

基于 SPSS 的城市区域地下水变异 Factor Analysis 过程

何兴江¹, 张信贵², 易念平²

(1. 成都理工大学, 成都 610059; 2. 广西大学, 南宁 530004)

[摘要]对需要大量运算的统计问题, 运用 SPSS 软件可以快速得到演算结果, 从而彻底消除重复而单调的过程, 研究人员可更多地专注于统计模型的建立。作者建立了城市建设引发城市区域的地下水缓慢变异的模型, 调用 SPSS 的 Factor Analysis 过程, 对城市区域地下水场变异起重要的组分进行提取, 以起重要作用的 K、Na、Ca、Mg、Fe、Mn 组分最少组合来作为评价地下水变异的指标体系。对南宁市区地下水中特征点进行初步定量分析表明, 城市区域地下水年变异水平是城市的建设与发展作用的结果, 随时间增长变异水平呈增长之势。从历史背景下研究城市区域地下水缓慢变异程度及趋势, 对城市地下水演化规律的认识有重要的指导意义。

[关键词]SPSS 软件 地下水环境变异 水化学组分 评价指标 因子分析

[中图分类号]P618 **[文献标识码]**A **[文章编号]**0495-5331(2006)01-0093-04

1 SPSS 简介

SPSS (Statistics Package for Social Science) 是目前世界上最优秀的统计软件之一, SPSS 涵盖统计分析的方方面面, 如: 方差分析、回归分析、相关分析、t 检验、非参数检验、因子分析、聚类分析、距离分析、多变量方差分析、非线性回归分析、曲线估计等等, 已广泛应用于自然科学、社会科学中, 其中涉及的领域包括工程技术、应用数学、经济学、商业、金融、生物学、医疗卫生、体育、心理学、农林等等, 可以说只要有对各种数据如数值型、字符型、逻辑型等进行统计分析的地方, 就有 SPSS 的用武之地。SPSS 除功能强大、应用广泛外, 还易学易用, 在面向用户的使用方面主要有以下三个突出优势: Windows 的窗口方式和界面友好的对话框; 得出的结果均以直观易懂的图表示; 拥有全面生动的帮助。另外, 对工程技术人员来说, 可使其能针对所研究的问题, 建立统计模型正确地解决问题, 对需要大量运算的统计问题运用高精的统计分析工具来快速得到演算结果, 彻底消除重复而单调的计算过程, 使工程技术人员更多地专注于统计模型的建立与模型的修改工作上^{[1][2]}。

本文不重在 SPSS 软件的介绍, 而是通过 SPSS

的 Factor Analysis 过程研究城市地下水随城市建设发生变异^{[5][6][7]}的统计模型的建立与模型分析结果的解释与评判上。

2 SPSS 的 Factor Analysis 过程

2.1 Factor Analysis (因子分析)

因子分析方法最早是 1904 年由德国心理学家 C. Spearman 提出的, 1957 年由美国 W. C. Krumbein 引进地质学, 主要用于研究沉积学, 1962 年 J. Imbrie 和 E. G. Purdy 加以完善和发展。因子分析是用于研究一组变量的相关性或研究相关矩阵内部结构的一种多元统计分析方法, 它将多个变量综合成为少数的“因子”, 也就是在较少损失原始数据信息的前提下, 用少量的因子去代替原始的变量, 从而达到对原始变量的分类, 揭露原始变量之间的内在联系。一般来说, 研究人员收集数据时希望得到的尽可能多, 而在分析、综合时又希望尽可能少, 因子分析提供了一条科学、逻辑的途径, 可把大量的原始数据精简, 以利于进行综合分析, 这种精简是以不影响预分析问题主要结论的精确性为前提, 或者说是在不损失分析问题的信息前提下进行的。同时可避免掺杂主观、片面的思维, 推理分析从而得到正

[收稿日期]2005-05-13; **[修订日期]**2005-12-26; **[责任编辑]**陈仁俊。

[基金项目]教育部科技研究重点项目 (批准号: 03105)、国家自然科学基金项目 (批准号: 40062002) 联合资助。

[第一作者简介]何兴江 (1966 年-), 男, 1989 年毕业于中国地质大学, 获学士学位, 现在成都理工大学攻读博士学位。

确的结论。并可以在复杂的成因环境中研究单一过程的性质与特征^{[3][4]}。

2.2 Factor Analysis(因子分析)的实施步骤

①对原始数据标准化,标准化的目的是为了消除原始数据量纲和数量级的影响。采用 Z - Score 变换进行标准化,其标准化公式为:

$$Z_{ij} = \frac{x_{ij} - \bar{x}_j}{S_j}$$

式中:

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij} \quad S_j^2 = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2$$

$\bar{x}_j$ 为样本均值, S_j^2 为标准差, Z_{ij} 为标准化后的数值。

②计算标准化数据的相关系数矩阵

③用雅可比方法求相关系数矩阵的特征根及相应的特征向量

④确定主成分个数:选择特征根大于某个值时的个数做为主成分个数或计算各主成分的方差贡献率及累积方差贡献率,当累积方差贡献率大于一定要求(累积方差贡献百分比)时来确定主成分个数。选择主成分个数的目的在于将为数较多的变量转化为少数几个综合变量,而且尽可能多地保留原始数据的信息,从而减少综合评价的工作量,所以一方面要保持主成分个数尽可能少,另一方面累积方差贡献率要尽可能地大。

⑤对主成分的含义进行解释

⑥由主成分与主成分累积方差贡献率计算主成分加权综合评价,并进行排序与和比较。

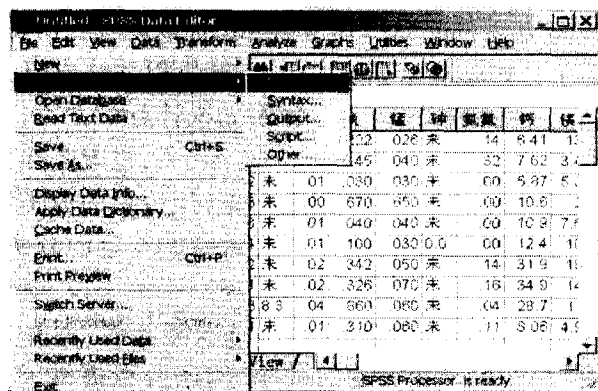


图1 打开数据表

2.3 SPSS 的 Factor Analysis 主要过程(见图 1~4)

3 实例分析 - 以南宁市地下水环境为例

3.1 Factor Analysis(因子分析)的内容

①选取南宁市某区域一点 10 年来地下水化学

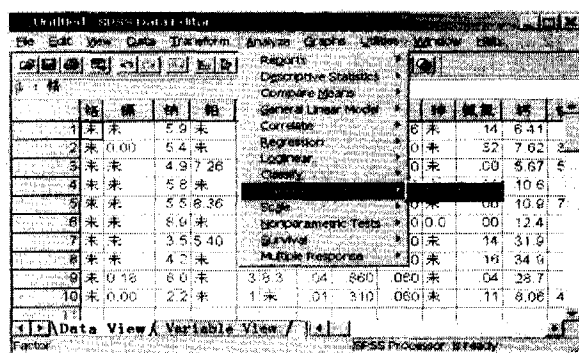


图2 对数据表中的数据进行 Factor Analysis 分析

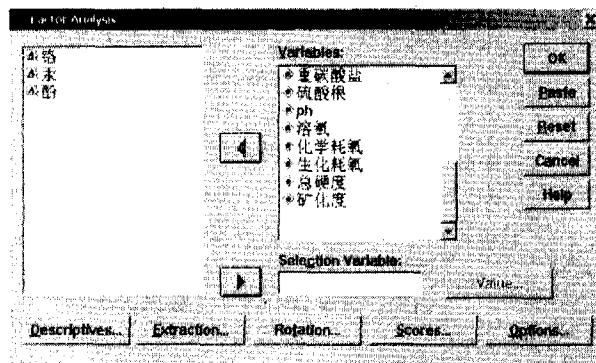


图3 Factor Analysis 过程

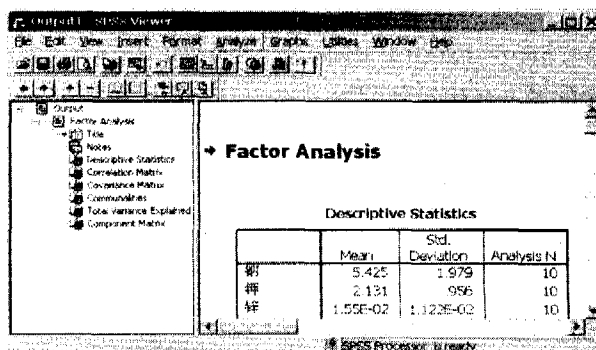


图4 分析结果

组分做为样本总体^[8](包含 18 个水化学组分变量)。然后从样本中抽取若干子样组合(采用九种不同化学组分的组合)。其中九种组合为:a)全部组分的组合;b)阴离子、阳离子、氨氮和 pH 的组合;c)阳离子、氨氮和 pH 的组合;d)阳离子和 pH 的组合;e)阳离子;f)阳离子(去掉 Zn);g)阴离子、氨氮、pH 的组合;h)阴离子和 pH 的组合;i)阴离子、氨氮。抽取不同子样组合即不同水化学组分组合的目的是为了寻找包含地下水变异信息最多的子样组合,从而用最少的组分的组合作为评价地下水变异的指标体系。

对子样组合进行 KMO (Kaiser - Meyer - Olkin Measure of Sampling Adequacy) 和球形 Bartlett 检验,选择通过检验并获得 KMO 值较高的子

样作为评价地下水变异的主要影响因子。结合因子分析结果,提取主要影响因子——主要的水化学组分的组合。

②对提取的地下水化学组分做主成分分析。并以主成分方差累积贡献率为权数,用主成分计算年地下水变异的主成分加权综合评价。

③调用 SPSS (Statistics Package for Social Science) 程序的 Factor Analysis 过程进行分析。

3.2 Factor Analysis 的过程与组分的提取

3.2.1 因子分析与组分提取过程

I) 水化学组分因子的背景值与基本统计量

II) 在总体样本中抽取不同子样,构成不同的子样组合

在总体样本中抽取不同子样,构成九种不同的子样组合:a)全部组分的组合;b)阴离子、阳离子、氨氮和 pH 的组合;c)阳离子、氨氮和 pH 的组合;d)阳离子和 pH 的组合;e)阳离子;f)阳离子(去掉 Zn);g)阴离子、氨氮、pH 的组合;h)阴离子和 pH 的组合;i)阴离子、氨氮。抽取不同子样组合即不同水化学组分组合的目的是为了寻找包含地下水变异信息最多的子样组合,从而用最少的组分的组合作为评价地下水变异的指标体系。

III) 九子样组合的因子分析的 KMO 与 Bartlett 检验

KMO (Kaiser - Meyer - Olkin Measure of Sampling Adequacy) 检验是用于比较观测相关系数值与偏相关系数值的一个指标,表示变量间的偏相关是否很小,其值愈接近 1,表明对这些组分变量进行因子分析的效果愈好。

球形 Bartlett 检验是将总样本随机分解成若干子样本,然后分别计算各子样本的方差,经子样本方差分析判断是否来至于方差相同的总体,即方差的同质性问题——子样本方差与总样本方差无显著性差异。采用来自于同一总体样本的九个不同子样的抽取方法,通过对九子样组合对地下水变异影响的方差分析,比较各种组合在反映地下水变异的准确度方面有无差异。

取 1% 显著性水平,计算子样抽样分布服从自由度 $\alpha - 1$ 的卡方检验量,将计算值与卡方分布的临界值比较,如果计算值大于 1% 显著性水平下临界值,则在该水平下拒绝方差具有同质性的假设,即方差具有齐性。

若计算值小于 1% 显著性水平下临界值时,则

在该水平下接受方差同质性假设,说明子样本抽取的组合方式对反映地下水变异无显著性差别,说明 10 年来地下水化学组分的变化,可由其中抽取的若干小样本来反映,即方差具有同质性。对方差齐性的子样本,反映抽取的子样本方差与总样本方差不一致,即子样本与总样本有显著性差异,方差齐性的子样本不能反映总样本的变化。

因此在进行分析时,对具有方差齐性的子样本首先剔除,选择具有方差同质性的样本,从经济角度选择子样本容量小但不能太小的水化学组分为评价指标体系。

IV) K、Na、Ca、Mg、Fe、Mn 组分变量的主成分分析

V) 地下水年变异水平的综合评价地下水年变异水平的定量分析采用如下方法:将原始指标通过恰当的线性组合成主成分;以各主成分的各自方差贡献率为权数的线性组合求出年综合评价。

从 KMO 值看,阴离子与 pH 的组合的 KMO 值最大,按 KMO 值的大小来讲,取 g) 种组合作为因子分析的评价指标体系是比较适宜,但考虑到现代测试手段的限制以及经济上的原因,选取阳离子组合比较简便、经济,故选择阳离子组合中 KMO 值最大的一种组合作为评价指标体系是适宜的。因此选择 f) 组合——即 K、Na、Ca、Mg、Fe、Mn 组合方式做为反映地下水变异的评价指标体系。

3.2.2 K、Na、Ca、Mg、Fe、Mn 组分变量的主成分分析

A 区,K、Na、Ca、Mg 组分的系数为正,系数也相差不大,表明对地下水变异程度起着同向的、相当的作用,而 Fe、Mn 的系数为负,表明对地下水变异起着负向的、相当的作用。

从第一主成分分析发现,K、Na 作用是正向的,Ca、Mg 与 Fe、Mn 作用则互为反相:当 Ca、Mg 是正向时,Fe、Mn 则是负向,反之亦然。在第一主成分中出现 Ca、Mg 和 Fe、Mn 两个组分组合变量负荷(系数)高正值和高负值时,则应将其视为两个组分组合看待,反映出两个组分组合间存在着负相关关系。

主要是 K^+ 、 Na^+ 离子为易溶盐成分,当环境变异时是正向增长,而 Ca、Mg 为难溶盐主要离子成分,Fe、Mn 在土中主要是胶结状态,环境变异时,环境变异对其溶解有偏向性与优先选择权。

从第二、第三主成分来看,K、Na、Ca、Mg、Fe、Mn 除在第一主成分出现一次且有较高的因子负荷外,

在第二、三主成分上也出现不同的较高的因子负荷,说明地下水变异不是一次环境变异的结果,反映三区域曾经历过多次环境变异的结果。

另外,从主成分方差分析,3 个主成分的方差的大小反映出 3 个主成分已能完全刻画原始指标信息(主成分方差越大,说明包含各组分的原始信息越多)。

3.2.3 地下水年变异水平的综合评价

图 5 为年地下水变异水平趋势分析图。通过定量分析可以看出,年地下水变异水平随城市建设时间增长,地下水变异逐渐增加。

4 结 语

1) SPSS 功能强大、应用广泛、易学易用,对需要大量运算的统计问题,运用 SPSS 软件可以快速得到演算结果,彻底消除重复而单调的计算过程,研究人员可更多地专注于统计模型的建立与统计结果的评判上。

2) 建立了城市建设引发城市区域的地下水缓慢变异的因子分析模型,调用 SPSS 的 Factor Analysis 过程,对城市区域地下水场变异起重要的组分进行定量提取,以起重要作用、组分最少的 K、Na、

Ca、Mg、Fe、Mn 组合来作为评价地下水变异的指标体系。对南宁市区地下水中特征点进行初步定量分析表明,城市区域地下水年变异水平是城市的建设与发展作用的结果,随时间增长变异水平呈增长之势。

3) 文中仅对南宁市区某地下水特征点,运用十年来城市区域的地下水数据进行了因子分析模型的建立与计算,从历史背景下研究城市区域地下水缓慢变异程度及趋势。尽管分析模型不能完全代表“面”上的地下水变异情况,但为下一步从“面”上分析地下水变异情况提供了有益的方法与分析基础。

[参考文献]

- [1] 林杰斌,刘明德编著. SPSS10.0 与统计模式建构[M]. 北京:科学出版社,2002,1.
- [2] 三味工作室编写. SPSS V10.0 for Windows 实用基础教程[M]. 北京:希望电子出版社,2001.
- [3] 胡键颖,冯泰编著. 实用统计学[M]. 北京:北京大学出版社,1997.
- [4] 张崇甫,陈述云,胡希铃编著. 统计分析方法及其应用[M]. 重庆:重庆大学出版社,1995,9.
- [5] 吴恒,张信贵,易念平. 城市建设与地质环境的关系及研究要点[J]. 地质科学,1999,34(1),162~170.
- [6] 吴恒,张信贵,易念平,等. 水土作用与土体细观结构研究[J]. 岩石力学与工程学报,2000,19(2),199~204.
- [7] 吴恒,张信贵,易念平,等. 城市环境下的水土作用对土强度的影响[J]. 岩土力学,1999,20(4),25~30.
- [8] 吴恒,张信贵,易念平,等. 南宁市地下水化学场分区特征及其主要影响因素[J]. 广西科学,2000,7(1),30~34.
- [9] 广西壮族自治区南宁市地质条件系列图集. 广西壮族自治区地质矿产局,1988.

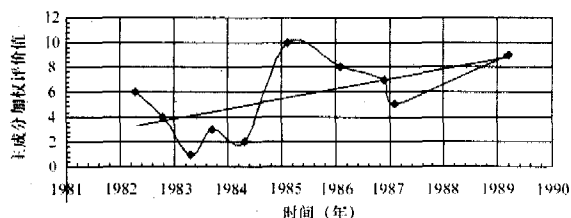


图 5 年地下水变异水平趋势分析图

CITY GROUNDWATER VARIATION FACTOR ANALYSIS PROCESS BASE ON SPSS

HE Xing-jiang¹, ZHANG Xin-gui², YI Nian-ping²

(1. Chengdu University of Technology, Chengdu 610059;

2. College of Civil Engineering, Guangxi University, Nanning 530004)

Abstract: For a heavy statistics operation, applying SPSS software can get results quickly. Clearing repeating and dull calculation process, one can concentrate on establishing statistics model. This paper sets up groundwater slow variation model, and applies SPSS program Factor Analysis process to draw some important factors of water variation, considering K, Na, Ca, Mg, Fe, Mn's least combinations as evaluating index system of water changes. Analyzing the character points, A area (concentrate-exploiting groundwater city proper), B area (scatter-exploiting groundwater city proper), and C area (scatter-exploiting groundwater suburban) in the Nanning City, show that the groundwater variation level is the result of city construction and development. Investigating slow change extent and groundwater trend from history background has instruction significance for understanding groundwater movement discipline. Though different area has different groundwater environment change and value index system, the means and methods can offer some helps.

Key words: SPSS software, groundwater environment variation, water chemistry factor, evaluating index, factor analysis