

供 中医学类、中药学类、药学类 等专业用

主编

蔡

晶

魏

高文

医学统计学实战指导

人民卫生出版社
PEOPLE'S MEDICAL PUBLISHING HOUSE

供中医学类、中药学类、药学类专业用

医学统计学实战指导

主 审 何 雁

主 编 蔡 晶 魏高文

副主编 李国春 武 松

编 委 (按姓氏笔画排序)

王丽梅	孔丽娅	刘志臻	许 茜	孙春阳
李国春	杨雪梅	步怀恩	吴建军	陈梅妹
陈婷婷	武 松	林有志	俞向梅	徐 刚
黄品贤	葛 亮	韩 梅	谢海林	赖新梅
蔡 晶	魏高文	魏歆然		

学术秘书 许 茜 葛 亮

人民卫生出版社

图书在版编目(CIP)数据

医学统计学实战指导/蔡晶,魏高文主编.—北京:
人民卫生出版社,2016

ISBN 978-7-117-23030-8

I. ①医… II. ①蔡…②魏… III. ①医学统计-统
计学-医学院校-教材 IV. ①R195.1

中国版本图书馆 CIP 数据核字(2016)第 184882 号

人卫智网	www.ipmph.com	医学教育、学术、考试、健康, 购书智慧智能综合服务平台
人卫官网	www.pmph.com	人卫官方资讯发布平台

版权所有,侵权必究!

医学统计学实战指导

主 编:蔡 晶 魏高文

出版发行:人民卫生出版社(中继线 010-59780011)

地 址:北京市朝阳区潘家园南里 19 号

邮 编:100021

E-mail: pmph @ pmph.com

购书热线:010-59787592 010-59787584 010-65264830

印 刷:北京机工印刷厂

经 销:新华书店

开 本:787×1092 1/16 印张:8

字 数:190 千字

版 次:2016 年 9 月第 1 版 2016 年 9 月第 1 版第 1 次印刷

标准书号:ISBN 978-7-117-23030-8/R · 23031

定 价:23.00 元

打击盗版举报电话:010-59787491 E-mail: WQ @ pmph.com

(凡属印装质量问题请与本社市场营销中心联系退换)

前言

统计学是一门古老而又新兴的科学与艺术。随着社会经济的发展与科学技术的进步,统计理论与统计技术的深度和广度也在与时俱进。21 世纪以来,电子计算机技术的飞速发展令大数据挖掘时代扑面而来,使统计数据的搜集、整理、分析、存贮、传递等过程逐步电子化,极大地提高了统计工作的效能,扩大了统计技术的应用领域,使得以往高高在上的统计学日益普及,成为普通研究者都能自我实现的“简单过程”。

对于非统计学专业的学习者来说,统计学繁杂的计算公式往往令人望而却步,而统计软件的使用虽然看似简单,但其中依然存在许多技巧,需要用心学习与勤于练习才能掌握。我们身为一线教师,在长期的医学统计学教学实践中,深深体会到统计学理论教学和统计软件应用实践的重要性,二者不可偏废。但纵观大量的统计教材和统计软件教材,大多数是厚厚的一大本,从理论到应用,虽然洋洋洒洒几百页,内容丰富,但实战技巧却点拨不够到位,难免压抑初学者的学习信心,不利于统计技术的普及和推广。因此,我们着手编写了这本医学统计学实战指导,按照统计思维的步骤介绍相关的统计分析技术,突出简单实用的特点,大大降低学习的难度,使读者能以较短的时间、比较轻松地掌握常用的统计分析技术。在统计软件的选择上,我们选用了业界认可且界面友好、相对简单易学的 SPSS 软件。如研究工作需要,可在此基础上进一步学习 SAS、STAT、MINITAB、JMP、S-PLUS 等统计软件。

本书包括基础统计方法选择、计量资料的统计描述、均数比较、方差分析、 χ^2 检验、非参数检验、直线相关与回归共 7 章。各章均结合中医药研究的实例,介绍各种统计分析方法在 SPSS 统计软件中的实现步骤和结果解读技巧。每章均附有练习题,供学习者参照练习。本书可作为学习医学统计学的各专业本科生、研究生的实验指导,也可供广大医务工作者总结临床经验、论文写作及科研工作时参考。

本书是全国各中医院校统计学教师集体智慧的结晶,其编写得到了各高等中医药院校和人民卫生出版社的关怀和鼎力支持。在此一并表示深深的感谢!由于编者学识有限,难免存在不足和纰漏,恳请广大读者批评指正。

编者

2016 年 2 月

目 录

第一章 基础统计方法选择	1
第二章 计量资料的统计描述	3
第三章 均数比较	10
第一节 样本均数与总体均数比较的 t 检验	10
第二节 配对设计资料比较的 t 检验	15
第三节 完全随机设计两样本均数比较的 t 检验	23
第四章 方差分析	27
第一节 完全随机设计多样本单因素方差分析	27
第二节 随机区组设计资料的双因素方差分析	33
第三节 重复测量计量资料方差分析	45
第五章 χ^2 检验	59
第一节 四格表 χ^2 检验	59
第二节 行列表 χ^2 检验	66
第三节 配对 χ^2 检验	70
第六章 非参数检验	76
第一节 单样本秩和检验	76
第二节 配对样本秩和检验	81
第三节 两样本秩和检验	86
第四节 多样本秩和检验	94
第七章 直线相关与回归	103
第一节 直线相关	103
第二节 直线回归	112

第一章 基础统计方法选择

医学统计学是运用概率论与数理统计的原理及方法,结合医学实际,研究数字资料的收集、整理、分析与推断的一门学科,在医学科研中的应用非常广泛,是医学各相关专业、各层次学生的一门必修课程。但对于医学生而言,医学统计学不是数学,是提高医学科研能力的必需工具,一定要以理解为主,医学统计学的思维贯穿科研的整个过程。

1. 先确定研究目的,根据研究目的选择方法。不同研究目的采用的统计方法不同,常见的研究目的主要有三类:一是差异性研究,即比较组间均数、率等的差异,可用的方法有 t 检验、方差分析、 χ^2 检验、非参数检验等;二是相关性分析,即分析两个或多个变量之间的关系,可用的方法有相关分析;三是影响性分析,即分析某一结局发生的影响因素,可用的方法有线性回归、logistic 回归、Cox 回归等。

2. 明确数据类型,根据数据类型进一步确定方法。不同数据类型采用的统计方法也不同。计量资料可用的方法有 t 检验、方差分析、非参数检验、线性相关、线性回归等。分类资料可用的方法有 χ^2 检验、对数线性模型、logistic 回归等。图 1-1 ~ 图 1-4 简要列出了不同研究目的、不同数据类型常用的统计分析方法。

3. 选定统计方法后,需要利用统计软件具体实现统计分析过程。SPSS 中,不同的统计方法对应不同的命令,只要方法选定,便可通过对应的命令辅之以相应的选项实现统计结果的输出。

4. 统计结果的输出并非数据分析的完成。一般统计软件都会输出很多结果,需要从中选择自己需要的部分,并做出统计学结论。但统计学结论不同于专业结论,最终还需要

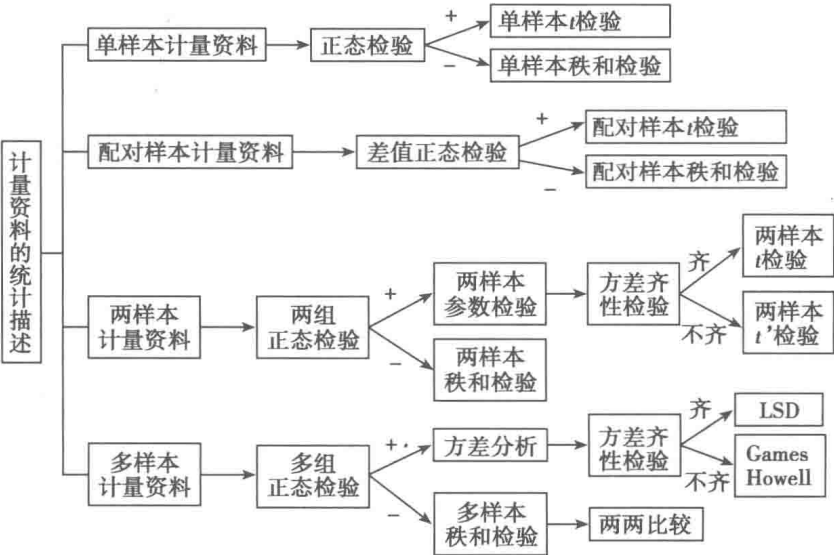


图 1-1 计量资料的统计描述

结合实际做出合理解释。

以下分别是计量资料的统计描述(图 1-1)、双变量相关的统计描述(图 1-2)、计数资料的统计描述(图 1-3)以及等级资料的统计描述(图 1-4)。

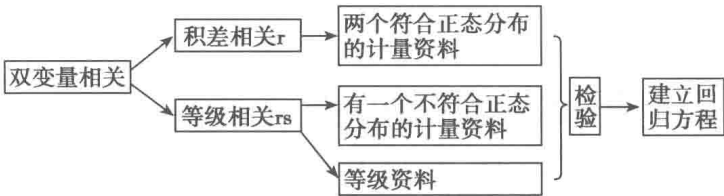


图 1-2 双变量相关的统计描述

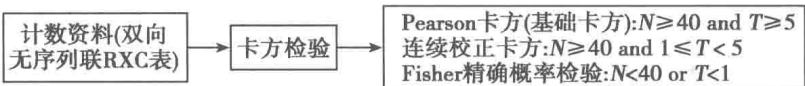


图 1-3 计数资料的统计描述



图 1-4 等级资料的统计描述

第二章 计量资料的统计描述

计量资料的统计描述包括集中趋势、离散趋势和分布形态。符合正态分布的计量资料的统计描述可用算术平均数和标准差表示,不符合正态分布的计量资料的统计描述则运用中位数和四分位数间距表示。

例 2-1 抽样调查某地 120 名 18~35 岁健康男性居民血清铁含量($\mu\text{mol/L}$),数据如下,试编制血清铁含量的频数分布表,计算算术均数、中位数、众数、极差、四分位数间距、方差、标准差、偏度系数、峰度系数,绘制频数分布图及正态曲线。

19.12	15.83	19.26	14.37	9.97	23.90	25.61	17.55	24.22	12.67
16.72	19.89	17.19	27.81	19.82	17.51	23.29	19.50	20.18	23.77
18.61	12.65	19.83	19.47	17.48	19.09	18.37	19.69	22.55	7.42
12.50	20.40	23.02	21.62	16.10	17.14	17.67	18.36	24.66	15.31
18.12	19.08	19.32	20.13	17.45	11.74	24.75	21.00	18.48	23.07
22.73	14.94	23.02	15.94	27.90	21.61	13.12	20.75	15.51	17.40
17.25	19.59	18.54	16.79	21.69	26.02	16.52	14.77	26.13	16.32
8.40	13.17	16.99	21.53	11.09	21.46	19.38	17.98	29.64	20.30
20.87	12.73	24.13	14.27	11.75	18.89	17.08	21.31	24.36	21.65
14.89	17.32	17.40	20.52	24.52	8.65	19.53	23.12	18.89	18.46

SPSS 操作步骤:

1. 数据录入 在“变量视图”以“铁含量”为变量名,切换到“数据视图”输入铁含量的 120 个数据,建立 120 行(120 个观测对象)1 列(1 个变量名)的数据文件,保存名称为铁含量,见图 2-1。

在图 2-1“度量标准”下的单元格中,单击定位后,单元格右侧出现按钮,点击该按钮,在下拉列表框中选择度量类型。

度量:定量变量,如身高、体重、血压等测量值。

序号:等级变量(半定量,有序分类),如疗效记录:痊愈、好转、稳定、无效。

名义:定性变量或字符串变量,如血型记录为 A 型、B 型、AB 型、O 型;姓名为陈某。

变量类型为数值时,度量类型默认值为度量,变量类型为字符串时,其默认值为名义。

2. 正态性检验 ①菜单行点击“分析→描述统计→探索”,将铁含量送入因变量列表→在“绘制”栏中带检验的正态图→继续→在“输出”栏中选择“二者都”(默认)→确定,见图 2-2;②输出结果:如图 2-3,当样本含量 ≥ 50 时,选择 Kolmogorov-Smirnov 统计量,当样本含量 < 50 时,选择 Shapiro-Wilk 统计量。本例样本例数 ≥ 50 ,选择 K-S 结果, $P=0.200$ 。

答: H_0 :铁含量数值服从正态分布。

H_1 :铁含量数值不服从正态分布。



图 2-1 定义及输入数据



正态性检验

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	统计量	df	Sig	统计量	df	Sig
铁含量	.054	120	.200 [*]	.988	120	.348

a. Lilliefors 显著水平修正

*. 这是真实显著水平的下限

图 2-3 正态性检验结果

 $\alpha=0.05$ (双侧建议)
$$Z = 0.054, P = 0.200 > 0.05$$

结论:在 $\alpha=0.05$ 的检验水平,不拒绝 H_0 ,可认为铁含量服从正态分布。

3. 频数分析 ①菜单行点击“分析→描述统计→频率”,将铁含量送入“变量”框,见图 2-4;②在“统计量”的对话框下选择四分位数、均值、中位数、众数、标准差、最大值、最小值、方差、范围、偏度、峰度→点击“继续”,见图 2-5;③在“图表”的对话框下选择“直方图”,在直方图上显示正态曲线(S)→点击“继续”,见图 2-6;④回到“频率”对话框,点击“确定”。

4. 结果解读 见表 2-1、图 2-7。



图 2-4 频数分析命令

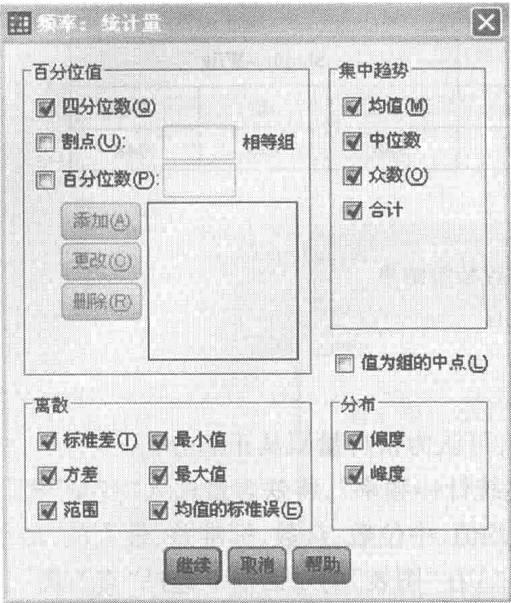


图 2-5 统计量的选择

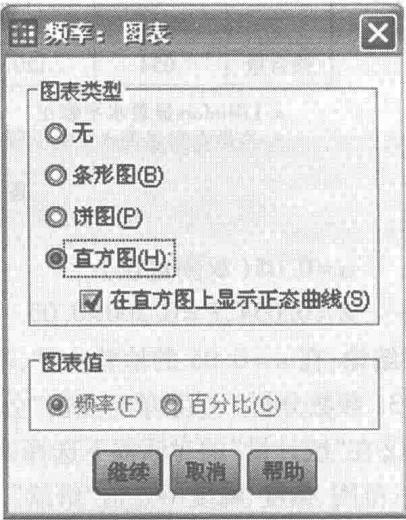


图 2-6 频率-图表的选择

表 2-1 统计量

铁含量			
有效	120	峰度	-.243
缺失	0	峰度的标准误	.438
均值	18.2845	全距	21.24
均值的标准误	.41368	极小值	8.40
中值	18.5400	极大值	29.64
众数	18.89	和	2194.14
标准差	4.53164	百分位数	
方差	20.536	25	14.8900
偏度	-.109	50	18.5400
偏度的标准误	.221	75	21.5125

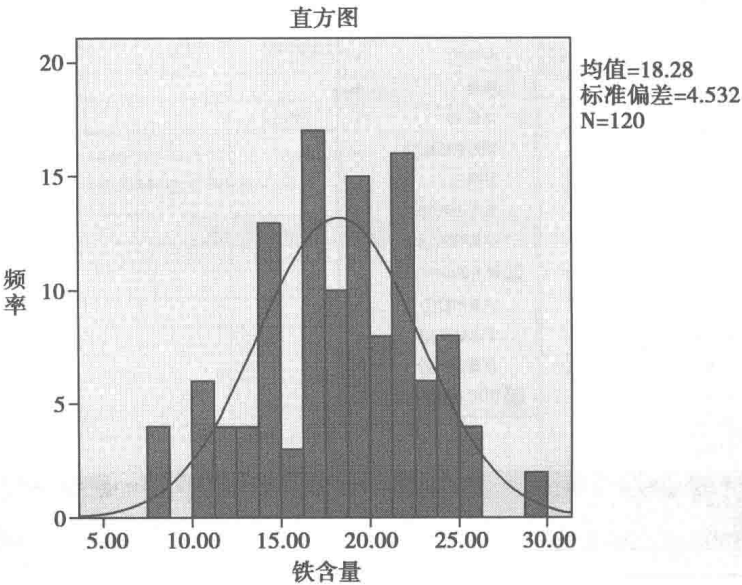


图 2-7 直方图及正态曲线



如表 2-1,统计描述:有效数 120,算术均数 18.28、中位数 18.54、众数 18.89、极差 21.24、四分位数间距 6.6225、方差 20.54、标准差 4.53、偏度系数 -0.109、峰度系数 -0.243。

练习题

1. 某地用随机抽样方法检查了 140 名正常成年男子的红细胞数($\times 10^{12}/L$),检测结果如下,试编制红细胞数的频数分布表,计算算术均数、中位数、众数、极差、四分位数间距、方差、标准差、偏度系数、峰度系数,绘制频数分布图及正态曲线。

4.76 5.26 5.61 5.95 4.46 4.57 4.31 5.18 4.92 4.27 4.77 4.88
 5.00 4.73 4.47 5.34 4.70 4.81 4.93 5.04 4.40 5.27 4.63 5.50
 5.24 4.97 4.71 4.44 4.94 5.05 4.78 4.52 4.63 5.51 5.24 4.98
 4.33 4.83 4.56 5.44 4.79 4.91 4.26 4.38 4.87 4.99 5.60 4.46
 4.95 5.07 4.80 5.30 4.65 4.77 4.50 5.37 5.49 5.22 4.58 5.07
 4.81 4.54 3.82 4.01 4.89 4.62 5.12 4.85 4.59 5.08 4.82 4.93
 5.05 4.40 4.14 5.01 4.37 5.24 4.60 4.71 4.82 4.94 5.05 4.79
 4.52 4.64 4.37 4.87 4.60 4.72 4.83 5.33 4.68 4.80 4.15 4.65
 4.76 4.88 4.61 3.97 4.08 4.58 4.31 4.05 4.16 5.04 5.15 4.50
 4.62 4.73 4.47 4.58 4.70 4.81 4.55 4.28 4.78 4.51 4.63 4.36
 4.48 4.59 5.09 5.20 5.32 5.05 4.41 4.52 4.64 4.75 4.49 4.22
 4.71 5.21 4.94 4.68 5.17 4.94 5.02 4.76

答:

2. 某地 100 例 30 ~40 岁健康男子血清总胆固醇值 (mg/dl) 测定结果如下, 试编制血清总胆固醇的频数分布表, 计算算术均数、中位数、众数、极差、四分位数间距、方差、标准差、偏度系数、峰度系数, 绘制频数分布图及正态曲线。

202 165 199 234 200 213 155 168 189 170 188 168 184 147 219 174 130 183 178
174 228 156 171 199 185 195 230 232 191 210 195 165 178 172 124 150 211 177
184 149 159 149 160 142 210 142 185 146 223 176 241 164 197 174 172 189 174
173 205 224 221 184 177 161 192 181 175 178 172 136 222 113 161 131 170 138
248 153 165 182 234 161 169 221 147 209 207 164 147 210 182 183 206 209 201
149 174 253 252 156

答:

3. 2014 年某市 100 名孕妇产前检查次数资料如下, 试编制产前检查次数的频率分布表, 检查次数请按 0、1、2、3、4、5、>5 编制。

5 3 5 7 3 3 2 4 2 7 5 5 2 5 1 4 4 4 6 5
2 1 5 5 0 5 4 3 4 5 2 5 12 5 5 4 0 4 5 4
1 3 4 3 6 2 4 1 4 4 5 4 4 2 3 4 1 4 4 3
2 5 5 1 5 5 4 4 5 4 4 7 4 4 5 3 7 8 3 3
6 5 1 3 2 6 5 3 0 4 6 5 2 3 6 4 3 3 4 6

答:



4. 试应用频率表法计算中位数、算术均数、四分位数间距、方差、标准差。

组段	频数
12 ~	1
24 ~	7
36 ~	11
48 ~	11
60 ~	7
72 ~	5
84 ~	4
96 ~	2
108 ~	2

答:

5. 某地 200 例正常成人血铅含量($\mu\text{mol/L}$)的频数分布如下。

(1) 简述该资料的分布特征。

(2) 若资料近似呈对数正态分布,试分别用百分位数法和正态分布法估计该地正常成人血铅值的 95% 参考值范围。

血铅含量	频数
0.00 ~	7
0.24 ~	49
0.48 ~	45
0.72 ~	32
0.96 ~	28
1.20 ~	13
1.44 ~	14
1.68 ~	4
1.92 ~	4
2.16 ~	1
2.40 ~	2
2.64 ~	1

答:

第三章 均数比较

均数比较是计量资料常用的统计分析方法。按资料的设计类型可分为单样本均数与总体均数比较、配对设计资料均数比较、完全随机设计两独立样本均数比较、完全随机设计多样本均数比较、随机区组设计资料均数比较、重复测量计量资料均数比较等；分别对应于样本均数与总体均数比较的 t 检验或秩和检验、配对设计资料均数比较的 t 检验或秩和检验、完全随机设计两独立样本均数比较的 t 检验或秩和检验、完全随机设计多样本均数比较的方差分析或秩和检验、随机区组设计资料均数比较的方差分析或秩和检验、重复测量计量资料均数比较的方差分析或秩和检验。在具体统计分析方法的选择上则还需要根据数据是否符合正态分布选择；符合正态分布的选择 t 检验、方差分析，不符合正态分布的选择秩和检验。其中 t 检验在本章进行介绍，方差分析见第四章，秩和检验见第六章。

第一节 样本均数与总体均数比较的 t 检验

例 3-1 已知某地成年男子，通过尿酸酶法测得血尿酸均数为 $400\mu\text{mol/L}$ ，现在该地邻近海边区域随机调查 20 名健康成年男子，测得其血尿酸值分别为 (410, 415, 422, 418, 419, 420, 425, 430, 411, 415, 428, 433, 427, 411, 421, 429, 435, 424, 427, 440)，请根据此判断海边区域成年男子的血尿酸是否与该地成年男子有所不同？

SPSS 操作步骤：

- 1. 数据录入 ①点击“变量视图”界面进入变量定义，在“名称”列下输入“血尿酸”一个变量。②在“组别”变量行中点击“值”，默认为“无”；“值标签”默认为“无”，见图 3-1。③点击“数据视图”，输入全部的数据，见图 3-2。

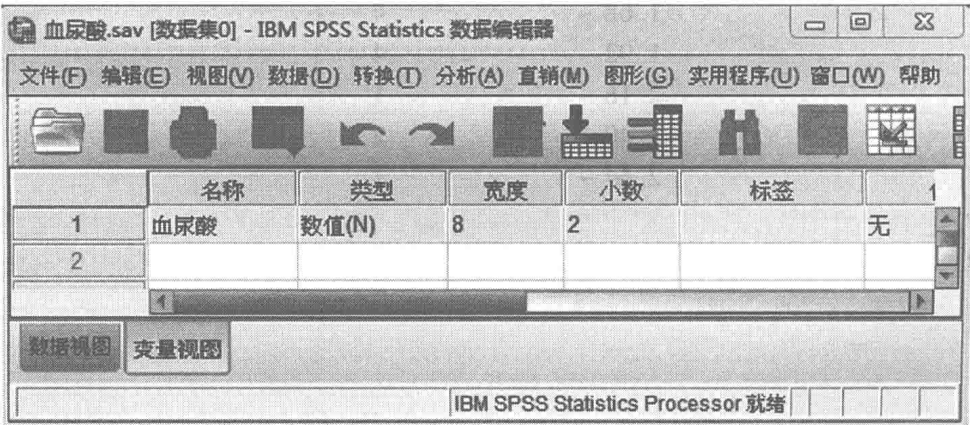
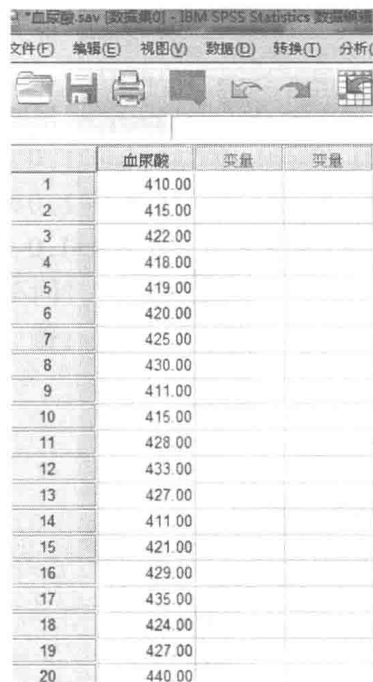


图 3-1 定义变量



	血尿酸	变量	变量
1	410.00		
2	415.00		
3	422.00		
4	418.00		
5	419.00		
6	420.00		
7	425.00		
8	430.00		
9	411.00		
10	415.00		
11	428.00		
12	433.00		
13	427.00		
14	411.00		
15	421.00		
16	429.00		
17	435.00		
18	424.00		
19	427.00		
20	440.00		

图 3-2 数据录入

2. 统计分析 ①在“数据视图”中,点击“分析”,选择“比较平均值”→“单样本 T 检验”,见图 3-3。②将左侧框内变量“血尿酸”,通过选择箭头选入检验变量(T)框内,在检验值(V)框内输入该地区血尿酸平均值“400”,点击“确定”,见图 3-4。

3. 运行结果 见图 3-5。

答:根据正态性检验(步骤略),结论: $P>0.05$,符合正态分布,因此采用单样本 t 检验。

H_0 :海边区域成年男子的血尿酸与该地成年男子相同。

H_1 :海边区域成年男子的血尿酸与该地成年男子不同。

$\alpha=0.05$ (双侧检验)

$t=12.340$ $df=19$ P 显示为 $0.000<\alpha$

结论:在 $\alpha=0.05$ 的检验水平,拒绝 H_0 ,接受 H_1 ,差异具有统计学意义,可认为海边区域成年男子的血尿酸与该地成年男子有差别。



图 3-3 样本分析方法选择

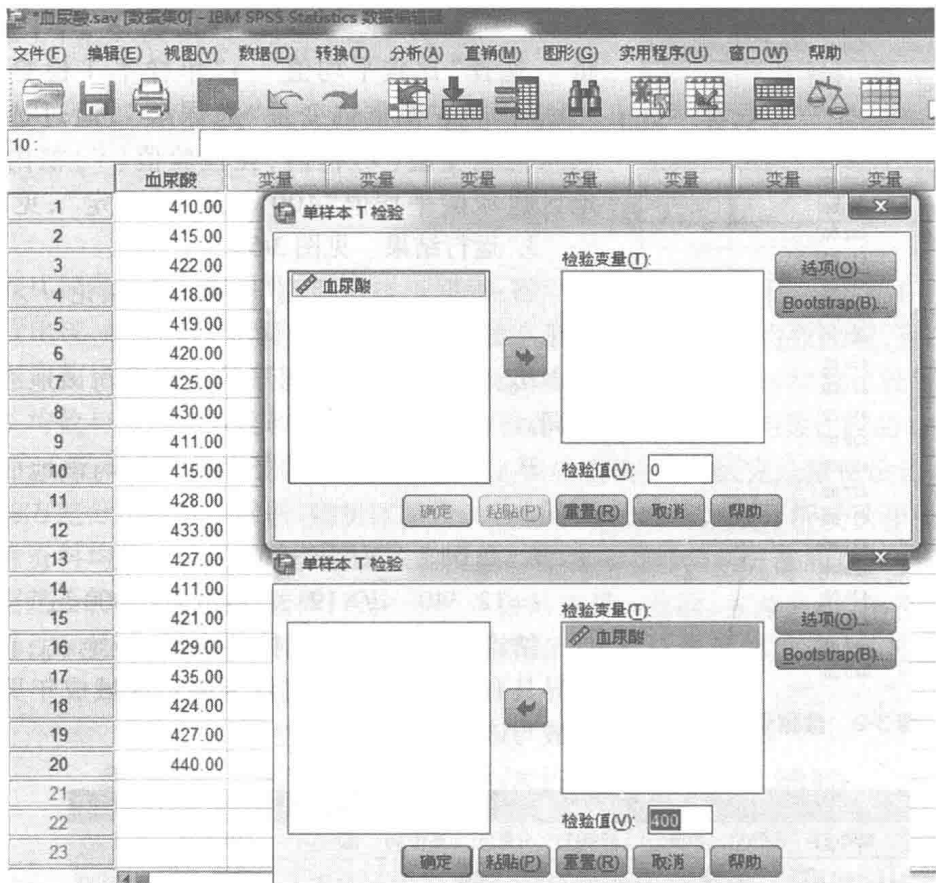


图 3-4 检验变量选择和检验值输入

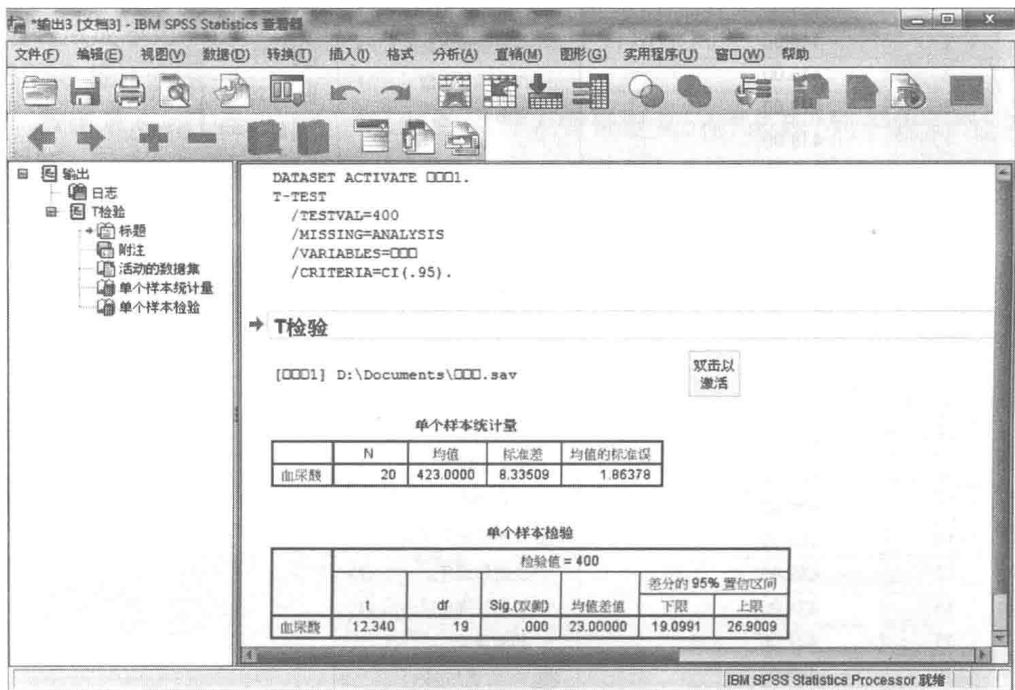


图 3-5 检验结果