第 18 卷 第 2 期
1999 年 5 月

人 类 学 学 报
ACTA ANTHROPOLOGICA SINICA

Vol. 18, No. 2
May, 1999

用聚类分析方法对土家族和瑶族人发中 7 种元素综合指标的比较研究

杨若明* 王振英

(中央民族大学生物化学系、预科部, 北京 100081)

摘 要

对我国土家族、瑶族青年人体头发中钙、镁、铁、锰、铜、锌、铬等 7 种金属元素进行了测定。用聚类方法对元素含量的综合指标进行了考查、比较研究。结果表明, 土家族与瑶族青年人体头发中这 7 种元素的综合指标存在明显的差别, 民族之间区别较大。

关键词 头发, 元素, 聚类分析, 土家族, 瑶族

1 前 言

为了研究同一国家不同民族间人体头发中微量及常量元素含量是否有显著差异以进行民族之间的比较研究, 本文作者曾首次对我国 17 个民族人发中 7 种元素含量进行了测定, 并分别就每一种元素的含量对 17 个民族进行了方差分析(多重比较 T 法)(马萨特(比)等, 1990), 对不同地区同一民族人发中元素含量进行了单因素方差分析(王振英等, 1995a、b), 均得到了有价值的结果。为了进一步探讨多种元素含量在整体水平上的差别情况, 本文作者又首次用聚类方法对不同民族之间头发中多种元素指标进行了比较研究的尝试(杨若明等, 1994)。本文对土家族和瑶族人体头发中钙、镁、铁、锰、铜、锌、铬等 7 种元素含量应用聚类分析方法进行综合评价, 以进行民族之间的比较研究。聚类分析结果表明, 在七种元素含量的整体水平上, 达到了民族之间的较好区分。

2 实验部分

2.1 发样来源及采集

来自民族主要聚居区的中央民族大学一年级土家族和瑶族新生, 年令 17 至 21 岁, 身体健康。采样时用不锈钢剪刀采枕部头发 2—3 克。

2.2 发样的预处理及测定

发样用洗涤剂浸泡充分后, 用蒸馏水, 二次去离水冲洗充分后, 烘干, 称重, 硝酸分

收稿日期: 1997-12-12
本工作获 96 年国家民委科技进步二等奖, 本文为部分工作内容。

解，定容，以备使用。

用美国 Psseries 1000 型 ICP-AES 仪，对发样中 Ca、Mg、Fe、Mn、Cu、Zn 及 Cr 等 7 种元素含量进行测定，所用试剂均为光谱纯级，水为二次去离子水。

3 结果和讨论

对 20 个样本中 7 种元素测定的结果列于表中 (因篇幅限制，表略)。其中，土家族 1—10 克，瑶族 11—20 克。

3.1 数据规范化

实验测得的原始数据含 20 个样本，每个样本有 7 个指标分别与测定的 7 种元素对应。由于不同元素在含量的数量级上有差别。例如，Ca 含量在 450—1700ppm 范围，而 Cr 的含量仅在 1—10ppm 数量级。为了分类，可以认为样品中 Ca 含量 100ppm 数量级的差别与 Cr 含量 0.1ppm 的数量级的差别同等重要。但是，如果用原始数据计算相似性量度，Cr 含量之间的差别与 Ca 含量之间差别比较，将显得微不足道。因此，为了消除变量变化总幅度的影响，在进行聚类分析之前，应对原始数据进行变换以达到数据规范化。

实验测得的数据构成原始数据阵 X (N×P)，N 为样本数，每个样本有 P 个指标。为了使 P 个指标等权，对原始数据阵的每列进行标准化处理。本文采用 Z-变换作标准处理 (杨若明等，1999)，又称自身规范化。是用 S_j 为单位将 X_{ij} 表示为 Z_{ij}，计算公式为:

$$Z_{ij} = \frac{X_{ij} - \bar{X}_{\cdot j}}{S_j} \quad (i = 1, 2, \cdots, N;\ j = 1, 2, \cdots, P) \tag{1}$$

式中: $\bar{X}_{\cdot j} = \frac{1}{N}\sum_{i=1}^{N} X_{ij}$ $\quad S_j = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(X_{ij} - \bar{X}_{\cdot j})^2}$

经 Z-变换后的标准化值构成一个新的数据阵 Z (N×P)，见表 1 所示。

表 1 标准化值

N \ P	1	2	3	4	5	6	7
1	-0.129	0.213	-0.833	-0.808	-0.761	-0.710	0.969
2	0.403	0.057	-0.870	-0.797	0.184	0.774	0.938
3	-0.208	-0.805	-0.632	-0.716	0.000	0.677	1.273
4	2.511	0.582	-0.928	-0.394	-1.595	0.645	0.738
5	0.546	-0.484	-0.942	-0.856	-0.675	-0.645	0.796
6	0.359	-0.446	-0.842	-0.606	-1.160	-0.194	0.665
7	1.166	0.213	0.026	-0.859	-0.564	1.677	0.896
8	-1.375	-1.061	-0.693	-0.339	-0.534	-1.355	0.681
9	-0.825	-1.008	-0.658	0.818	0.307	-1.064	0.692
10	-0.009	-0.719	-0.912	-0.882	-1.178	0.935	0.631
11	-0.261	-0.071	0.382	2.430	2.147	-0.710	0.269
12	0.300	0.582	0.004	0.388	0.736	1.645	-0.888
13	-0.799	-0.270	0.180	0.636	0.982	0.903	-1.580
14	-1.080	-0.750	-0.215	-0.635	-0.699	-0.258	-0.523
15	1.941	2.457	2.684	0.653	0.491	0	0.038
16	-0.498	-0.629	1.338	0.884	1.472	-0.322	-0.377
17	0.385	0.753	1.162	1.860	0.429	-0.452	-0.573
18	-0.758	2.457	-0.096	-0.392	0.859	-1.097	-1.015
19	-0.920	-0.369	0.373	-1.115	0.613	1.322	-1.935
20	-0.755	-0.815	1.487	0.810	-1.215	-1.619	-1.727

3.2　相似性矩阵的获得

为达到按民族不同划分类型的目的，本文对样品聚类。分别选用相关系数及欧氏距离作为相似性量度。用 r_{kl} 表示第 k 个样品与第 1 个样品的相关系数，计算方法如下:

$$r_{kl}=\frac{\sum_{j=1}^{P}(Z_{kj}-\overline{Z_{k\cdot}})(Z_{lj}-\overline{Z_{l\cdot}})}{\sqrt{\sum_{j=1}^{P}(Z_{kj}-\overline{Z_{k\cdot}})^2\cdot\sum_{j=1}^{P}(Z_{lj}-\overline{Z_{l\cdot}})^2}} \tag{2}$$

式中

$$\overline{Z_{k\cdot}}=\frac{1}{P}\sum_{j=1}^{P}Z_{kj},\quad \overline{Z_{l\cdot}}=\frac{1}{P}\sum_{j=1}^{P}Z_{lj}$$

Z_{kj}，Z_{lj} 为矩阵 Z (N × P) 中第 i= k 及 i= l 行的变量。

计算出数据阵的全部 r_{kl} 值之后，形成一个由相关系数组成的实对称矩阵 R，称为样品相关系数矩阵，亦称相关矩阵，相似性矩阵。其中，主对角元素均为 1，即 r_{ii}= 1 (i= 1, 2, …, N)，r_{kl} 的值在- 1 与+ 1 之间，根据 r_{kl} 接近于 1 的程度，可决定样品 k 与 l 之间的相似程度，从而对样品进行分类。

$$R=\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1N} \\ r_{21} & r_{22} & \cdots & r_{2N} \\ \cdots & \cdots & \cdots & \cdots \\ r_{N1} & r_{N2} & \cdots & r_{NN} \end{bmatrix}$$

用 d_{kl} 表示第 k 个与第 l 个样品之间的欧氏距离，计算公式为:

$$d_{kl}=\sqrt{\sum_{j=1}^{P}(Z_{kj}-Z_{lj})^2} \tag{3}$$

式中 Z_{kj}，Z_{lj} 为数据阵 Z (N × P) 第 i= k 及 i= l 行的变量。

计算出数据阵的全部 d_{kl} 值之后，构成一个由距离组成的实对称矩阵 D，称为样品的距离矩阵，也称为相似性矩阵。其中，主对角线上的元素均为 0，d_{kl} 均为正值。d_{kl} 值越小，表示差别越小，以此来对样品分类。

以 7 号和 10 号样品为例，相关系数 $r_{7,10}$ 与距离 $d_{7,10}$ 的详细计算过程见表 2 及表 3。

表 2　7 号样品与 10 号样品的相关系数计算

i \ j	1	2	3	4	5	6	7	$\overline{Z_i}$	$\sum_{j=1}^{7}(Z_{ij}-\overline{Z_i}\cdot)^2$
样品 7	1.166	0.213	0.026	- 0.859	- 0.564	1.677	0.896	0.365	
样品 10	- 0.009	- 0.719	- 0.912	- 0.882	- 1.178	0.935	0.631	- 0.305	
$Z_{7,j}-\overline{Z_7}\cdot$	0.801	- 0.152	- 0.339	- 1.224	- 0.929	1.312	0.531		5.144
$Z_{10,j}-\overline{Z_{10}}\cdot$	0.296	- 0.414	- 0.607	- 0.577	- 0.837	1.240	0.936		4.136

$$r_{7,10}=\frac{\sum_{j=1}^{7}(Z_{7,J}-\overline{Z_7}\cdot)(Z_{10,J}-\overline{Z_{10}}\cdot)}{\sqrt{5.114\times4.136}}=0.899$$

表 3 7 号与 10 号样品的距离计算

i \ j	1	2	3	4	5	6	7	$\sum$	$\sqrt{\sum}$
7	1.166	0.213	0.026	-0.859	-0.564	1.677	0.896		
10	-0.009	-0.719	-0.912	-0.882	-1.178	0.935	0.631		
$Z_{7,j}-Z_{10,j}$	1.157	0.932	0.938	0.023	0.614	0.742	0.265		
$(Z_{7,j}-Z_{10,j})^2$	1.157^2	0.932^2	0.938^2	0.023^2	0.614^2	0.742^2	0.265^2	4.0854	2.021

$$d_{7,10}=\sqrt{\sum (Z_{7,j}-Z_{10,j})^2}=\sqrt{4.0854}=2.021$$

求得数据阵的全部相关系数，并列成表 4 的相关系数矩阵。求得数据阵的全部欧氏距离后，列成表 5 的相似性（距离）矩阵D。

3.3 聚类谱系图的获得

3.3.1 相关系数法

依据表 3 的相关系数矩阵R，找出其中最接近于 1 的数值，如 $r_{7,10}=0.899$，$r_{2,3}=0.829$，$r_{5,6}=0.864$，$r_{17,20}=0.866$，$r_{12,19}=0.814$ 等等；按照接近于 1 的程度将样本连接起来，但连线不应出现倒转，由此而得到图 1 所示的谱系图。

3.3.2 欧氏距离法

依据表 5 的欧氏距离矩阵D，找出其中最小数值，如 $d_{5,6}=0.751$，$d_{1,5}=0.998$，$d_{2,3}=1.166$，$d_{12,13}=1.176$，$d_{6,10}=1.252$，$d_{3,10}=1.421$ 等等。按照最短距离将样本连接起来，同样连线不应出现倒转，由此而得到图 2 所示的谱系图。

3.4 聚类划分

为了达到聚类划分的目的，我们可以依次切断聚类谱系图中的最高连线。

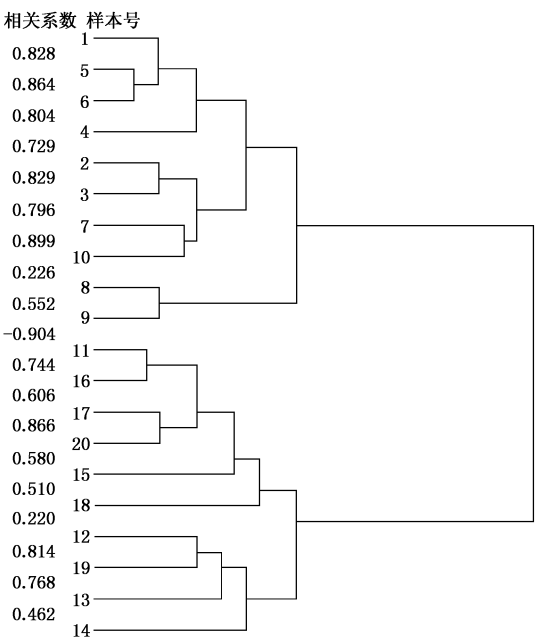

图 1 由相关系数法求得的头发样本的聚类谱系图（1—10 号土家族，11—20 号瑶族）

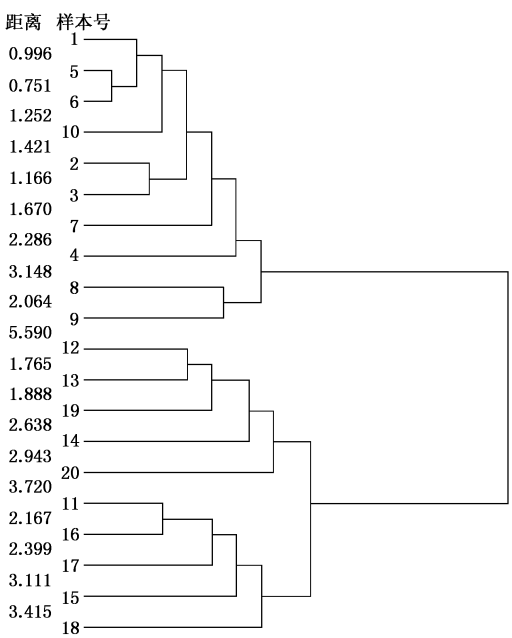

图 2 由距离法求得的头发样本的聚类谱系图（1—10 号土家族，11—20 号瑶族）

表 4　相关系数矩阵 R

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	1.0																			
2	0.601	1.0																		
3	0.495	0.829	1.0																	
4	0.510	0.511	0.214	1.0																
5	0.828	0.688	0.589	0.720	1.0															
6	0.742	0.633	0.580	0.804	0.864	1.0														
7	0.374	0.745	0.602	0.729	0.536	0.683	1.0													
8	0.196	-0.303	0.170	-0.475	0.031	-0.030	-0.344	1.0												
9	0.136	-0.119	0.226	-0.399	0.123	-0.009	-0.568	0.552	1.0											
10	0.449	0.783	0.769	0.624	0.564	0.757	0.899	-0.175	-0.204	1.0										
11	-0.381	-0.416	-0.310	-0.679	-0.389	-0.550	-0.904	0.288	0.778	-0.693	1.0									
12	-0.636	0.044	-0.185	-0.073	-0.568	-0.411	0.159	-0.762	-0.520	0.049	-0.118	1.0								
13	-0.860	-0.399	-0.355	-0.619	-0.880	-0.793	-0.380	-0.300	-0.086	-0.384	0.413	0.787	1.0							
14	-0.314	-0.153	0.239	-0.503	-0.481	-0.245	0.080	0.481	-0.045	0.194	-0.120	0.118	0.329	1.0						
15	-0.084	-0.537	-0.745	0.104	-0.176	-0.209	-0.151	0.002	-0.543	-0.497	-0.233	-0.125	-0.145	-0.186	1.0					
16	-0.684	-0.668	-0.336	-0.867	-0.637	-0.765	-0.745	0.461	0.412	-0.705	0.744	-0.004	0.589	0.315	0.025	1.0				
17	-0.533	-0.948	-0.886	-0.377	-0.595	-0.563	-0.791	0.116	0.188	-0.805	0.606	-0.013	0.363	-0.113	0.488	0.545	1.0			
18	0.028	-0.275	-0.589	-0.297	-0.350	-0.489	-0.461	-0.176	-0.221	-0.633	0.152	0.136	0.156	-0.248	0.510	0.052	0.325	1.0		
19	-0.699	-0.064	-0.120	-0.388	-0.673	-0.592	0.111	-0.352	-0.553	-0.073	-0.150	0.814	0.786	0.462	0.026	0.312	-0.080	0.141	1.0	
20	-0.532	-0.959	-0.733	-0.363	-0.555	-0.460	-0.569	0.388	0.054	-0.623	0.353	-0.157	0.243	0.220	0.580	0.586	0.866	0.061	-0.001	1.0

表 5 欧氏距离矩阵 DR

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	0																			
2	1.856	0																		
3	1.920	1.166	0																	
4	3.142	2.847	3.513	0																
5	0.998	1..759	1.794	2.778	0															
6	1.110	1.764	1.726	2.576	0.751	0														
7	2.856	1.670	2.198	2.286	2.687	2.411	0													
8	2.505	3.433	2.844	5.056	2.739	2.748	4.247	0												
9	2.448	2.958	2.498	4.358	2.488	2.586	4.129	2.064	0											
10	1.972	1.660	1.421	2.924	1.772	1.252	2.032	3.243	3.148	0										
11	4.583	4.352	4.360	5.819	4.650	4.754	5.148	4.214	2.927	5.198	0									
12	3.700	2.618	3.155	3.930	3.701	3.484	2.705	4.427	3.799	3.273	3.732	0								
13	3.991	3.421	3.598	5.104	4.016	3.846	3.934	3.831	3.281	3.725	3.312	1.765	0							
14	2.163	3.051	2.352	4.295	2.276	2.046	3.443	1.943	2.362	2.163	4.451	3.254	2.715	0						
15	5.173	4.992	5.509	4.821	5.326	5.301	4.433	5.623	5.696	5.761	4.795	4.133	4.989	5.592	0					
16	3.920	3.684	3.804	5.373	3.928	3.971	4.243	3.013	2.700	4.248	2.167	2.967	2.193	3.133	4.310	0				
17	3.939	3.939	4.183	4.584	4.061	3.957	4.155	3.842	3.310	4.453	2.399	2.932	2.840	3.720	3.111	2.195	0			
18	3.586	3.955	4.523	5.198	4.137	4.242	4.727	4.274	4.124	4.748	4.276	3.578	3.592	3.720	4.307	3.829	3.415	0		
19	4.119	3.775	3.598	5.231	4.116	3.947	3.795	4.089	4.225	3.550	4.943	2.450	1.888	2.638	5.504	3.323	4.180	3.939	0	
20	4.213	5.002	4.782	5.590	4.252	4.070	5.276	2.954	3.971	4.612	4.556	4.539	3.640	2.943	5.299	3.294	3.219	4.443	4.148	0

在由相关系数法得到的谱系图 1 中, 我们先切断数值为- 0.904 的连线, 这时得到 (1, 5, 6, 4, 2, 3, 7, 10, 8, 9) 和 (11, 16, 17, 20, 15, 18, 12, 19, 13, 14) 两大类, 再切断 0.226 和 0.220 的两条连线 (因为这两条线的相似性属于同一级区间), 得到了 (1, 5, 6, 4, 2, 3, 7, 10)、(8, 9)、(11, 16, 17, 20, 15, 18) 和 (12, 19, 13, 14) 四类。依次再切断 0.462 的连线, 得到五类。相继切断连线可得到如下结果:

(1, 5, 6, 4, 2, 3, 7, 10, 8, 9) (11, 16, 17, 20, 15, 18, 12, 19, 13, 14) (k= 2)
(1, 5, 6, 4, 2, 3, 7, 10) (8, 9) (11, 16, 17, 20, 15, 18) (12, 19, 13, 14) (k= 4)
(1, 5, 6, 4, 2, 3, 7, 10) (8, 9) (11, 16, 17, 20, 15, 18) (12, 19, 13) (14) (k= 5)
(1, 5, 6, 4, 2, 3, 7, 10) (8, 9) (11, 16, 17, 20, 15) (18) (12, 19, 13) (14) (k= 6)
(1, 5, 6, 4, 2, 3, 7, 10) (8) (9) (11, 16, 17, 20, 15) (18) (12, 19, 13) (14) (k= 7)
(1, 5, 6, 4, 2, 3, 7, 10) (8) (9) (11, 16, 17, 20) (15) (18) (12, 19, 13) (14) (k= 8)

同样道理, 在由欧氏距离法得到的谱系图图 2 中, 先切断数值为 5.590 的最高连线, 得到两大类, 再切断 3.720 的连线, 得到三大类, 依次切断 3.415 连张, 切断 3.148 和 3.111 的连线, 相继再切断连线, 得到如下结果:

(1, 5, 6, 10, 2, 3, 7, 4, 8, 9) (12, 13, 19, 14, 20, 11, 16, 17, 15, 18) (k= 2)
(1, 5, 6, 10, 2, 3, 7, 4, 8, 9) (12, 13, 19, 14, 20) (11, 16, 17, 15, 18) (k= 3)
(1, 5, 6, 10, 2, 3, 7, 4, 8, 9) (12, 13, 19, 14, 20) (11, 16, 17, 15) (18) (k= 4)
(1, 5, 6, 10, 2, 3, 7, 4) (8, 9) (12, 13, 19, 14, 20) (11, 16, 17) (15) (18) (k= 6)
(1, 5, 6, 10, 2, 3, 7, 4) (8, 9) (12, 13, 19, 14) (20) (11, 16, 17) (15) (18) (k= 7)
(1, 5, 6, 10, 2, 3, 7, 4) (8, 9) (12, 13, 19) (14) (20) (11, 16, 17) (15) (18) (k= 8)

3.5　结论

在两种方法获得的聚类谱系图中, 切断最高连线, 都可以得到两大类, 每类中各包括 10 个样本, 分别是来自土家族及瑶族的人发样品, 1—10 号为土家族样品归为一类, 11—20 号为瑶族样品归为另一类。

这表明, 依照头发中钙、镁、铁、锰、铜、锌和铬等 7 种元素的含量, 在整体水平上比较, 土家族与瑶族之间存在着明显的差别。

3.6　讨论

两种方法聚类, k= 2 时, 每大类均包括并且只包括了一个民族, 误判率为 0。

分别选择相关系数和欧氏距离为相似性量度值, 这是由于在大多数情况下, 距离能更好地反映差别, 而相关系数则能更好地发现相似性。本文用两种方法聚类分析的结果是基本一致的。尤其是 k= 2 时的结果完全一致, 这更提高了结论的可信度。

在两个聚类谱系图中, 依次考察了 k 为 3、4、5、6、7、8 时的分类情况, 发现两种方法分类在 k 值增大时, 有所不同, 但是, 总存在着一些共同的稳健类, 例如, (1, 2, 3, 4, 5, 7, 10) (11, 16, 17) 和 (12, 13, 19), 说明这些样本在 7 种元素含量总体水平上的接近程度更大。对照样本来源, 发现这些含在稳健类中的样本, 并非一定来自同一地区, 而是含有来自不同地区的同一民族。例如, 1, 2, 3, 4, 5, 7, 10 号的土家族样本分别来自湖北、湖南、四川和贵州; 11 号和 16, 17 号是分别来自广西和湖南的瑶族; 12, 13, 19 号则包含了广东、广西和贵州 3 个省份。从另一方面看, 同是来自贵州的 6, 7, 8 (土家族) 与 15, 19 (瑶族) 的差别都很大。这应源于民族的不同。这一现象本文作者曾在单元素水

平上作过比较。结论一致。

本文研究结果更进一步表明，我国不同民族间在人发所含重要元素的综合指标上存在着明显的差别，更多研究将会表明这一结论的普遍意义。

致谢：中央民族大学科研处、学生处、各系提供发样的同学；北京市理化分析测试中心的田小青、潘品良同志；北京大学化学与分子工程学院童沈阳教授；国家环保局全浩教授。

参 考 文 献

马萨特（比）等著 1990. 聚分析法解析分析化学数据. 北京，化学工业出版社.

王振英，杨若明等 1995a. 我国不同地区回族青年人发中微量元素含量的测定和比较研究. 人类学学报 14 (2): 176—181.

王振英，杨若明等 1995b. 民族与性别对汉、回族人发中 7 种元素含量的影响. 微量元素与健康研究，12 (3): 43—46.

杨若明等 1994. 中国不同民族人发中元素的测定和比较研究. 中央民族大学学报，3 (2): 49—56.

杨若明，王振英 1999. 对苗族畲族人体头发中 7 种金属元素综合指标的聚类分析比较研究. 中央民族大学学报（自然科学版），8 (1): 58—62

THE COMPARATIVE RESEARCH OF ELEMENTAL CONCENTRATIONS IN THE HUMAN HAIR FROM THE TUJIA NATION AND YAO NATION BY THE USE OF CLUSTER ANALYSIS

Yang Ruoming Wang Zhenying

(*Central University for Nationalities, Beijing* 100081)

Abstract

In this paper, the concentrations of five trace elements and two common elments existing in the hair of young people from the Tujia Nation and Yao Nation were determined by inductively coupled plasma atomic emission spectrometer (ICP-AES). A Data matrix of concentration of elements existing in the hair from two nationalities is evaluated comprehensively by using cluster analysis with satisfactory results which shows that there is considerable difference between these notionalities

Key words Hair, Element, Cluster Analysis, Tujia Nation, Yao Nation