

<<应用多变量统计分析>>

图书基本信息

书名：<<应用多变量统计分析>>

13位ISBN编号：9787030321756

10位ISBN编号：7030321758

出版时间：2011-8

出版时间：科学出版社

作者：孙尚拱

页数：339

版权说明：本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问：<http://www.tushu007.com>

## <<应用多变量统计分析>>

### 内容概要

《应用多变量统计分析》由孙尚拱编著，介绍了多变量统计分析的基本理论及其各种常用模型。全书共有11章，内容包括绪言，矩阵的某些补充知识，多元正态分布，假设检验，多元线性模型，实用多元线性回归与典则相关分析，判别分析，主成分分析与因子分析，隐变量分析，聚类分析，生存分析。

书中配有大量例题、习题，并且例题都写出了sas计算程序。

随书配的光盘中提供了书中大量数据的电子版，以方便读者使用。

《应用多变量统计分析》可供普通高等院校数学、应用数学、统计学等各专业高年级本科生及研究生作为教材使用，也可供相关专业研究人员参考使用。

<<应用多变量统计分析>>

书籍目录

- 前言
- 第1章 绪言
- 第2章 矩阵的某些补充知识
- 第3章 多元正态分布
- 第4章 假设检验
- 第5章 多元线性模型
- 第6章 实用多元线性回归与典则相关分析
- 第7章 判别分析
- 第8章 主成分分析与因子分析
- 第9章 隐变量分析
- 第10章 聚类分析
- 第11章 生存分析
- 参考文献
- 附录

## &lt;&lt;应用多变量统计分析&gt;&gt;

## 章节摘录

第1章 绪言统计学是收集及分析统计数据的学科。

随着社会及科技的快速发展，特别是统计软件的发展，对统计数据的工作已变得越来越容易，但对统计分析的要求也越来越高及深入。

统计分析工具在日新月异地进步，可是人们的思想常带有一定的惯性，不一定都与技术的进步同步。自然或社会现象及经济生活中的各种变量之间往往存在很大的相关性或依赖性。

它们的变化往往是彼此相关联的，而人们习惯的分析总是把相关的变量割裂开来，彼此独立地去分析每一个变量，这就是“多变量问题的单变量分析法”。

为了提示这种分析法存在的问题，现举例说明如下：例1.120世纪70年代初，北京市高血压防治组发现：北京炊事员的高血压患病率很高，觉得不可思议，于是重新设计了一个方案，于1974年抽查了北京市916名炊事员，调查15个变量：性别、年龄、工种、工龄、做炊事工作前的工种、一天工作的时间、班次、常在高温下工作、食量、嗜咸、素食、肾炎史、家族史、超重及是否发胖。

指标是舒张压[1]。

先用多重回归分析法找出对舒张压有显著影响的变量为年龄，体胖，肾炎史，性别，工种，家族史，嗜咸。

再用单变量分析法，即把每一个变量与舒张压作统计分析。

与上述回归分析不一致的是：“超重、工龄、素食”在单变量分析法中对舒张压有显著的影响；而在多重回归分析法中有显著影响的“家族史”在单变量分析法中对舒张压没有显著的影响。

文献[1]逐一分析了它们发生不一致的原因。

举几例说明如下：(1) 工龄对于血压的影响。

在单变量分析法中，可以把工龄与血压的关系作成表1.1的形式。

从表1.1可见：高血压患病率随工龄的增高而升高。

使用单因素统计的 $2 \times 4$ 列联表中的独立性检验公式，可算得？

$\chi^2=33:9$ ，自由度为3，查表可得 $p \leq 0.001$ 。说明高血压患病率在不同工龄的群体中有非常显著的差别（ $p \leq 0.001$ ）。

若同时对此数据作另一形式的处理，如表1.2所示，其中限定工人的年龄在40岁以上，则由于例数减少，只好对工龄分得粗一些，作成两水平：15年及以下与15年以上。

从表1.2可见，不同工龄段的高血压患病率（24.5%与26.0%）相差很小（？

$\chi^2=0:0.88$ ； $p > 0.75$ ）。

这就说明，表1.1中“工龄增加时高血压患病率也随之增加”的结果实际是一种假象。

造成这种假象的根本原因是未把隐藏在工龄背后更本质的因素（年龄）控制住。

也就是说，对同一批数据，即使使用同一形式的统计法，也可得出完全相反的结论。

但表1.2的方法已包含有多因素分析的思想，所以表1.2自然比表1.1合理一些。

(2) 体重与体胖对于血压的影响。

从表1.3的表面上来看，表中仅考察超重与不超重。

但实际上，表1.3中未把另外的因素（如年龄、体胖等）对于血压的影响扣除。

表1.4和表1.5仅是扣除了体胖的影响。

结果说明，表1.3中超重对高血压的影响实际上是体胖的影响结果。

表(3) 家族史对高血压的影响。

表1.6是单因素下家族史对高血压的影响列联表。

结果是非常不显著的，这与线性回归的结果不一致。

但表1.6中未把另外14个因素对血压可能的影响扣除，即由于这些因素的相互干扰（如比例分配不匀等）也可能造成了表1.6的假象。

而高血压有遗传性，这一点在目前也早有定论了。

对素食变量的情况的分析也有类似的问题。

把有相关性的变量割裂开来，把多变量问题简单地拆成很多单变量的分析法绝不限于中国。

## &lt;&lt;应用多变量统计分析&gt;&gt;

1978年8月9日，光明日报刊登一则科技信息：某单位对美国20个城市作饮水氟化研究，10个城市的饮水进化氟化处理，而另外10个城市未氟化作为对照，得到结论：“饮水氟化有致癌作用”。

但过了一个时期，光明日报又登载了相反的文章。

原来前述论文发表后，受到了美国癌肿协会和国皇家统计协会的怀疑。

他们派人对该批数据重新作统计分析，得到结论：饮水氟化没有发现有致癌作用，相反地，却略有保护作用。

两个结论差别如此之大，根源在于第一分析法完全是类似于表1.1的单因素分析法，而后一个结论则把两组城市中种族、生活环境上的不同所产生的影响尽可能地扣除，再去比较两组城市的癌症的患病率。

上述两例虽发生在20世纪70年代，但至今，国内外的刊物、报刊上单纯用单因素分析法公布结论仍是相当普遍。

也就是说，多变量统计分析的知识在很多人的头脑中仍是相当缺乏。

多变量统计分析所使用的数据远比单变量分析法所用的数据更易取得，只要在相同的条件下把与问题有关或可能有关的变量尽可能多地记录即可。

它远比单因素分析法必须控制（或固定）其他因素而去记录某个因素（或对它做试验）要容易得多，因为其他因素应控制或固定在什么水平上，不同的控制或固定法对结果可能会有很大的影响。

但多变量分析也有其缺点即必须使用计算机统计软件；理解计算结果要有一定的抽象思维能力；它不及单因素分析法那样直观、明了。

这些大概是因为不是每一个实际工作者都在使用或懂得多变量统计法的原因吧。

统计数据的收集常需要一定时间的周密设计再去抽样或做试验。

在过去相当长的时间内，对统计资料的分析大多停留在计算均值、百分比或加一点简单的组间比较。

我们花了大量的精力、财力（有时达几十亿人民币）去调查或做试验，为什么只仅仅去计算几个均值及百分比？

根据作者几十年的工作经验，我们认为“一年取样，十年分析”。

也就是说，如果样本的取得用了一年时间，但真正要对一批合格的大样本资料作较为全面、认真而细致的分析工作，用十年的时间是不算多的。

这主要是因为“变量与变量”，“人与变量”及“人与人”之间的关系实在是相当复杂，要充分揭露它们之间的各种内在的统计联系，绝不可能在几天或几个月就可以完成。

特别地，人的认识总是在不断地进步，因此，对资料的内在信息的提取也随时间在不断地提高。

可惜的是，即使在目前计算机高度普及的情况下，人们对资料中信息的提取仍然是普遍地提取不足。

这种对统计信息的提取不足的根源是实际工作者对多变量统计的内容了解很少，自然也就应用得少了。

例如，对中医理论的解析问题。

应该说，中医理论的实质应该是几千年以来中医大夫“经验性的多变量统计分析”的结果。

现在，高度计算机化了的计方法的应用仍不很充分。

## <<应用多变量统计分析>>

### 编辑推荐

《应用多变量统计分析》重点是介绍多变量统计的分析方法，但书内有大量的实例及练习题，它对初学者是极有帮助的。

书内的这些数据不仅可用于验证《应用多变量统计分析》的例子，也可进一步用于做各种统计分析工作。

<<应用多变量统计分析>>

版权说明

本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问:<http://www.tushu007.com>