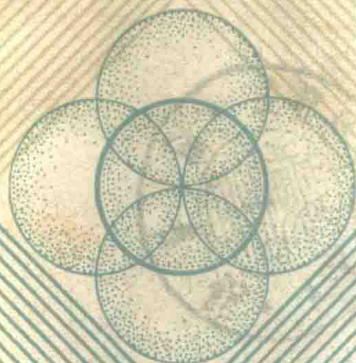


多元分析方法 及其应用

丁士晟 编著



吉林人民出版社

多元分析方法及其应用

丁士晟 编著

吉林人民出版社

多元分析方法及其应用

丁士晟 编著

*

吉林人民出版社出版 吉林省新华书店发行
长春新华印刷厂印刷

*

850×1168毫米32开本 24印张 530,000字

1981年12月第1版 1981年12月第1次印刷

印数：1—4,290册

统一书号：13091·92 定价：2.50元

内 容 简 介

本书介绍了多元分析方法及其应用,包括作者二十年来在这方面的
的工作成果,如用方差分析检验周期、用权重法建立 0.1 回归方程、
点聚图的分析 and 检验。

全书共分十章,收集了一些较复杂和较新的多元分析方法,如聚
类分析、模糊聚类、信息分析、第二类错误的检验、多因变量逐步回
归、逐步多级判别。同时又从实用角度出发,选择了一些例子,由浅入
深具体说明多元分析的方法,并详细列出了计算步骤,使读者通过范
例所示,逐步掌握各种方法解决具体问题。为了便于读者应用,本书
附录中有 28 个统计用表。

本书可供气象、农学、水文、地质、地震、医学等部门有关科技
人员、研究人员以及大专院校师生参考。

前 言

近几年来,多元分析在气象、农学、水文、地质、地震、医学等方面得到广泛应用,有必要写一本多元分析方法及其应用的书籍。本书试图尽量收集多元分析方法及其应用的最新成就,同时又从实用角度出发,选择了一些例子,由浅入深具体地说明多元分析方法,并详细地列出其计算步骤,使读者通过范例所示,能逐步掌握各种方法,解决具体问题。本书一共有171个例子,其中大部分是通例,虽然有一部分是气象方面的例子,但从事其它方面工作的同志也很容易理解。为了便于读者应用,本书附录中有28个统计用表,其中一部分是作者自己编算的,一部分录自 *Beyer, W. H.* 和山内二郎编著的统计用表,也有一部分录自常用数理统计用表等有关统计书。

全书共分十章,各章重点内容是:第一章叙述概率统计基本概念及数据处理;第二章介绍了各种统计检验,特别是第二类错误检验的方法和步骤;第三章是方差分析,从方差分析原理讲到多因素方差分析,同时为了科学安排试验,介绍了正交设计;第四章叙述各种回归分析方法,有复杂的多因变量逐步回归,也有简单有效的权重回归;第五章介绍了各种判别分析,有逐步多级判别分析和简便实用的点聚图的检验和分析;第六章介绍了信息分析及其应用;第七章叙述聚类分析,包括模糊聚类;第八章是滑动和插值;第九章讲的是拟合和配置实验曲线;第十章系统地介绍了过程分析,包括用方差分析方法检验周期。

李麦村同志审阅了本书，并提出了宝贵意见，吉林省气象科学研究所的领导和同志曾给作者鼓励和支持，在此一并致谢。由于作者水平有限，希望读者对本书的缺点和错误提出批评和指正。

丁 士 晟

1979年6月

目 录

第一章 概率统计基础.....	1
第一节 频率与概率	1
第二节 分布律	12
第三节 数字特征	22
第四节 数据处理	39
第二章 统计检验	44
第一节 统计检验概述	44
第二节 t 检验	47
第三节 χ^2 检验	54
第四节 F 检验	70
第五节 符号检验	76
第六节 秩和检验	80
第七节 游程检验	83
第三章 方差分析	88
第一节 方差分析原理	88
第二节 单因素方差分析	91
第三节 双因素方差分析	100
第四节 多因素方差分析	112
第五节 正交设计	129
第六节 极差的运用	143
第四章 回归分析	155
第一节 相关系数	155
第二节 一元线性回归	177
第三节 二元线性回归	187
第四节 多元线性回归	197

第五节	逐步回归	206
第六节	逐次回归	218
第七节	概率回归	223
第八节	0,1 回归	234
第九节	数量化方法	241
第十节	多因变量逐步回归	247
第十一节	非线性回归	251
第五章	判别分析	259
第一节	二级判别	259
第二节	多级判别	275
第三节	逐步判别	291
第四节	序贯判别	300
第五节	点聚图的检验和分析	309
第六章	信息分析	315
第一节	熵	315
第二节	条件熵	323
第三节	信息	331
第四节	熵在气象中的应用	341
第五节	信息在气象中的应用	346
第六节	信息时空分布	356
第七章	聚类分析	362
第一节	聚类分析概述	362
第二节	聚类统计量	363
第三节	系统聚类法	384
第四节	逐步聚类法	395
第五节	逐个聚类法与分解法	406
第六节	最优分割法	409
第七节	筛选因子分类法	421
第八节	模糊聚类	437
第八章	滑动与插值	445
第一节	试验曲线的平滑	445

第二节	有周期曲线的滑动	464
第三节	场的平滑	469
第四节	一维插值	477
第五节	二维插值	487
第九章	曲线与曲面拟合	495
第一节	已知函数的曲线拟合	495
第二节	未知函数的曲线拟合	505
第三节	曲面拟合	523
第十章	过程分析	540
第一节	趋势分析	541
第二节	阶段分析	545
第三节	周期分析	548
第四节	自相关分析	566
第五节	马尔可夫链分析	575
第六节	多维时间序列分析	591

附 录

表 1	正态分布表	598
表 2	正态分布的双侧分位数(u_α)表	604
表3—1	极差 $d(n, l)$ 和 $\phi(n, l)$ 表	605
表3—2	极差 $d'(n, l)$ 和 $\phi'(n, l)$ 表	606
表 4	t 分布的双侧分位数(t_α)表	607
表 5	t 分布 β 检验表	609
表 6	χ^2 分布的上侧分位数(χ_α^2)表	613
表 7	χ^2 分布 β 检验表	615
表 8	2×2 最大 Y 值检验表	616
表 9	F 检验的临界值 (F_α) 表	649
表10	F 分布 β 检验表	659
表11	符号检验表	666
表12	秩和检验表	667
表13	游程检验表	668

表14	正交拉丁方表	673
表15	正交表	677
表16	多重比较中的 q 表	692
表17	相关系数检验表	694
表18	权重表	695
表19	点聚图检验表	697
表20	点聚图分析表	707
表21	$-P\log_2 P$ 表	709
表22	熵值 H 表	714
表23	$\log_2 N$ 表	717
表24	相关系数信息表	719
表25	年天气日数熵值 H 表	720
表26	月天气日数熵值 H 表	724
表27	正交多项式表	730
表28	皮尔森Ⅲ型分布表	752

参 考 文 献

第一章 概率统计基础

第一节 频率与概率

一 几个基本概念

自然界的现象可以分为两大类。一类是必然现象，这些现象符合因果规律，具有必然性。如在一个大气压情况下，温度低于 0°C ，水就要结冰，温度高于 100°C ，水就要沸腾。另一类是随机现象，在同样条件下重复试验或观测中，有时出现有时不出现具有偶然性的现象，又称为随机事件，简称为事件。如下雨同时打雷，就是随机事件，因为下雨时有时打雷有时不打雷。

在试验和观测中有若干随机事件 $A_1, A_2, \dots, A_i, \dots, A_n$ 中至少发生某一事件所组成的事件 C ，称为事件 C 是 $A_1, A_2, A_3, \dots, A_i, \dots, A_n$ 诸事件之和。并记为

$$C = A_1 + A_2 + \dots + A_i + \dots + A_n = \sum_{i=1}^n A_i \quad (1.1)$$

这就是事件的和。

例 1—1：如进行两次预报大风试验， A_1 表示只报对一次， A_2 表示两次均报对， A_3 表示两次均没有报对。

最多报对一次的事件 $C_1 = A_1 + A_3$

至少报对一次的事件 $C_2 = A_1 + A_2$

如果有一组事件 $A_1, A_2, A_3, \dots, A_i, \dots, A_n$ 同时发生的事件记为 C ，则事件 C 为 $A_1, A_2, A_3, \dots, A_i, \dots, A_n$ 诸事件之积，并记为：

$$C = A_1 \cdot A_2 \cdot A_3 \cdots A_i \cdots A_n = \prod_{i=1}^n A_i \quad (1 \cdot 2)$$

例 1—2：如果 A_1 表示下暴雨事件， A_2 表示有洪峰事件， A_3 表示水库放流事件， C 事件表示下暴雨、有洪峰同时水库放流事件，则有

$$C = A_1 \cdot A_2 \cdot A_3$$

称 C 事件是事件 A_1, A_2, A_3 三个事件之积。

如果有 A, B 两个事件，并把不可能事件记为 $\vee$ ，把必然事件记为 $\sqcup$ 。

若 $A \cdot B = \vee$ ，则事件 A 和事件 B 为互不相容事件。

若 $A \cdot B = \vee$ ，同时有 $A + B = \sqcup$ 则事件 A 和事件 B 称为互逆事件，事件 A 称为事件 B 的逆事件，事件 B 也称为事件 A 的逆事件。

$$\text{并记为 } \overline{A} = B \quad \overline{B} = A$$

下面介绍几个概率论、数理统计常用名词。

总体：由同性质成员或个体所构成的集体称为总体，它是我们研究对象的全体。总体所包含的成员往往是无穷的，是设想的或抽象的全体。总体又称为母体、母群、全群。

样本：由总体中抽出部分同性质的成员或个体所组成的集体称为样本。在实际工作中不可能得到总体的全部信息，只能用样本来推测总体的性质，处理实际问题往往和样本打交道。样本又称为子群、子样。

个体：总体或样本中某一个成员称为个体。每次观测或测量得到的数据就是一个个体的数据。

例 1—3：某河流解放后才有水文观测资料，由 1949 年到

1978年有30年流量资料，这30年水文资料就是样本。而某河流流量的总体是指有地球以来该河流的流量情况，既包括有资料的30年流量情况，也包括没有水文观测的历史时代的流量情况。而每次水文观测就是一个个体。

样本容量：由观测或试验得到的样本所包含个体的数量称为样本容量。一般统计上把样本容量大于30称为大样本，样本容量小于30称为小样本。

频数：某一样本在某一条件下出现的次数称为频数，记为 M_i 。

例 1—4：如某森林测量一地块的树木直径，其直径 ≤ 10.0 厘米的有60棵树，直径在 10.1 到 20.0 厘米的有80棵树，直径 20.1 到 30.0 厘米的有100棵树，直径 ≥ 30.1 厘米的有50棵树。可以记为 $d \leq 10.0$ 厘米时频数 $M_1 = 60$ ， $10.1 \leq d \leq 20.0$ 时频数 $M_2 = 80$ ， $20.1 \leq d \leq 30.0$ 时频数 $M_3 = 100$ ， $d \geq 30.1$ 时频数 $M_4 = 50$

频率：某一样本在某一条件下出现的次数与样本容量的比值称为频率，并记为 f_i 。

$$f_i = \frac{M_i}{N} \quad (1.3)$$

其中 N 为样本容量， M_i 为频数。频率是描述某一样本在某一条件下出现的可能性。

例 1—5：据例 1—4 计算频率可以有

样本容量： $N = 60 + 80 + 100 + 50 = 290$

$$d \leq 10.0 \text{ 厘米的频率 } f_1 = \frac{M_1}{N} = \frac{60}{290} = 0.207$$

$$10.1 \leq d < 20.0 \text{ 厘米的频率 } f_2 = \frac{M_2}{N} = \frac{80}{290} = 0.275$$

$$20.1 \leq d < 30.0 \text{ 厘米的频率 } f_3 = \frac{M_3}{N} = 0.345$$

$$30.0 \text{ 厘米} \leq d \text{ 的频率 } f_4 = \frac{M_4}{N} = 0.172$$

频率是小于或等于 1 的正数。即

$$0 \leq f \leq 1 \quad (1.4)$$

频数和频率在气象中应用很多，在气表中就有风向频数和频率，不同雨量出现的频数和频率，不同天气现象出现的频数等等。

大量的试验证明，事件的频率具有一定稳定性。当样本容量很大时，频率在某一数值周围摆动，事件发生的可能性可以用一个数来表示。这个描述事件 A 在试验中发生可能性程度小于 1 的正数叫随机事件的概率，记作 $P(A)$ 。

可以认为当样本容量 $N \rightarrow \infty$ 时，事件的频率就是概率。

$$f(A)_{N \rightarrow \infty} = P(A) \quad (1.5)$$

当 N 很大时，频率 $f(A)$ 是概率 $P(A)$ 很好的近似值。

上面介绍了由频率出发来定义概率，下面再介绍一下概率的古典定义。假设试验的一切可能结果中，可以由 N 个互不相容事件，并且是等可能事件所组成的完备事件组，而其中有 M 个事件是有利随机事件 A 的，则随机事件 A 的概率等于有利于随机事件 A 出现的次数 M 与事件的总数 N 的比值。

$$P(A) = \frac{M}{N} \quad (1.6)$$

例 1—6：如某地 16 个风的方位是等可能性的，求偏南风的概率。偏南风包括 $ESE, SE, SSE, S, SSW, SW, WSW$ ，七个方位，偏南风的概率 $P = \frac{7}{16} = 0.4375$

如果我们在事件 B 已经发生条件下计算事件 A 的概率，则这种概率叫做事件 A 在事件 B 已发生的条件下的条件概率，记作 $P(A|B)$ 。

例 1—7：如果打雷的事件记为 B ，在打雷时下雨事件 A 的

概率为0.32, 则这种概率为打雷情况下下雨的条件概率, 记作 $P(A|B) = 0.32$ 。

二 概率加法定理

概率加法定理: 两互不相容事件的和的概率, 等于这二事件的概率的和。事件 A 、事件 B , 若 $A \cdot B = \emptyset$ 则 $P(A+B) = P(A) + P(B)$ (1.7)

证明如下:

如有 N 个事件组成完备事件组, 其中有 M_1 个有利事件 A , M_2 个有利事件 B , 并且事件 A 和事件 B 为互不相容事件, 所以对 A 和 B 事件的有利事件共 $M_1 + M_2$ 个, 于是有:

$$\begin{aligned} P(A+B) &= \frac{M_1 + M_2}{N} = \frac{M_1}{N} + \frac{M_2}{N} \\ &= P(A) + P(B) \end{aligned}$$

例 1—8: 如 A 、 B 是互不相容事件, 事件 A 的概率 $P(A) = 0.21$, 事件 B 的概率 $P(B) = 0.12$, 则这二事件的概率的和 $P(A+B) = P(A) + P(B) = 0.21 + 0.12 = 0.33$

同样对于多个互不相容事件的概率加法定理为: 有限个互不相容事件之和的概率等于这些事件的概率的和。

有事件 $A_1, A_2, A_3, \dots, A_i, \dots, A_n$ 为互不相容事件, $A_1 \cdot A_2 \cdot A_3 \cdot \dots \cdot A_i \cdot \dots \cdot A_n = \emptyset$ 则

$$\begin{aligned} &P(A_1 + A_2 + A_3 + \dots + A_i + \dots + A_n) \\ &= P(A_1) + P(A_2) + P(A_3) + \dots + P(A_i) \\ &\quad + \dots + P(A_n) \end{aligned} \quad (1.8)$$

对于事件 $A_1, A_2, A_3, \dots, A_i, \dots, A_n$ 组成互不相容的完备事件组, 即

$$P(A_1) + P(A_2) + P(A_3) + \dots + P(A_i)$$

$$+ \cdots + P(A_n) = 1$$

$$A_1 \cdot A_2 \cdot A_3 \cdots A_i \cdots A_n = \bigvee \quad \text{则}$$

$$\begin{aligned} & P(A_1 + A_2 + A_3 + \cdots + A_i + \cdots + A_n) \\ &= P(A_1) + P(A_2) + P(A_3) + \cdots + P(A_i) \\ &+ \cdots + P(A_n) = 1 \end{aligned} \quad (1.9)$$

对于互逆事件 A 和 $\bar{A}$ ，因为有

$$P(A) + P(\bar{A}) = 1$$

$$A \cdot \bar{A} = \bigvee$$

$$\text{则 } P(A + \bar{A}) = P(A) + P(\bar{A}) = 1$$

也即互逆事件的概率和等于 1。

例 1—9：如 A 、 B 、 C 是互不相容事件，事件 A 的概率 $P(A) = 0.21$ ，事件 B 的概率 $P(B) = 0.12$ ，事件 C 的概率 $P(C) = 0.23$ 。 A 、 B 、 C 这三事件的概率的和

$$P(A + B + C) = P(A) + P(B) + P(C)$$

$$= 0.21 + 0.12 + 0.23 = 0.56$$

例 1—10：如某人做水文预报 20 次报对 15 次，报错 5 次，那末预报 3 次有报错的概率是多少。

$$\text{报错一次的概率 } P(A_1) = \frac{C_5^1 C_{15}^2}{\sum_{i=0}^3 C_5^i C_{15}^{3-i}} = \frac{525}{1140} = 0.46$$

$$\text{报错二次的概率 } P(A_2) = \frac{C_5^2 C_{15}^1}{\sum_{i=0}^3 C_5^i C_{15}^{3-i}} = \frac{150}{1140} = 0.13$$

$$\text{报错三次的概率 } P(A_3) = \frac{C_5^3 C_{15}^0}{\sum_{i=0}^3 C_5^i C_{15}^{3-i}} = \frac{10}{1140} = 0.01$$

报错的概率：

$$P(A_1 + A_2 + A_3) = P(A_1) + P(A_2) + P(A_3) = 0.60$$

三 概率乘法定理

概率乘法定理：二事件的积的概率等于其中一事件的概率与另一事件在上一事件已发生条件下的条件概率的乘积。

$$\begin{aligned} P(A \cdot B) &= P(A) \cdot P(B|A) \\ &= P(B) \cdot P(A|B) \end{aligned} \quad (1.11)$$

证明如下：

如有 N 个事件组成完备事件组，其中有利于事件 A 的有 M_1 个，有利于事件 B 的有 M_2 个，而有利于事件 $A \cdot B$ 的有 M 个，显然 $M \leq M_1$ ， $M \leq M_2$ 。如果已知事件 A 发生，则表明有利于事件 A 的 M_1 个基本事件中必有一件发生。所以这时有利于事件 B 的基本事件仅有 M 个。

$$P(B|A) = \frac{M}{M_1} = \frac{\frac{M}{N}}{\frac{M_1}{N}} = \frac{P(AB)}{P(A)}$$

可得 $P(AB) = P(A)P(B|A)$

同理有 $P(AB) = P(B)P(A|B)$

同样可以推广到多个互不相容事件的概率乘法定理。

有限个事件的积的概率等于这些事件概率的乘积，其中每一事件的概率是在它前面一切事件都已发生条件下的条件概率。

$$\begin{aligned} &P(A_1 A_2 A_3 \cdots A_i \cdots A_n) \\ &= P(A_1)P(A_2|A_1)P(A_3|A_1 A_2) \cdots \\ &P(A_n|A_1 A_2 \cdots A_{n-1}) \end{aligned} \quad (1.12)$$

例 1—11：某地7、8月打雷的概率为0.2，下暴雨的概率为0.05，下暴雨打雷的概率为0.02，求打雷条件下暴雨的概率和