R型聚类法的构建和求解

1）用相关系数进行变量的相关性度量。

记变量的取值(j=1,2,...,m),则可以用两变量与的样本相关系数作为它们的相似性度量，即

=

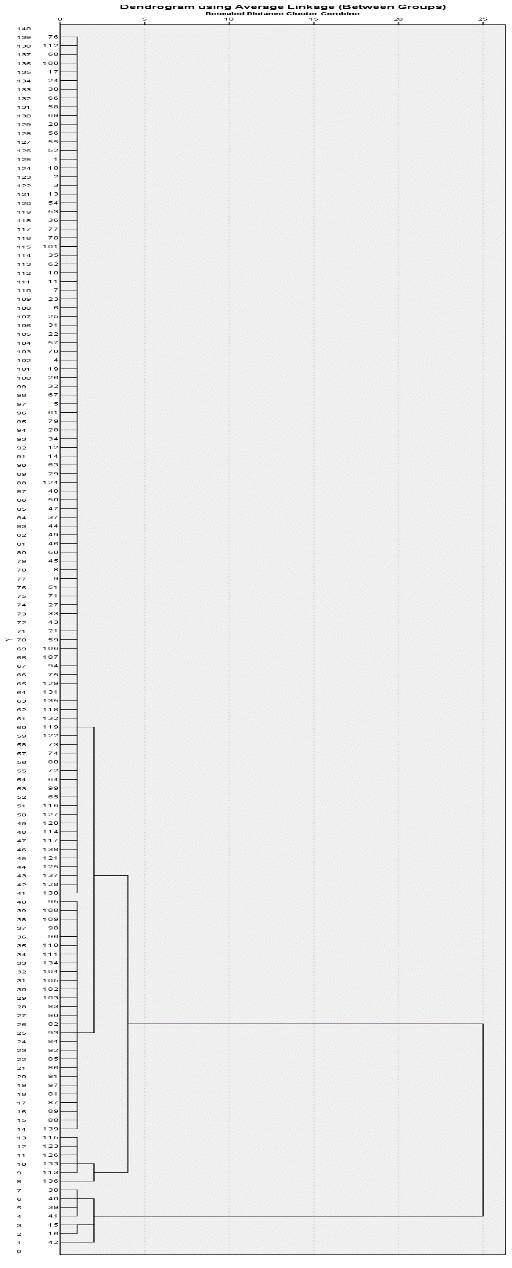
2)用最短距离法进行变量的聚类，定义两类变量的距离为

R（）=min||

式中：或=1-,这时，R（）与这两类中的相似性最大的两个变量间的相似性度量值相关。

3）现对OH州的139个人口普查社会经济数据指标进行相关分析，得到各因素之间的相关系数表。从表中可以看出139个人口普查社会经济指标中存在强烈的相关性，这也印证了进行R型聚类分析的必要性，指标过多，指标之间的相关度高，给系统分析和建模带来了很大的困难，故使用R型聚类法按照变量的相似关系把它们聚合成若干类，进而找出影响阿片类药物传播的主要因素。

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  | … |  |  |  |
|  | 1 | 0.9945 | 0.9888 | … | 0.8368 | 0.8117 | 0.6925 |
|  | 0.9945 | 1 | 0.9967 | … | 0.8297 | 0.8282 | 0.6598 |
|  | 0.9888 | 0.9967 | 1 | … | 0.8087 | 0.8179 | 0.6654 |
| … | … | … | … | … | … | … | … |
|  | 0.8368 | 0.8297 | 0.8087 | … | 1 | 0.7261 | 0.6413 |
|  | 0.8117 | 0.8282 | 0.8179 | … | 0.7261 | 1 | 0.3962 |
|  | 0.6925 | 0.6598 | 0.6654 | … | 0.6413 | 0.3962 | 1 |

4）其中,,…,分别表示139个人口普查社会经济指标，用最大系数法对139个变量进行系统聚类，分类结果如图所示。

用,,…表示聚类图上的各个节点，记={,,…}。

当距离值ƒ=2时，分为五类，即

={},={},={},

={},

={

,}

={}

其中=-（+++）.分析知第一类反映了生育率信息，第二类主要反映了Average household size和Average family size信息，第三类主要反映了Ancestry是欧洲和美国有关的信息，第四类主要反映了出生地和语言有关的信息，第五类主要反应了Year of entry和受教育程度等有关信息。这样我们就从139个指标中选取5个分析指标，分别为,,,,.