

· 合理用药 ·

doi:10.3969/j.issn.1006-4931.2018.13.024

中医方剂研究中数据统计分析方法应用解析

梁欢, 兰鸿[△]

(湖北省十堰市太和医院·湖北医药学院附属医院, 湖北 十堰 442000)

摘要:目的 探讨数据统计分析方法在中医方剂中的应用。方法 收集、整理近年来有关统计方法在方剂“证-药-方”中的应用文献,并进行归纳、分析。结果 整理了方剂在聚类分析、主成分分析、因子分析和关联规则分析中的应用研究,且从证、药、方3个方面阐述中医理论在方剂中的指导作用。结论 运用数据统计分析方法对中医方剂证候特点、本质等核心问题进行分析,可发现隐藏在其中的规律或数据间的关系,为临床辨证论治、遣药组方提供了依据。

关键词:聚类分析;主成分分析;因子分析;关联规则;方剂

中图分类号:R289;R288

文献标识码:A

文章编号:1006-4931(2018)13-0075-03

Application and Analysis of Data Statistical Analysis Method in Traditional Chinese Medicine Prescription

Liang Huan, Lan Hong

(Shiyan Taihe Hospital Affiliated to Hubei University of Medicine, Shiyan, Hubei, China 442000)

Abstract: Objective To investigate the data statistical analysis method in the application of the traditional Chinese medicine(TCM) prescription. **Methods** The application researches of statistical methods in syndrome, drugs and prescriptions were collected and summarized in recent years. **Results** The application of prescriptions in cluster analysis, principal component analysis, factor analysis and association rule analysis was summarized. It elaborated that the TCM theory guided the prescription from the syndrome, drugs and prescriptions three aspects. **Conclusion** The data statistical analysis method is used to analyze the core issues of TCM syndrome characteristics, essence and so on, in order to find the relationship between the hidden rules and the data. The results can provide the basis for clinical syndrome differentiation and construction of prescription.

Key words: cluster analysis; principal component analysis; factor analysis; association rules; prescription

第一作者:梁欢(1988-),女,硕士研究生,药师,主要从事门诊中药房调剂工作,(电子信箱)lianghuan198810@163.com。

[△]通信作者:兰鸿(1976-),男,硕士研究生,副主任药师,主要从事药房管理工作,(电子信箱)lh76117611@163.com。

- 1-13.
- [2] Ashour HM, El - Sharif A. Microbial spectrum and antibiotic susceptibility profile of gram - positive aerobic bacteria isolated from cancer patients[J]. Journal of Clinical Oncology, 2007, 25(25): 5763 - 5769.
- [3] 尹剑辉. 临床抗生素不合理用药影响因素分析及干预对策探究[J]. 医药前沿, 2017, 7(11): 126 - 127.
- [4] Pletzer D, Hancock RE. Antibiofilm peptides: potential as Broad - Spectrum agents[J]. J Bacteriol, 2016, 198(19): 2572 - 2578.
- [5] 蒲聪颖, 李珊, 张媛媛. 肿瘤耐药性产生的分子机制研究进展[J]. 四川生理科学杂志, 2015, 37(2): 89 - 93.
- [6] 张美莲, 张海茂, 赵白云, 等. 恶性肿瘤患者肺部感染的病原菌分布及耐药性分析[J]. 中华医院感染学杂志, 2013, 23(21): 15354 - 15356.
- [7] 刘艳萍, 张俊华, 郭本成. 中晚期肿瘤患者肺部感染的病原菌分布[J]. 中华医院感染学杂志, 2014, 24(10): 2383 - 2385.
- [8] 郭彩霞, 石红梅, 李飞, 等. 晚期肺癌患者肺部感染病原菌与药敏分析[J]. 中华医院感染学杂志, 2016, 26(16): 3702 - 3704.
- [9] 杨妮珍, 胡少春, 童先丽, 等. 肿瘤患者痰标本分离出革兰阴性杆菌的耐药性分析[J]. 检验医学与临床, 2016, 13(Z2): 82 - 84.
- [10] 杨爱平, 雷和月, 刘军, 等. 恶性肿瘤患者侵袭性真菌感染菌种分布及耐药性分析[J]. 中国微生物学杂志, 2014, 26(1): 79 - 81.
- [11] 王怡云, 姚伯程, 胡红莲. 肿瘤患者医院感染病原菌分布及其耐药性研究[J]. 国际检验医学杂志, 2013, 34(4): 485 - 487.
- [12] 叶彩玲, 江奕楠, 潘扬勋, 等. 肿瘤患者医院感染革兰阳性球菌的分布与耐药性分析[J]. 中华医院感染学杂志, 2016, 26(11): 2522 - 2525.
- [13] 陆娟, 王忠明, 顾建兴. 2014年江苏省无锡市肿瘤医院临床分离病原菌的分布及耐药分析[J]. 中国现代药物应用, 2015, 9(11): 37 - 39.
- [14] 罗庆, 王伟, 白权菊. 贵州省肿瘤医院2012年病原菌分布及耐药分析[J]. 实验与检验医学, 2015, 33(3): 370 - 373.
- [15] Jia B, Raphenya AR, Alcock B, et al. CARD 2017: expansion and model - centric curation of the comprehensive antibiotic resistance database[J]. Nucleic Acids Res, 2017, 45(D1): D566 - D573.

(收稿日期:2017-12-28;修回日期:2018-03-20)

方剂是中医理、法、方、药的重要组成部分,本质表现为方与方、方与药、药与药、药与剂量,以及方药与证候交叉错综的多维关联与对应,具有复杂性和非线性动态变化的特性,具有以“方-药-证”为核心的多维结构。由于方剂中蕴涵海量数据,缺乏科学的信息处理方法,大量方剂长期处于分散流失状态,难以进行系统、深入的研究,以至传统方药理论难有突破和发展^[1]。数据统计分析是同时分析和处理多组变量,从整体把握事件的特征和发生规律的统计分析方法,其核心内容是总体参数估计的修正和统计推断,具体表现出来就是各类统计方法,如聚类分析、主成分分析、因子分析、关联规则分析等^[2]。本研究中首次从聚类分析、主成分分析、因子分析、关联规则4个方面,从证、药、方3个层次进行数据统计分析,以发现在中医方剂中的应用规律,现介绍如下。

1 聚类分析

聚类分析是一种探索性分析,能从样本数据出发,自动进行分类,不必事先给出分类标准。聚类方法可将一组或一群症状自然地聚为一族或一群,从而揭示一组症状与证候之间的密切关系。聚类分析按其分类目的,可分为两大类:一是R型聚类,又称为指标聚类,其目的是将指标降维,从而选择有代表性的指标;二是Q型聚类,又称为样品聚类,其目的是找出样品的共性。在应用方剂研究时,多选用Q型聚类。根据聚类样品的不同进行分类。

对证候的聚类:中医在辨证论治的过程中,四诊信息量非常大,对同一病种的辨证分型亦不相同,为了明确各分类证候的属性,可在大规模流行病学调查基础上采用聚类分析,对收集到的症状的属性进行归纳和分类,然后依据专业知识找出比较公认的中医证型。阎玥等^[3]采用聚类分析方法探讨感冒后咳嗽中医证候的分布特点,得到感冒后咳嗽常见证候类别有风邪恋肺、肺气虚、阴津亏虚、脾虚、肺热、痰浊、风寒袭表、肺卫亏虚、肝火上炎和肾虚。孙文军等^[4]运用聚类分析研究广泛性焦虑症的中医证候学规律,总结了肝郁化火、肾精亏虚、心胆气虚、心血亏虚、气血两虚、痰湿困脾、阴阳两虚、气郁血瘀8个证候类型,为最终制订《广泛性焦虑症的中医证候诊断标准》提供了依据。巩璇^[5]根据症状、体征、舌脉等四诊信息进行聚类分析,归纳出2型糖尿病中医证候类型为脾肾气虚证、肝肾阴虚证、阴虚热盛证、气阴两虚夹痰湿证、阳虚血瘀证5个证候类型。

对方的聚类:对治疗某种病的大量方剂进行研究时,从众多方剂的几大类型推测该病治疗的基本方,可

基于方中药物的属性对方聚类。吕晓东等^[6]在总结特发性肺纤维化的中医治法方药规律时,利用聚类分析对107首中药复方进行治法分析和方药配伍研究,结果7类中有6类复方主要体现了益气的治法,有5类主要体现了活血的治法,有4类主要体现了养阴止咳平喘的治法,可见益气、活血、养阴、止咳平喘是众医家的最主要共识。杨雯晴等^[7]对筛选出的315首治疗头痛的方剂进行分析,提取出核心药物组合26个,并在此基础上进行聚类,得到新处方13个。吴嘉瑞等^[8]采用中医传承辅助系统平台软件对筛选出的颜正华教授治疗风湿痹证102首处方进行关联规则和复杂系统熵聚类的用药规律研究,挖掘出22个核心组合和11首新处方。

对药物的聚类:如想了解治疗某病的药物种类时,可对类方中所有药物聚类,可基于药物功效或性味归经。若分析某一个方剂时,可对该方所包含的药物聚类,一般根据药物的性味归经聚类。刘嘉辉等^[9]对国医大师治疗慢性乙型肝炎的常用药物进行聚类分析为5类,I类和II类都是益气健脾的药物为主,III类为利水祛湿之品,IV类为活血祛瘀之品,V类则为清热解毒之品。

2 主成分分析和因子分析

主成分分析与因子分析都是寻求少数几个变量(或因子)来综合反映全部变量(或因子)的大部分信息,变量虽然较原始变量减少,但所包含的信息量可占原始信息的85%以上,且这些新变量彼此互不相关,消除了多重共线性^[10]。陈光兰等^[11]采用主成分分析法对治疗幽门螺杆菌感染的常用12种中药方剂进行定量评价,结果发现主要中药为清热解毒药、温里药、活血化瘀药、补益药和止血药的组合。王春晓等^[12]采用主成分分析法分析颈椎病中医证型分布特点,对98个常见症状进行分析,结果其基本病机是以“邪实”为主,主要表现为气滞、湿热、寒湿、痰湿,并见血瘀之邪,提示主成分分析用于中医证型的分类研究具有一定科学性。孙大志等^[13]分析胃癌的中医证候构成及其分布规律,因子分析结果显示为证候或病机,胃气上逆、痰湿内蕴、脾气亏虚、瘀血内阻、脾胃虚寒、肝气犯胃、气血两虚、内热和脾胃气滞。金香兰等^[14]通过因子分析得出了高血压病的主要证候要素是气虚、火、血瘀、痰,次要证候要素是阴虚、血虚;主要涉及脏腑为肝、肾两脏,次要脏腑为心。

3 关联规则分析

关联规则是通过数据挖掘某事物与其他事物之间的联系,寻求相互依存性及关联程度。通过分析不同类型方剂数据集的关联可得到中药与中药间的相关性,从而发现目标方剂集的核心药对或药组。其应用主要

有3类。

药对的发掘与研究:利用关联规则的频繁项集探寻方剂中的高频药组。陈芳等^[15]探讨四物汤方中药物间的配伍关系,采用“以药类方”的方法,从中医方剂数据库中检索得到含有四物汤其中任意3味药物的方剂842首,建立数据库,采用Apriori算法,研究类方中药物间配伍关系。结果发现,作为四物汤类方核心的当归、地黄、川芎、白芍4味药中,当归与其他药物的联系最紧密,表现为四物汤类方的中心环节,“当归-地黄”药对则为方中的中心药对。

以药聚方:以药为类,搜集方剂建立数据库,然后从中发掘出用药规律。吴嘉瑞等^[16]应用关联规则Apriori算法研究含甘草处方中单味药物频次、药物组合频次、关联规则与核心药物组合等。结果,含甘草处方常用于治疗咳嗽、胃痛、感冒等病证,高频次药物包括甘草、陈皮、赤芍、茯苓、丹参等,高频次药物组合包括“甘草,赤芍”“甘草,连翘”“甘草,陈皮”“甘草,金银花”等,置信度 ≥ 0.9 的关联规则包括“连翘 $\rightarrow$ 甘草”“金银花 $\rightarrow$ 甘草”“连翘,金银花 $\rightarrow$ 甘草”“桔梗 $\rightarrow$ 甘草”等,治疗咳嗽的关联规则核心药物组合与止咳散有一定相似性,治疗胃痛的关联规则核心药物组合与香砂养胃丸有一定相似性。

以病(证)统方:以病(证)为类,搜集方剂建立数据库,然后从中发掘出用药规律。艾军等^[17]以大样本、多中心的480例小儿肺炎病例为数据源,进行症状关联规则运算,结果小儿肺炎的基本病机为化热、气郁与生痰。风热犯肺证与痰热闭肺证与热相关的病机之间具有较高的关联度;与郁相关的病机之间,痰热闭肺证比风热犯肺证关联度高;与痰相关的病机之间,痰热闭肺证比风热犯肺证关联度高。吴嘉瑞等^[18]采用关联规则Apriori算法研究补气类中成药组方规律,结果处方中药物组合置信度较高的有“川芎、熟地黄 $\rightarrow$ 当归”“白芍、熟地黄 $\rightarrow$ 当归”“白芍,茯苓 $\rightarrow$ 当归”“川芎,白芍 $\rightarrow$ 当归”。

方剂是中医辨证论治的完整体现,充满非线性思维,但数据统计分析方法能以线性和非线性方式解析数据,且能进行高层次的知识整合,又善处理模糊和非量化数据,对集成方剂文献信息、评价方剂效应以及中医药知识发现具有重要指导作用。由于各种统计分析方法自身的特点,在具体的应用中又都存在着各自的局限,通过几种方法联合应用,可获得更加广阔的数据和规律,在实际应用中要灵活掌握。

参考文献:

- [1] 郭锡琴,方向明. 数据挖掘技术在方剂学配伍研究中的应用探讨[J]. 安徽中医学院学报,2009,28(6):10-12.
- [2] 朱小梅. 多元统计分析方法在宏观经济分析中的应用[J]. 赤峰学院学报(自然科学版),2013,29(20):5.
- [3] 阎 玥,王辛秋,史 琦,等. 基于聚类分析的感冒后咳嗽中医证候研究[J]. 世界中西医结合杂志,2015,10(6):741-743.
- [4] 孙文军,曲 森,唐启盛. 广泛性焦虑症中医证候的聚类研究[J]. 中华中医药杂志,2014,29(7):2298-2300.
- [5] 巩 璇. 基于聚类分析的2型糖尿病证候规律的研究[D]. 大连:大连医科大学,2014.
- [6] 吕晓东,刘 创,庞立健,等. 基于数据挖掘的特发性肺纤维化中医治法方药研究[J]. 中华中医药杂志,2015,31(7):2509-2513.
- [7] 杨雯晴,李运伦,丁学义,等. 基于数据挖掘的治疗头痛方剂药物规律分析[J]. 中国中医基础医学杂志,2014,20(2):231-234.
- [8] 吴嘉瑞,郭位先,张 冰,等. 基于关联规则和复杂系统熵聚类的颜正华教授治疗风湿痹证用药规律研究[J]. 中华中医药杂志,2014,29(3):696-699.
- [9] 刘嘉辉,韦志辉,何洁茹,等. 基于数据挖掘的国医大师治疗慢性乙型肝炎用药规律研究[J]. 中国中医药科技,2016,23(2):248-249.
- [10] 王 芳. 主成分分析与因子分析的异同比较及应用[J]. 统计教育,2003(5):14.
- [11] 陈光兰,陈成龙,汪望月,等. 基于聚类分析和主成分分析探讨治疗幽门螺杆菌感染的中药方剂的用药规律[J]. 医学研究杂志,2013,42(6):108-111.
- [12] 王春晓,谢兴文,李 宁. 聚类分析与主成分分析在颈椎病中医证型规范化研究中的应用[J]. 中国组织工程研究与临床康复,2011,15(43):8083-8088.
- [13] 孙大志,修丽娟,施 俊,等. 胃癌中医证候多元分析[J]. 中国中医药信息杂志,2016,23(10):16-20.
- [14] 金香兰,张允岭,陈志刚,等. 运用因子分析探讨原发性高血压病证候要素[J]. 北京中医药大学学报,2011,34(2):131-134.
- [15] 陈 芳,朱 敏,尚尔鑫,等. 基于Apriori算法的四物汤类方组方特点分析[J]. 中华中医药杂志,2011,26(2):249-252.
- [16] 吴嘉瑞,郭位先,张晓滕,等. 基于数据挖掘的国医大师颜正华含甘草处方用药规律研究[J]. 中华中医药杂志,2016,31(6):2313-2315.
- [17] 艾 军,汪受传,戴 铭,等. 基于症状关联规则的小儿肺炎中医证候病机学研究[J]. 北京中医药大学学报,2016,39(2):140-143.
- [18] 吴嘉瑞,金燕萍,张 冰,等. 基于关联规则与熵聚类的补气类中成药组方规律研究[J]. 世界中医药,2016,11(1):172-175.

(收稿日期:2018-02-08;修回日期:2018-03-06)

本栏目

国药控股云南有限公司

协办