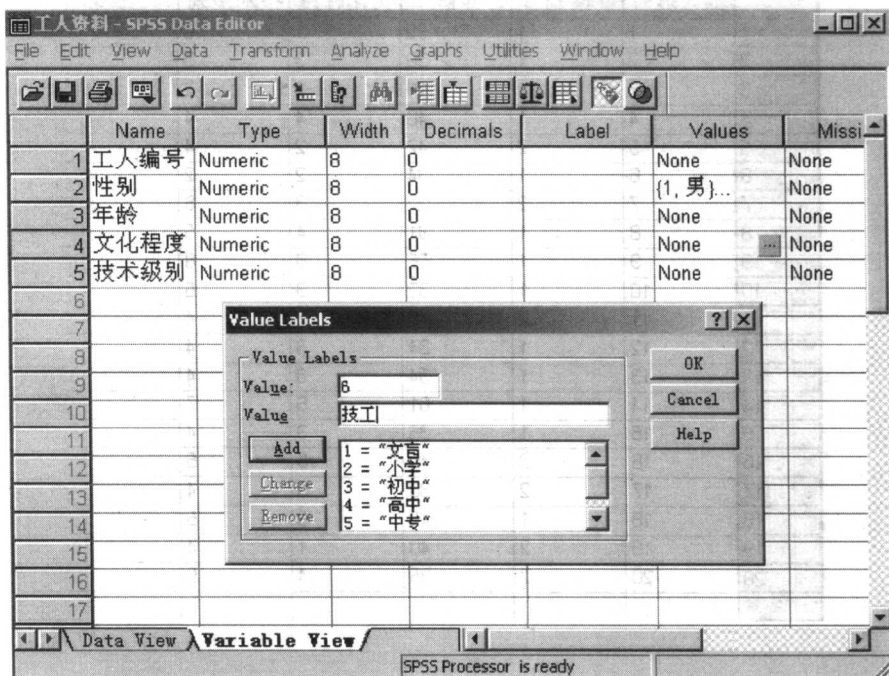
File Edit View Data Transform Analyze Graphs Utilities Window Help

20 : 工人编号 20

	工人编号	性别	年龄	文化程度	技术级别	var
1	1	1	52	1	6	
2	2	1	30	3	3	
3	3	1	19	3	2	
4	4	1	46	4	4	
5	5	2	47	2	4	
6	6	1	34	2	2	
7	7	2	22	3	8	
8	8	1	31	4	5	
9	9	1	55	3	8	
10	10	1	32	3	5	
11	11	2	49	5	4	
12	12	1	34	3	4	
13	13	1	34	3	4	
14	14	1	61	6	7	
15	15	1	36	3	4	
16	16	2	40	6	5	
17	17	2	30	5	4	
18	18	1	45	5	5	
19	19	2	40	1	4	
20	20	2	35	4	5	
21						



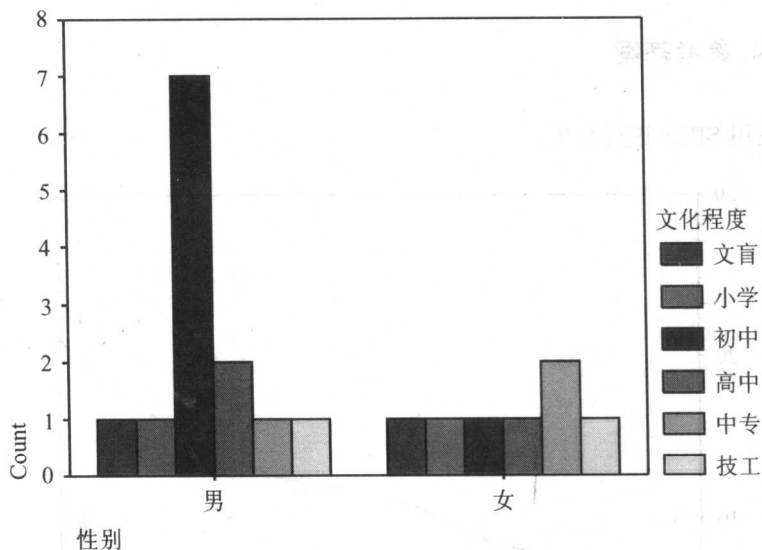
但这样编码也有一个问题，即当分类太多或数据很大时，很难区分编码的含义，例如上表和图中就很难看出文化程度编码的含义，SPSS 给出了确定这些编码含义的方法，即对其赋值，见下图：



按上面的步骤重新分析得下面的结果：

性别 * 文化程度 Crosstabulation

Count		文化程度						Total
		文盲	小学	初中	高中	中专	技工	
性别	男	1	1	7	2	1	1	13
	女	1	1	1	1	2	1	7
Total		2	2	8	3	3	2	20



三、练习

为了了解全国人均消费额(y)与人均国民收入(x)的关系,今调查了1984-1993年10个年份的数据,见表1.2,试分析它们之间的关系(数据说明:数据来源为《1994年中国统计年鉴》;文件名:收入与消费.sav)。

表 1.2 1984-1993年全国人均国民收入与人均消费额 (单位:百元)

年份 t	人均国民收入 x	人均消费金额 y
1984	5.44	3.29
1985	6.68	4.06
1986	7.38	4.51
1987	8.60	5.13
1988	10.69	6.43
1989	11.69	6.99
1990	12.51	7.13
1991	14.30	8.03
1992	17.26	9.47
1993	21.00	11.48