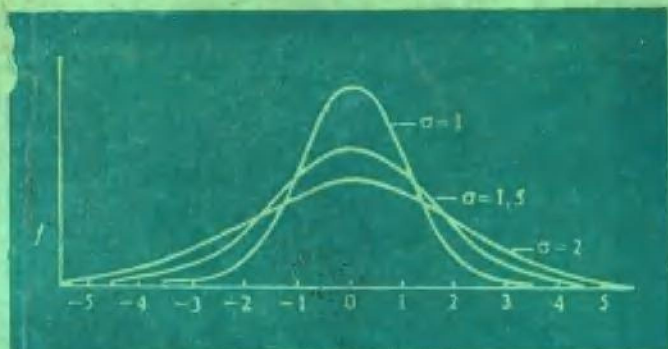




YIYONG TONGJI
FANG FA

医用统计方法

金丕焕 主编

上海医科大学出版社

8/K321.2

医 用 统 计 方 法

主编 金丕焕
副主编 詹绍康 张照寰

上海医科大学出版社

(沪)新登字 207 号

责任编辑 倪如晶

封面设计 吴 平

医用统计方法

主编 金丕焕

上海医科大学出版社出版发行

上海市医学院路 138 号

邮政编码 200032

新华书店上海发行所经销

商务印书馆上海印刷厂印刷

开本 787×1092 1/16 印张 34.75 插页 2 字数 845 千字

1993 年 2 月第 1 版 1993 年 2 月第 1 次印刷

印数 1—6500

ISBN 7-5627-0071-0/R·62

定价: 25.00 元

前 言

由上海医科大学卫生统计学与社会医学教研室根据历年在教学、科研中应用的统计方法而编写并于1979年出版的《医学统计方法》一书,深受广大读者欢迎,曾多次重印,至今已逾13年。近些年来,统计理论与方法发展十分迅速,更有效、更实用的新方法不断出现,多变量分析方法的应用日益普及,原书已不能满足广大读者的要求,迫切需要修改和补充。为此,我们通过两年多时间大量的工作,对内容作了大幅度的增补和修改,编写了本书。

本书主要是面向医学工作者的,所以仍保持浅显、易懂和实用的原则。即使是多变量分析的有关章节,也尽量避免较深的数学内容。如实属必要,也只是对复杂的数学运算作一些通俗的说明,以免读者为之困惑。多变量分析一般需由计算机来实现。本书重点在于讲清各种统计方法的意义、原理、计算原则、适用条件及结果的解释等。如果读者需要应用这些统计方法,可参阅本书的姐妹篇——《医学统计程序集》。它包括了本书中介绍的绝大部分统计方法的计算机程序,并备有可用于多种计算机的相应软盘和磁带,使用十分方便。

试验设计对科学研究的成败至关重要,本书加强了这一部分内容。鉴于计算机技术已相当普及,多变量分析广泛应用,因此增加了主成分分析、因子分析和聚类分析三章;考虑到流行病学研究中有关统计方法不仅用于传染病流行情况的研究,而且频繁地用于各种非传染病的研究工作,这里不仅加入常用的Mantel-Haenzel分析和相对危险度分析等内容,而且还编写了logistic回归分析和比例风险模型这两章。为适应对随访资料作分析的需要,在生存率统计一章中进行了很大的扩展,以实例讲解了小样本资料分析、时序检验(logrank test)、Gehan比检验和Kaplan-Meier分析等方法。由于对应变量作变换后用线性最小二乘法作曲线拟合的方法在理论上有些缺陷,这部分内容也有较大的变动。此外,还增加了圆分布资料分析和非参数统计分析中的多重比较方法、等级相关分析及柯尔莫哥洛夫—斯米尔诺夫检验、Bayes定理的应用和正交试验设计方差分析等章节。

本书的编写者在各章之末都已写明。此外,钱大钧老师参加了把旧式计量单位转换为法定计量单位等的计算机程序编写及运算核对工作;朱颖亚老师完成了汉英、英汉索引的计算机处理工作;谢玲玉、陶建国及李梅芳等同志参加了图表绘制工作,都为本书的完成作出了贡献。上海医科大学出版社的编辑对书稿进行了十分认真细致的审阅和推敲,减少了原稿不少业务上和版面上的问题,特此一并感谢。

在《医学统计方法》一书发行的13年中,不少读者对该书内容提出了宝贵的意见,在此我们表示衷心的感谢。由于我们的知识和经验有限,本书难免还有各种问题,恳切希望读者继续提出意见和建议。

金丕焕
1992年11月

目 录

前言

第一章 绪论	1
第一节 统计方法在医学科学研究及疾病防治工作中的作用	1
第二节 统计研究的对象	1
第三节 几个基本概念	2
一、总体与样本	2
二、抽样误差	2
三、随机化	2
四、概率	3
五、统计量与参数	3
第四节 正确运用统计学的观点与思想方法	3
第二章 统计资料的收集和整理	5
第一节 统计资料的来源	5
一、报表资料	5
二、报告卡(单)	5
三、日常医疗卫生工作记录	6
四、实验数据	7
五、现场调查资料	7
第二节 统计资料的记录和贮存	7
一、资料的记录	7
二、资料的贮存	9
第三节 统计资料的检查和整理	11
一、资料的检查	11
二、缺项与差错	11
三、资料的整理	12
第三章 平均数与标准差	15
第一节 算术均数和几何均数	15
一、算术均数	15
二、几何均数	18
第二节 中位数和百分位数	20
一、中位数	20
二、百分位数	21
第三节 标准差	22

一、标准差的意义	22
二、不分组资料的标准差计算	23
三、离均差平方和的简化算法	23
四、分组资料的标准差计算	25
五、标准差的应用	25
第四章 均数的抽样误差与 t 分布	30
第一节 均数的抽样误差	30
一、抽样与抽样误差	30
二、均数的标准误	30
第二节 总体均数的可信区间与 t 分布	31
一、大样本资料均数的可信区间	31
二、 t 值与 t 分布	32
三、小样本资料均数的可信区间	33
第五章 两均数差别的统计意义检验	34
第一节 统计意义检验的基本步骤	34
第二节 样本均数和总体均数差别的统计意义检验	35
第三节 配对样本的 t 检验	36
一、自体比较(自身对照比较)的 t 检验	37
二、配对实验的 t 检验	38
第四节 两样本均数差别的统计意义检验	39
一、两样本均数差别的 t 检验	39
二、关于非正态分布资料均数差别的检验	40
第五节 方差不齐时两样本均数差别的统计意义检验(t' 检验)	41
第六节 两种检验与两类错误	42
一、单侧检验与双侧检验	42
二、第一类错误和第二类错误	43
三、差别统计意义的正确理解	44
第七节 正态性检验	44
一、正态概率纸法	44
二、偏度与峰度检验	45
三、 W 检验	48
四、 D 检验	49
第六章 方差分析(一)	50
第一节 方差分析的基本概念	50
第二节 完全随机设计资料的方差分析	51
一、各组样本含量相等	51
二、各组样本含量不相等	53
第三节 随机区组设计资料的方差分析	54
第四节 均数间的相互比较	56

一、各组样本含量相等	56
二、各组样本含量不相等	57
第五节 几个处理组均数与一个对照组均数比较	58
第六节 拉丁方设计资料的方差分析	60
第七节 方差齐性检验	62
一、各组样本含量相等	63
二、各组样本含量不相等	64
第八节 近似 F 检验	65
第九节 方差分析中的数据转换	66
一、平方根转换	67
二、反正弦函数转换	68
三、对数转换	69
第七章 方差分析(二)	71
第一节 2×2 析因实验设计	71
第二节 “ 2^3 设计”或“ $2 \times 2 \times 2$ 设计”	73
第三节 次级组样本含量不相等	76
一、两因素各次级组样本含量不相等, 但呈一定的比例	76
二、次级组样本含量不相等且不成比例	78
第四节 不同模型资料的 F 检验	83
第八章 正交试验设计方差分析	85
第一节 两水平正交试验	85
一、一些基本概念	85
二、正交试验及其分析的基本步骤	87
三、无重复两水平正交试验	87
四、有重复两水平正交试验	89
第二节 三水平及其他正交试验	90
一、三水平正交试验	90
二、在两水平正交表上安排四水平的因素	92
第三节 正交试验中的一些应用问题	96
一、参数估计	96
二、关于特定条件的选择	97
三、两水平正交表的简易制作法及交互作用列的确定	98
第九章 直线回归与相关	100
第一节 直线回归	100
一、概述	100
二、直线回归分析的具体步骤和方法	101
三、大样本资料的回归分析	103
第二节 直线回归分析中的误差及可信区间	105
一、标准估计误差	105

二、实测值围绕回归线的离散度	106
三、回归系数的标准误	106
四、 $\bar{y}$ 的标准误	107
五、 μ_y 的可信区间	107
六、 $\hat{y}_i$ 的标准误	107
七、 μ_{y_i} 的可信区间	108
八、截距 a 的误差及总体参数 α 的可信区间	108
九、单一个体 y_i 值的预测	108
第三节 回归系数和截距的统计意义检验	110
一、回归系数的 t 检验	110
二、回归系数的方差分析	111
三、两个回归系数差别的统计意义检验	112
四、截距的统计意义检验	114
五、两条回归线高度差别的统计意义检验	114
第四节 相关	115
一、相关系数的意义	115
二、相关系数的计算方法	116
三、相关系数的统计意义检验—— t 检验	117
四、两个相关系数差别的统计意义检验	118
五、相关和回归的关系	119
第十章 多元线性回归和曲线回归	120
第一节 多元线性回归	120
一、多元回归分析的原理和多元回归方程的求法	120
二、多元回归分析中的假设检验	122
三、标准化偏回归系数	123
四、偏回归系数的统计意义检验	124
五、复相关系数和偏相关系数	125
第二节 曲线回归	128
一、概述	128
二、医学上常用曲线的配合方法	130
三、曲线配合的拟合度——相关指数	134
第十一章 协方差分析	136
第一节 协方差分析法的意义和应用	136
第二节 完全随机设计资料协方差分析	137
第三节 协方差分析方法简化算法	141
第十二章 相对数分析	143
第一节 相对数	143
一、构成比	143
二、率	143

三、构成比和率的应用	144
第二节 率的差别的统计意义检验	146
一、率的标准误	146
二、总体率的可信区间	147
三、样本率与总体率差别的统计意义检验	147
四、两个样本率差别的统计意义检验	148
第三节 率的标准化法	149
一、标准化率的计算方法	149
二、标准的选择	152
三、标准化率差别的统计意义检验	153
第四节 重复检验结果总阳性率的计算	156
一、总阳性率的计算	156
二、标准误的计算	157
三、两个总阳性率的比较	157
第十三章 卡方检验(一)	159
第一节 四格表资料的 χ^2 检验	159
第二节 行 $\times$ 列表的 χ^2 检验	162
一、检验方法	162
二、检验结果的说明	163
第三节 四格表精确检验法	164
一、有实际数为零的情况	164
二、没有实际数为零的情况	165
三、查表法	167
第四节 配对计数资料的 χ^2 检验	167
一、两种处理方法的比较	168
二、两种以上处理方法的比较	168
第十四章 卡方检验(二)	170
第一节 $2\times c$ 表 χ^2 的计算公式和 χ^2 值的分割	170
一、 $2\times c$ 表 χ^2 的几种计算公式	170
二、 χ^2 值及自由度的分割	172
第二节 组内分组资料的卡方检验	172
第三节 内部构成不同的两个率差别的统计意义检验	174
一、加权 χ^2 检验法(Cochran 检验)	174
二、 χ^2 值相加法	176
三、 χ 值相加法	177
第四节 分表理论数合并法	177
一、 2×2 表理论数合并法	178
二、行 $\times$ 列表理论数合并法	179
第五节 行 $\times$ 列表的分割	180

一、 $2 \times k$ 表的分割	180
二、行 $\times$ 列表的分割	182
第六节 $2 \times c$ 表线性回归的统计意义检验	183
一、按性质分组的资料	183
二、按数量分组的资料	185
三、百分率很小(或很大)时的资料	186
四、药物疗效比较	186
第十五章 二项分布	188
第一节 概述	188
第二节 二项分布的条件	189
第三节 二项分布的形态	190
第四节 二项分布在医学上的应用	191
一、应用二项分布的展开, 计算概率大小, 观察疗效	191
二、计算总体率的可信区间	191
三、研究患某病的家族集积性	191
四、混合样品的分析	193
第十六章 普哇松分布	195
第一节 普哇松分布的配合	195
第二节 配合适度检验	197
第三节 普哇松分布的一些特征	197
一、普哇松分布的方差	197
二、普哇松分布的形态	198
三、普哇松分布的可加性	198
四、总数的分布	198
第四节 总体均数的可信区间	200
第五节 用普哇松分布对聚集性作研究	201
第六节 普哇松计数差别的统计意义检验	202
一、样本计数和总体均数差别的统计意义检验	202
二、两样本计数差别的统计意义检验	203
三、多个样本计数差别的统计意义检验	204
第七节 稀释法估计细菌数	205
一、只有一种稀释度时细菌数的估计	205
二、有几种稀释度时细菌数的估计	206
第十七章 圆分布资料分析	208
第一节 基本概念	208
第二节 位置、时间变换为角度	209
第三节 平均角及其假设检验	210
第四节 两样本的比较	212
第五节 多个样本的比较	214

第十八章 非参数统计(一)	215
第一节 成对资料的检验	215
一、符号检验	215
二、符号等级检验(Wilcoxon 法)	217
第二节 成组资料的检验	218
一、两样本等级和检验(Wilcoxon, Mann and Whitney 法)	218
二、中位数检验	219
第三节 完全随机化设计资料的检验	220
一、 H 检验(Kruskal and Wallis 法)	220
二、各处理组间相互比较	222
三、中位数检验法	223
第四节 等级分组资料的检验	223
第五节 随机区组设计资料的检验	224
一、 M 检验(Friedman 法)查表法	224
二、 F 检验	226
三、各组间相互比较	226
四、中位数检验	227
第十九章 非参数统计(二)	228
第一节 等级相关	228
一、Spearman 等级相关	228
二、Kendall 等级相关	229
三、 $2 \times K$ 列联表的等级相关	231
四、 $R \times C$ 列联表的等级相关	233
五、等级的一致性检验	234
第二节 柯尔莫哥洛夫—斯米尔诺夫检验	235
一、样本分布与总体分布的比较	235
二、两个样本分布的比较	237
第二十章 Ridit 分析	239
第一节 样本与总体比较	239
一、样本与总体差别的统计意义检验	239
二、用 S_R^2 的近似值作统计意义检验	241
三、多个样本与某总体比较	242
四、应用本法的注意点	242
第二节 两组平均 Ridit 值比较	243
一、两组平均 Ridit 值差别的统计意义检验	243
二、精确法检验	244
第三节 多组平均 Ridit 值的比较	244
第二十一章 序贯检验	246
第一节 序贯检验的意义	246

第二节 单一处理效果的检验——单向检验	247
一、计数资料	247
二、计量资料	248
第三节 两种处理的配对比较——双向检验	250
一、计数资料	251
二、计量资料	253
第四节 闭锁型序贯检验	255
一、计数资料	255
二、计量资料	256
第二十二章 逐步回归分析	259
第一节 基本概念	259
第二节 计算方法和步骤	259
第三节 逐步回归分析注意点	264
第二十三章 判别分析	266
第一节 基本概念	266
第二节 两类判别, 两个指标	267
一、计算判别式	267
二、临界值	269
三、回代	269
四、判别能力的统计意义检验	269
五、各因素贡献率	270
六、图示法	270
七、新个体的判别	271
第三节 三个指标的两类判别	271
第二十四章 主成分分析	273
第一节 概述	273
第二节 二维主成分分析	274
第三节 主成分分析的一般步骤	277
第四节 主成分分析应用举例	279
第二十五章 因子分析	280
第一节 概述	280
第二节 因子分析的步骤	282
一、建立模型	282
二、解释共性因子的实际意义	287
第三节 因子得分	287
第四节 因子分析应用举例	288
第二十六章 聚类分析	290
第一节 概述	290
第二节 R 型聚类	291

第三节 Q 型聚类	294
第二十七章 BAXES 定理的应用	299
第一节 一个指标时的应用	299
第二节 先验概率对后验概率的影响	302
第三节 多个指标时的应用	303
第四节 似然度法	305
第五节 计量诊断表	306
第二十八章 生存率统计	309
第一节 小样本资料的生存率分析	309
一、原始资料的过录	309
二、寿命表的制作	310
三、生存率的图示	312
四、标准误	312
五、平均生存时间	313
第二节 生存率差别的统计意义检验	313
一、时序检验	314
二、Gehan 比分检验	317
第三节 分组资料的生存率分析	318
一、分组资料对中断观察者的处理	318
二、原始资料的过录	319
三、计算步骤	319
四、生存率曲线	320
五、标准误的计算	321
第四节 分组资料生存率差异的时序检验	321
一、计算方法	321
二、消除其他因素影响后的比较	322
三、同时消除两个因素的影响	323
四、两个以上水平的比较	323
五、趋势检验	324
第二十九章 半数效量	326
第一节 基本概念	326
第二节 目测法	327
一、基本原理	327
二、具体步骤	328
三、标准误和可信区间	329
四、注意点	329
第三节 加权直线回归法	329
一、加权直线回归线的求法	330
二、回归系数的 χ^2 检验	332

三、误差和可信区间	333
四、关于自然死亡率	334
第四节 Kärber 氏法(寇氏法)	334
第五节 序贯法	335
第六节 半数效量的有关应用	337
一、效力比与毒力比	337
二、两个治疗指数的比较	338
第三十章 流行病学调查资料分析	339
第一节 成组病例一对照研究资料	339
一、大样本四格表资料	340
二、小样本四格表资料	343
三、几个四格表资料的联合分析	345
四、公共比数比	346
五、公共比数比的可信区间	347
六、比数比的一致性检验	348
七、比数比的趋势检验	349
八、多个暴露水平—— $2 \times k$ 表资料的分析	349
第二节 配对病例一对照研究资料	351
一、配对的四格表	351
二、 $1:M$ 配对资料	352
三、对照数 M 不相等时的 $1:M$ 配对资料	357
第三节 定群研究资料分析	358
一、暴露人年数的计算	358
二、率的直接法标准化及其有关指标	360
三、率的间接法标准化和标准化死亡比	364
四、危险度分析	367
五、定群研究和生存分析	369
第三十一章 Logistic 回归分析	370
第一节 成组资料分析——非条件 logistic 回归	370
一、Logistic 回归分析的概念	370
二、建立 logistic 回归方程	371
三、四格表的 logistic 回归	371
四、 $2 \times 2 \times 2$ 表的 logistic 回归	373
五、 $2 \times 2 \times 3$ 表的 logistic 回归	376
六、 $2 \times 2 \times 2 \times 2$ 表的 logistic 回归	377
七、对于病例一对照研究资料的 logistic 回归的几点说明	378
八、自变量效应(即 logistic 回归系数)的统计意义检验	379
第二节 配对资料分析——条件 logistic 回归	384
第三十二章 比例风险模型——COX 回归	387

第一节 比例风险模型的概念和原理	387
第二节 回归系数的估计方法	389
第三节 自变量效应的统计意义检验	393
一、似然比检验	393
二、检验 x_1 和 x_2 的交互作用	394
三、检验自变量的效应	394
第四节 应用比例风险模型的注意点	395
第三十三章 抽样设计	396
第一节 单纯随机抽样	396
第二节 机械抽样(系统抽样)	397
第三节 分层抽样	398
第四节 整群抽样	400
第三十四章 研究设计	402
第一节 研究设计的基本内容	402
一、确定研究目的	402
二、确定研究对象	402
三、确定研究对象的数量	402
四、确定观察指标	402
五、设立适当的对照	403
六、随机化分组	404
七、制定较好的原始记录表	404
八、考虑好统计分析的方法	405
第二节 不同场合的研究设计	405
一、调查研究设计	405
二、临床研究设计	405
三、动物实验设计	406
第三节 研究设计的类型	406
一、完全随机化设计	406
二、随机区组设计	407
三、拉丁方设计	409
四、序贯设计	412
五、其他设计	412
第三十五章 样本大小与极端值处理	413
第一节 对平均数作抽样调查时的样本大小	413
第二节 对率作抽样调查时的样本大小	414
第三节 对样本均数与总体均数的差别作统计意义检验时所需样本大小	415
第四节 成组比较即两个样本均数的差别作统计意义检验时所需样本大小	416
第五节 成对比较(或自身前后对比)作差别的统计意义检验时所需样本大小	417
第六节 对两个率的差别作统计意义检验时所需样本大小	417

第七节 病例—对照研究中的样本大小	418
一、不配对的病例—对照研究所需的样本含量	418
二、配对的病例—对照研究所需的样本含量	421
第八节 极端值的处理	421
一、正态分布数据	422
二、普哇松分布数据	422
三、二项分布数据	423
第三十六章 统计表与统计图	424
第一节 统计表	424
一、统计表的结构与编制要求	424
二、统计表的种类	424
三、统计表的审查和修改	425
第二节 统计图	426
一、绘制统计图的要求	426
二、常用统计图的绘制方法	427
附录一 统计用表	435
附表 3-1 正态分布表	435
附表 4-1 t 值表	436
附表 5-1 百分率与概率单位对照表	437
附表 5-2 计算统计量 W 所必需的系数 a_i 值	437
附表 5-3 W 检验统计量 W 的 P 分位数	439
附表 5-4 D 检验统计量 y 的 P 分位数	439
附表 6-1 F 值表(双侧检验, 方差齐性检验用)	440
附表 6-2 F 值表(单侧检验, 方差分析用) $P=0.05$	441
附表 6-3 F 值表(单侧检验, 方差分析用) $P=0.01$	442
附表 6-4 各均数间相差有统计意义时所需之 Q 值	443
附表 6-5 Dunnett t 表(单侧检验)	444
附表 6-6 Dunnett t 表(双侧检验)	445
附表 6-7 χ^2 值表	446
附表 6-8 相当于百分比之角度	447
附表 8-1 $m=2$ 的正交表	449
附表 8-2 $m=3$ 的正交表	453
附表 8-3 $m=4$ 的正交表	455
附表 9-1 γ 值表	456
附表 9-2 由 γ 转 Z 值表	457
附表 13-1 $\lg n!$ 数值表	457
附表 13-2 四格表的统计意义检验用表(O 值表)	458
附表 15-1 百分率的可信区间	470
附表 16-1 普哇松分布用的上下可信区间	473

附表 17-1	Rayleigh's Z 的临界值表	473
附表 17-2	Watson and Williams 检验中校正因子	474
附表 17-3	Watson's U^2 检验的临界值表	475
附表 18-1	符号等级检验表(Wilcoxon 成对比较用)	477
附表 18-2	等级总和数临界值(双侧检验)	477
附表 18-3	H 值与概率对照表	478
附表 18-4	M 值的界限值($P=0.05$)	480
附表 19-1	等级相关系数的统计意义界限表	480
附表 19-2	样本与总体比较的 Kolmogorov-Smirnov 检验临界值表	481
附表 19-3	两样本比较的 Kolmogorov-Smirnov 检验临界值表	481
附表 21-1	计数资料单向序贯检验边界系数表	482
附表 21-2	计量资料单向序贯检验 $\frac{\theta_1 - \theta_2}{\sigma}$ 值表	484
附表 21-3	配对计数资料序贯检验边界系数表	485
附表 21-4	配对计量资料序贯检验边界系数表($\alpha = \beta = 0.05$)	485
附表 21-5	计数资料闭锁型序贯检验边界(U 、 L)和中间线(M 、 M')坐标	485
附表 21-6	计量资料序贯检验闭锁线坐标 n' 、 y'_n 数值	486
附表 29-1	概率单位与权重系数对照表	487
附表 29-2	作业用概率单位之极小值、极大值及全距	487
附表 30-1	SMR 的 95% 可信区间系数表	488
附表 33-1	随机数字表	489
附表 35-1	对平均数作抽样调查 S/δ 取不同的数值时所需样本大小(n) ($\alpha=0.05$)	490
附表 35-2	对平均数作抽样调查 S/δ 取不同的数值时所需样本大小(n) ($\alpha=0.01$)	490
附表 35-3	对率作抽样调查时所需样本大小(n) ($\alpha=0.05$)	491
附表 35-4	对率作抽样调查时所需样本大小(n) ($\alpha=0.01$)	491
附表 35-5	对样本均数与总体均数的差别作统计意义检验时所需样本大小(n)	492
附表 35-6	对两个样本均数的差别作统计意义检验时所需样本大小(n)	493
附表 35-7	对两个率的差别作统计意义检验时每组所需样本大小(n) $\alpha=0.05, \beta=0.10$	494
附表 35-8	对两个率的差别作统计意义检验时每组所需样本大小(n) $\alpha=0.01, \beta=0.05$	494
附表 35-9	T_{N1} 的临界值表(极端值处理用)	495
附表 35-10	T_{N2} 的临界值表(极端值处理用)	495
附表 35-11	T_{P1} 的 0.05 水平临界值	495
附表 35-12	T_{P1} 的 0.01 水平临界值	496
附表 35-13	T_{P2} 的 0.05 水平临界值	496
附表 35-14	T_{P2} 的 0.01 水平临界值	496