

DOI 10.19656/j.cnki.1002-2406.20230705

药 学 专 题

基于SPSS、R语言和C#的全蝎中重金属及有害元素检测 results 分析

梁瑞强¹, 颜子然², 曹进¹, 孙姗姗^{1✉}, 林瑞超³

(1. 中国食品药品检定研究院国家市场监管重点实验室(食品质量与安全), 北京 100050;

2. 蚂蚁云创数字科技(北京)有限公司, 北京 100089;

3. 北京中医药大学中药学院中药品质评价北京市重点实验室, 北京 102488)

【摘要】目的:探索并建立中药全蝎产地与其重金属及有害元素含量的关系,并运用计算机语言实现样本的产地推测功能。方法:将30批全蝎经过微波消解后使用电感耦合等离子体质谱仪分析铅、镉、砷、汞、铜元素的含量,检测结果经SPSS进行数据分析,根据数据分析结果运用R语言和C#开发具有全蝎样本产地识别功能的应用程序。结果:30批全蝎样品铅、镉、砷、汞、铜检测结果的中位值分别为0.208、1.710、0.567、0.010 4、170.0 mg/kg;经SPSS数据分析发现检测结果在一定程度上可以起到区分产地的效果;通过R语言结合C#编写的程序可以通过主成分参数对未知产地的全蝎进行产地推测。结论:全蝎中重金属及有害元素的含量与其产地存在关联性;使用R语言和C#所编写的产地推测软件具有进一步的开发价值。

【关键词】全蝎;元素;重金属;SPSS;R;C#

【引用格式】

梁瑞强,颜子然,曹进,等.基于SPSS、R语言和C#的全蝎中重金属及有害元素检测 results 分析[J].中医药信息,2023,40(7):30-35.

LIANG R Q, YAN Z R, CAO J, et al. Analysis of heavy metals and harmful elements in scorpion based on SPSS, R language and C#[J]. Information on TCM, 2023, 40(7):30-35.

全蝎为东亚钳蝎(*Buthus martensii* Karsch)的干燥体,临床常用于治疗痉挛抽搐、疮疡肿毒、癰疽结核、风湿顽痹和顽固性偏头痛等疾病^[1]。该药材主要分布于长江以北地区,河南和山东地区是全蝎的古代原产地,江南地区(四川)属于人为引入^[2-5]。生长环境、生活习性、人工养殖方法以及后期饮片炮制加工工艺等因素都可能会对全蝎引入重金属和有害元素。重金属会对人体的正常生理功能和新陈代谢带来显著的伤害,加之临床上部分患者存在长期用药的情况,因此即使中药饮片中重金属及有害元素残留通常在痕量水平,也有可能对人体中蓄积而危害身体健康。2020年版《中华人民共和

国药典》全蝎项下没有对重金属及有害元素的限量和检测方法做出相关规定^[6]。因此有必要选择目前重点关注的铅(Pb)、镉(Cd)、砷(As)、汞(Hg)、铜(Cu)5种元素对全蝎药材样品进行测定。在以往的研究中,有研究者通过灰色关联度分析法对全蝎水分、总灰分、酸不溶性灰分及浸出物含量进行分析,发现不同产地的全蝎药材存在质量差异,说明产地可能对全蝎的质量产生影响^[7]。因此本研究也尝试分析并建立重金属及有害元素与全蝎产地的关系。

目前,对于重金属及有害元素的分析方法主要有原子吸收光谱法(AAS)、原子发射光谱法(AES)、原子

基金项目:国家中药标准化项目(ZYBZH-C-QIN-45)

第一作者简介:梁瑞强(1990-),男,主管中药师,主要研究方向:中药标准及标准物质研究。

✉通信作者简介:孙姗姗(1984-),女,副研究员,主要研究方向:食品安全检测、食品的真实性检验及过敏原的鉴定。

荧光谱法(AFS)、紫外-可见分光光度法(UV-Vis)、电感耦合等离子体发射光谱法(ICP-OES)和电感耦合等离子体质谱法(ICP-MS)。其中ICP-MS适合多元素同时分析且灵敏度高^[8]。本研究使用ICP-MS法对全蝎进行元素检测,与《中华人民共和国药典》规定的重金属及有害元素分析方法相符。

SPSS作为常用的统计分析软件,具有操作简便、功能强大、模块组合灵活等特点,可用于主成分分析、聚类分析、探索性分析和描述性统计等多种统计学分析^[9]。R语言是一种可视化编程语言,可以用于统计分析,使用灵活,开发简便^[10]。C#是专门为.NET应用开发的语言,适合用于桌面和Web应用程序开发,其语法简单,可跨平台,也是完全支持面向对象编程的语言^[11]。本文在全蝎重金属及有害元素检测结果的基础上,综合运用SPSS、R语言和C#,分析了全蝎产地和检测结果的联系,并开发具有可视化界面的产地推测软件。

1 材料与方法

1.1 仪器

电感耦合等离子体质谱仪(iCAP Q ICP-MS, Thermo 公司);分析天平(MSE224S-.CE, sartorius 公司);多功能消解仪(MARS, CEM 公司);赶酸炉(G-400, 上海屹尧仪器科技发展有限公司);酸清洗器(AC200M1343, Amerlab 公司);超纯水器(GWB SERISES, 普析通用公司)。

1.2 试剂和试液

硝酸(优级纯,国药集团化学试剂有限公司);30%过氧化氢(分析纯,国药集团化学试剂有限公司);1 000 μg/mL 铅元素标准溶液(GBW08619/17 032, 中国计量科学研究院);1 000 μg/mL 镉元素标准溶液(GBW08612/17 022, 中国计量科学研究院);1 000 μg/mL 砷元素标准溶液(GBW08611/17 032, 中国计量科学研究院);1 000 μg/mL 汞元素标准溶液(GBW08617/17 032, 中国计量科学研究院);铜(Cu, 1000 μg/mL, H2-CU03021 CGCU1);1 000 μg/mL 铟元素标准溶液(K2-IN653104, Inorganic Ventures 公司);1 000 μg/mL 锗元素标准溶液(GSB 04-1728-2004, 国家有色金属及电子材料分析测试中心);1 000 μg/mL 铋元素标准溶液(GSB 04-1719-2004, 国家有色金属及电子材料分析测试中心);实验用水由超纯水器制备。

重蒸硝酸:将优级纯硝酸加入酸清洗器,使用酸重蒸模式将优级纯硝酸重蒸并冷凝收集,即得。

5% 硝酸溶液:量取重蒸硝酸 5 mL,加水至 100 mL,混匀。

1.3 供试品

30批全蝎药材采购于中药材公司、药店或由制药公司提供,产地分布在河北、辽宁、山西、河南、山东 5 个省份(每个省份 6 批)。经中国食品药品检定研究院康帅副研究员鉴定为钳蝎科动物东亚钳蝎(*Buthus martensii* Karsch)的干燥体。

1.4 仪器条件

RF 发生器功率:1200 W;等离子体流速:13 L/min;载气流速:0.85 mL/min;辅助气流速:0.7 L/min;采样深度:8.0 mm;截取锥孔径:0.5 mm;雾化器:微流同心雾化器;雾化室温度:3℃;模式:KED;碰撞气:高纯氦气(>99.999%),流速:4.2 mL/min;蠕动泵转速:30 r/min;驻留时间:0.02 s。

1.5 标准溶液配制

分别精密移取适量 Pb、Cd、As、Hg、Cu 元素标准溶液,以 5% 硝酸溶液稀释,配制成含 Pb、Cd、As 元素的浓度分别为 1、5、10、30、50 μg/L, Cu 元素浓度分别为 10、50、100、300、500 μg/L 的四元素混合标准系列溶液。Hg 元素标准系列溶液单独配制,浓度为 0.2、0.4、0.6、0.8、1.0 μg/L。

1.6 供试品溶液制备

将全蝎供试品粉碎,准确称取供试品粉末 0.2 g,置于清洗好的聚四氟乙烯消解罐内。加入重蒸硝酸 5.0 mL,30% 过氧化氢 1.0 mL,缓慢振摇消解罐使消解液充分浸没供试品粉末,在通风橱中放置过夜预消解。旋紧罐盖,将消解罐放入微波消解仪中,消解供试品。消解仪的温度-时间程序见表 1。

表 1 微波消解仪的温度-时间程序

STAGE	POWER/W	RAMP	CONTROL/℃	HOLD
1	1600	05:00	120	08:00
2	1600	05:00	150	10:00
3	1600	15:00	180	20:00

消解完毕后,将消解罐取出冷却,开罐。将消解罐放入赶酸炉中 120℃加热至酸液量 2 mL 以下,用超纯水将剩余溶液转移定容至 25 mL 塑料容量瓶中。可根据样品中各元素的含量进行再稀释。

不加样品,按上述步骤同时做空白试验。

1.7 测定步骤

用质谱调谐液调谐仪器各项指标,使仪器灵敏度、氧化物、双电荷、分辨率等指标达到要求。同时引入标准工作溶液、供试品溶液与内标溶液[锗(Ge)、铟(In)、铋(Bi)各 10 μg/L]进行检测。

2 结果与分析

2.1 标准溶液测定结果

将配制得到的各浓度标准溶液进样分析,以每一浓度测得的3次读数与内标读数比值的平均值为纵坐标(Y),标准溶液浓度($\mu\text{g/L}$)为横坐标(X),由仪器自带软件进行线性拟合,建立标准曲线。各元素的线性方程见表2。可见各待测元素在所配制的浓度范围内线性关系良好。

表2 Pb、Cd、As、Hg、Cu元素的回归方程

元素	回归方程	r	线性范围/ ($\mu\text{g/L}$)
Pb	$Y=138\,909.082\,1X-151\,218.613\,6$	0.999 2	1~50
Cd	$Y=5\,131.824\,7X-439.143\,1$	0.999 9	1~50
As	$Y=841.472X-128.934\,4$	0.999 9	1~50
Hg	$Y=35\,989.308\,5X-3\,630.271\,5$	0.998 0	0.2~1.0
Cu	$Y=17\,416.465\,9X+24\,762.553\,3$	0.999 4	10~500

2.2 供试品检测结果

将30批全蝎的重金属及有害元素检测结果进行统计分析,见表3。从检测结果整体看,含量最高的是Cu元素,最低的是Hg元素,Cu元素在30批供试品中含量的差异最大。将不同来源的元素测定结果取均值

后发现辽宁产全蝎的Pb和Cu残留相对较高,山东产全蝎的Cd平均残留量在各产地全蝎之间最高,除辽宁、河南产全蝎外,其他来源全蝎的Hg残留量批间差异较大,见图1。观察采集的全蝎性状可以发现,30批全蝎药材中有11批表面可见盐粒或少量盐霜,例如山西产的6批全蝎有3批表面有盐粒,而另外3批表面未见盐粒或盐霜,考虑到全蝎的加工工艺可能导致无机盐残留,加之先前的研究发现供试品的总灰分与5种元素之间的多元线性回归分析结果显示具有较高的相关性($r=0.746$, $P=0.00$),因此相同产地的实验结果的偏差也可能与炮制加工的方法差异有关。以上这些结果提示对于全蝎的重金属及有害元素的污染与产地、加工方法之间的相关性值得做进一步研究。

表3 重金属及有害元素检测结果(mg/kg)

项目	Pb	Cd	As	Hg	Cu
最小值	0.092	1.190	0.235	0.001 7	94.2
最大值	0.485	4.490	1.310	0.084 1	307.0
均值	0.237	1.990	0.600	0.017 7	177.0
中位值	0.208	1.710	0.567	0.010 4	170.0
标准偏差	0.106	0.731	0.272	0.018 5	46.6
P_{95}	0.465	3.350	0.999	0.052 1	249.0

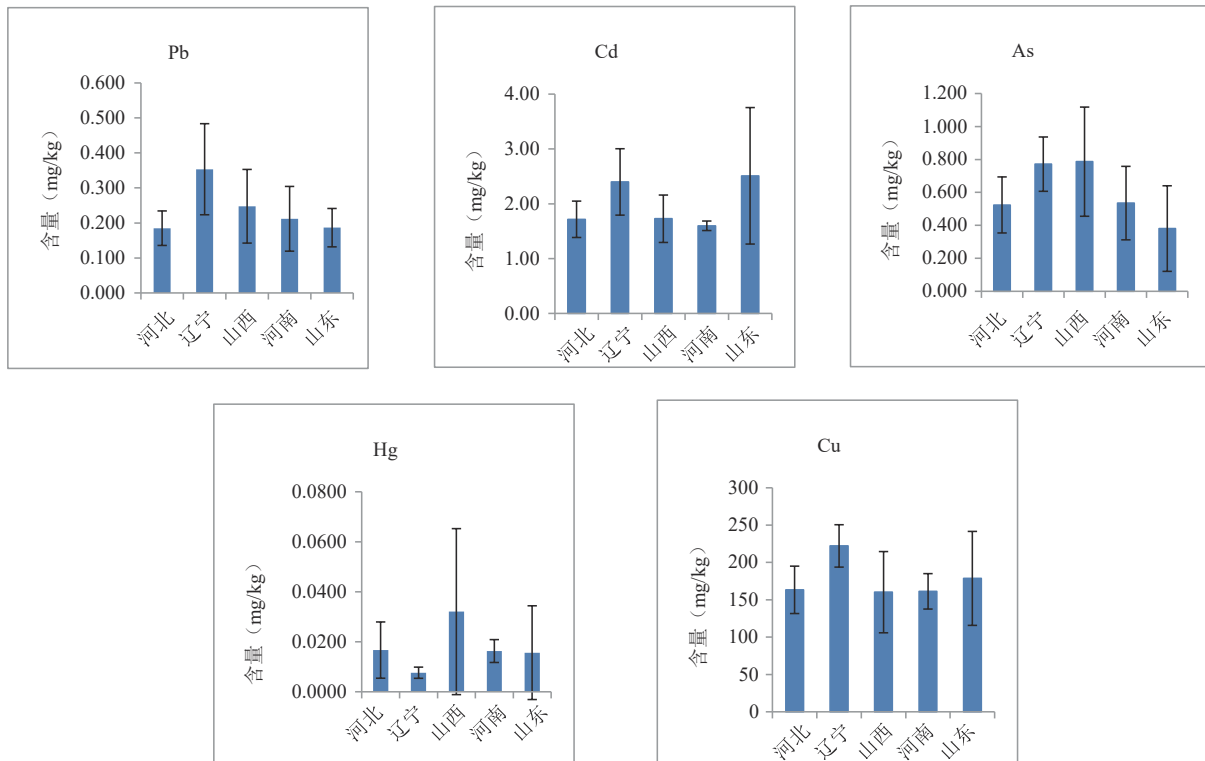


图1 不同产地全蝎样品的元素含量均值

2.3 SPSS数据分析

对30批全蝎供试品每批制备2份平行样并检测,共得到60条检测结果数据。使用SPSS软件对这些数据进行主成分分析,显示 Bartlett 的球形度检验显著性

P 值为 0.000 002,说明这些数据适合做因子分析($P<0.05$)。得到成分的特征值和贡献率见表4。通常认为特征值大于1的成分属于主成分。成分1、2、3的累积贡献率为 81.205%(>80%),说明可以通过分析这

三个成分达到检验目的^[12]。因此选择 1、2 和 3 三个成分作为主成分做进一步分析。

表 4 成分的特征值和贡献率表

成分	特征值λ	贡献率/%	累积贡献率/%
1	1.851	37.016	37.016
2	1.132	22.648	59.664
3	1.077	21.541	81.205
4	0.618	12.366	93.571
5	0.321	6.429	100.000

主成分分析的特征向量矩阵 U 与因子载荷矩阵 A、特征值 λ 的数学关系为： $U_i = A_i / \sqrt{\lambda_i}$ 。因此根据 SPSS 软件给出的因子载荷矩阵和表 4 中的特征值，计算得到特征向量结果见表 5。

表 5 因子载荷矩阵 A 和特征向量矩阵 U 表

元素	A ₁	A ₂	A ₃	U ₁	U ₂	U ₃
Pb	0.731	-0.209	0.399	0.537	-0.196	0.384
Cd	0.176	0.613	0.736	0.129	0.576	0.709
As	0.824	-0.086	-0.381	0.606	-0.081	-0.367
Hg	0.002	0.827	-0.460	0.001	0.777	-0.443
Cu	0.779	0.146	-0.136	0.573	0.137	-0.131

主成分的表达式为特征向量值分别乘以供试品对应被测成分结果的标准化值，然后求和。根据上述计

算结果，得到三个主成分 Y₁、Y₂ 和 Y₃ 的表达式。

$$Y_1 = 0.537 \times Z_{Pb} + 0.129 \times Z_{Cd} + 0.606 \times Z_{As} + 0.001 \times Z_{Hg} + 0.573 \times Z_{Cu} \quad (1)$$

$$Y_2 = -0.196 \times Z_{Pb} + 0.576 \times Z_{Cd} - 0.081 \times Z_{As} + 0.777 \times Z_{Hg} + 0.137 \times Z_{Cu} \quad (2)$$

$$Y_3 = 0.384 \times Z_{Pb} + 0.709 \times Z_{Cd} - 0.367 \times Z_{As} - 0.443 \times Z_{Hg} - 0.131 \times Z_{Cu} \quad (3)$$

根据公式计算 30 批供试品的 60 份主成分数值，部分结果见表 6。主成分分析可以起到降维的效果，从而综合考量各单因素产生的作用。将供试品中的 5 个元素的检测结果，降维成 3 个影响因素（三维），发现不同来源的全蝎供试品的主成分数值存在差异，如河北产全蝎的三个主成分大部分为负值，而辽宁、山西产全蝎主成分正值较多，同时来源全蝎的主成分的绝对值差异也较大。因此，可以尝试通过主成分数值对不同来源的全蝎进行产地推测。

将全部 60 份主成分数据进行聚类分析，得到树状图，见图 2。可见发生聚类的全蝎供试品大部分属于同一产地。如辽宁 1、辽宁 3 和辽宁 5 的供试品聚在一类，山东 4 和山东 6 聚在一类。因此相同产地全蝎主成分值更加接近，进一步说明可以利用主成分分析得到的数据进行产地推测。

表 6 供试品主成分部分数值表

供试品	Y ₁	Y ₂	Y ₃	供试品	Y ₁	Y ₂	Y ₃
河北 1-1	-0.250 1	-0.371 6	-0.158 2	山西 2-2	-0.328 3	-1.401 8	1.169 1
河北 1-2	-0.384 3	-0.168 0	-0.170 8	山西 3-1	-0.374 6	-1.616 6	-0.112 9
河北 2-1	-2.439 5	-0.504 8	-0.635 9	山西 3-2	-0.661 5	-1.524 8	-0.214 5
河北 2-2	-2.492 5	-0.538 3	-0.620 1	河南 1-1	-1.520 6	-0.162 5	-0.558 2
河北 3-1	-0.261 2	-0.264 9	-0.190 3	河南 1-2	-1.569 2	-0.015 3	-0.591 1
河北 3-2	-0.257 5	-0.236 0	0.002 8	河南 2-1	-0.000 1	-0.373 3	-0.763 5
辽宁 1-1	2.575 6	-0.021 8	1.548 5	河南 2-2	0.015 0	-0.122 8	-0.755 9
辽宁 1-2	2.923 3	0.240 7	1.574 7	河南 3-1	-0.301 4	-0.123 8	-0.618 4
辽宁 2-1	0.121 2	-0.760 4	-0.193 5	河南 3-2	-0.319 4	-0.085 5	-0.521 7
辽宁 2-2	0.189 1	-0.788 1	-0.272 8	山东 1-1	-1.451 1	-0.474 5	0.010 6
辽宁 3-1	2.919 9	-0.099 8	1.449 7	山东 1-2	-1.481 0	-0.541 3	-0.061 4
辽宁 3-2	3.247 2	0.259 3	1.083 8	山东 2-1	-1.077 6	-0.519 4	0.628 4
山西 1-1	1.961 4	2.282 3	-2.491 6	山东 2-2	-1.021 9	-0.446 5	0.800 2
山西 1-2	1.498 6	2.629 1	-2.483 2	山东 3-1	2.217 4	-0.588 6	-1.226 4
山西 2-1	-0.152 5	-1.346 4	0.986 0	山东 3-2	1.701 9	-0.624 2	-1.280 6

2.4 基于 R 语言和 C# 的产地推测

为了实现将主成分数值用于产地推测的功能，本研究使用 R 语言编写了产地推测程序，并利用 C# 编写了直观的界面化应用软件与 R 语言的代码衔接。开发环境为 Windows 7 和 Windows 10，使用的程序开发工具分别为 R (4.2.1) 和 Visual Studio (Community 2022)。

将供试品的主成分数值的数据导入 Excel 表格，表格中第 1 列为供试品来源，第 2~4 列分别对应主成分 Y₁、Y₂、Y₃ 的数值，直观起见将列标题分别命名为“主成分 1”“主成分 2”“主成分 3”。将 Excel 表另存为 .csv 格式，作为数据文件。

在用 R 语言编写的代码中，调用 .csv 主成分数值数据文件后，加载了数据包“dplyr”。由于原数据文件

中每个来源平行测定了两次,如“河北1-1”和河北“1-2”均为河北1来源,因此使用 substr() 函数截取“来源”列数据的前三个字符作为分组依据命名为“来源_新”列,再用 datamean = group_by(raw_data, 来源_新) %>% summarize_each(funs(mean)) 命令对主成分数值进行分组平均后赋值给新数据框“datamean”。之后用“for 循环”结合欧氏距离计算公式计算未知样品主成分数据与“datamean”中每一行主成分数据的距离,代码见图3。

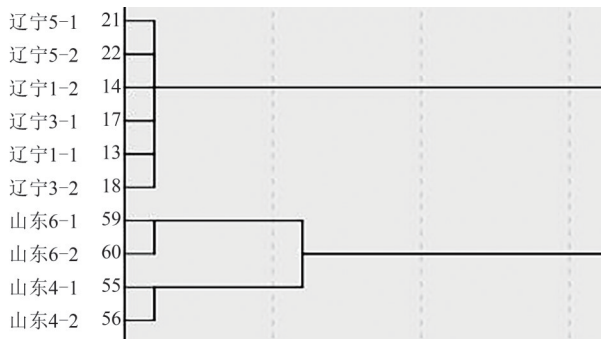


图2 全蝎供试品聚类树状图(部分)

```
find_best_c <- function(tv1, tv2, tv3){  
  min_value <- 9999  
  best_c <- ""  
  for(i in 1:length(datamean$来源_新)){  
    c <- datamean[i, ]$来源_新  
    value1 <- datamean[i, ]$主成分1  
    value2 <- datamean[i, ]$主成分2  
    value3 <- datamean[i, ]$主成分3  
  
    distance <- sqrt((value1-tv1) * (value1-tv1) + (value2-tv2) * (value2-tv2) +  
                    (value3-tv3) * (value3-tv3))  
    if(distance < min_value){  
      min_value <- distance  
      best_c <- c  
    }  
  }  
  cat('R-output 最近的来源是:', best_c, '最小的距离是:', min_value)  
}
```

图3 “for 循环”距离推算命令(截取)

tv1、tv2、tv3 为实际使用时输入的未知样品的主成分数值,先将 min_value 赋值为 9 999,进入循环后第一次计算的 distance 值一定小于 min_value 的初始值,循环继续,之后 distance 小于 min_value 时继续赋新值, distance 大于 min_value 时则忽略,直至达到 distance 的最小值,通过 cat() 命令输出结果。若得到的距离越近说明相似度越高,从而推测出最接近的全蝎来源。

上述功能的实现需要在 R 语言的开发环境中进行,为了方便使用,本研究在 Visual Studio 开发工具中利用 C# 语言“所见即所得”的特点制作了可视化界面,过输入供试品的主成分数值即可衔接到距离计算程序中并将输出的结果在界面上显示,例如输入 0.1011、0.62、-1.0516 三个数据后点击“计算”按钮,即可得到产地推测结果为“河北6”,已经计算得到的欧氏距离,见图4。同时为了增加界面的可用性,本研究在界面中添加了选择 R 语言版本、R 代码源文件、数据文件的功能见图5,以便在 R 开发环境版本、代码内容修改或

数据文件变化后仍能使用此界面。



图4 使用界面

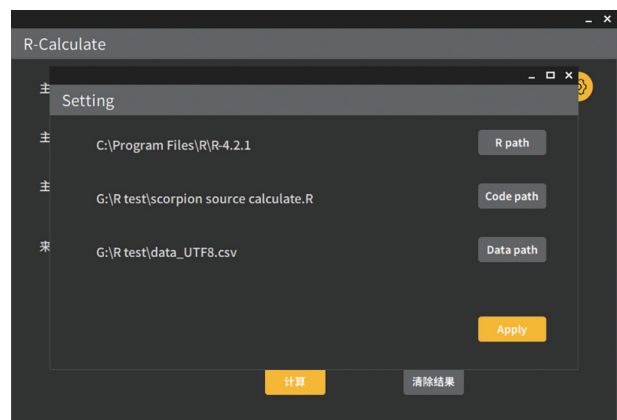


图5 条件设置界面

3 结论

本研究通过电感耦合等离子体质谱法(ICP-MS)测定了30批全蝎药材中铅、镉、砷、汞、铜的含量,从检测结果看不同来源的供试品在重金属及有害元素含量上存在差异。通过主成分分析将检测结果降维成3个影响因素,再经过聚类分析发现主成分数值可以用作产地推测。使用 R 语言结合 C# 开发出了通过计算和比较欧氏距离推测全蝎来源的软件并实现了界面化操作,可以实现推测未知来源全蝎的产地。

【参考文献】

- [1] 高学敏. 中药学[M]. 北京:中国中医药出版社,2002:481-483.
- [2] 杨祎辰,马存德,王二欢,等. 中药全蝎的本草考证及现代研究进展[J]. 安徽农学通报,2018,24(8):105-106.
- [3] 苏颂. 本草图经[M]. 合肥:安徽科学技术出版社,1994:525.
- [4] 李时珍. 李时珍医学全书·本草纲目[M]. 2版. 北京:中国中医药出版社,2015:1297-1298.
- [5] 唐慎微. 证类本草[M]. 北京:华夏出版社,1993:542.
- [6] 国家药典委员会. 中华人民共和国药典:一部[M]. 北京:中国医药科技出版社,2020:149-150.
- [7] 王聪聪,李峰,张美玲,等. 山东不同产地全蝎药材质量灰色关联度分析法综合评价[J]. 辽宁中医药大学学报,2016,18(10):42-45.

[8] 徐勤科,孙莺,张东,等. 超级微波消解-电感耦合等离子体质谱法测定化妆品中4种微量元素[J]. 分析测试技术与仪器,2021,27(2):144-148.

[9] 姜忠尉. 统计分析软件SPSS的特点和应用分析[J]. 中国证券期货,2012,20(4):291.

[10] 刘钰,余卓芮,刘岱宁. R语言可视化的优势及其在空间统计教学中的应用[J]. 高教论坛,2020,247(5):30-33.

[11] 张跟兄,富木仁. 浅析《C#语言》的优势及教学过程中的收获[J]. 科技与企业,2015,24(18):159.

[12] 徐琳,何洪源,侯超,等. 基于Python/MATLAB的常见茶叶中重金属元素的ICP-MS检验研究[J]. 茶叶通讯,2020,47(1):107-112.

(收稿日期:2023-01-04)

Analysis of Heavy Metals and Harmful Elements in Scorpion Based on SPSS,
R Language and C#

LIANG Ruiqiang¹, YAN Ziran², CAO Jin¹, SUN Shanshan^{1✉}, LIN Ruichao³

(1. Key Laboratory of Food Quality and Safety for State Market Regulation, National Institute for Food and Drug Control, Beijing 100050, China; 2. Ant Cloud Digital Technology (Beijing) Co., Ltd, Beijing 100089, China; 3. Beijing Key Lab for Quality Evaluation of Chinese Materia Medica, School of Chinese Materia Medica, Beijing University of Chinese Medicine, Beijing 102488, China)

【Abstract】 Objective: To explore and establish the relationship between the origin of scorpion and the content of heavy metals and harmful elements, and to realize the function of inferring the origin of samples by using computer language. Methods: After 30 batches of scorpions were digested by microwave, the contents of lead, cadmium, arsenic, mercury and copper were analyzed by inductively coupled plasma mass spectrometer. The test results were analyzed by SPSS. According to the data analysis results, an application program with the function of identifying the origin of scorpion samples was developed by using R language and C#. Results: The median values of lead, cadmium, arsenic, mercury and copper in 30 batches of scorpion samples were 0.208 mg/kg, 1.710 mg/kg, 0.567 mg/kg, 0.010 4 mg/kg, 170.0 mg/kg, respectively. SPSS data analysis found that the detection results could play a role in distinguishing the origin to a certain extent. The program compiled by R language and C# could speculate the origin of the whole scorpion with unknown origin through the principal component parameters. Conclusion: The content of heavy metals and harmful elements in scorpion is related to its origin. The software for predicting the origin has a further development value.

【Key words】 Scorpio; Element; Heavy metal; SPSS; R; C#