相关分析

相关分析研究的是两个变量的相关性,但你研究的两个变量必须是有关联的,如果你把历年人口总量和你历年的身高做相关性分析,分析结果会呈现显著地相关,但它没有实际的意义,因为人口总量和你的身高都是逐步增加的,从数据上来说是有一致性,但他们没有现实意义。

相关性分析和聚类分析一样，比较简单，数学建模中经常用，但是每次都只用一小步，或者只是对数据进行一下分析，根据分析的结果确定使用的方法，所以这些方法不要掌握的特别深，能会用SPSS实现就行。相关性分析可以是简单的理解为各个变量之间的相关程度。

相关性分析的SPSS操作不在演示，比较简单，各位学员操作一下就好了。

一般直接使用SPSS进行相关性分析就可以了。且一般这样认为：

0.8-1.0 极强相关

0.6-0.8 强相关

0.4-0.6 中等程度相关

0.2-0.4 弱相关

0.0-0.2 极弱相关或无相关

Sperman或kendall等级相关分析

Person相关（样本点的个数比较多）

Copula相关（比较难，金融数学，概率密度）

典型相关分析（因变量组Y1234，自变量组X1234，各自变量组相关性比较强，问哪一个因变量与哪一个自变量关系比较紧密？）

下面是一个典型相关性分析的MATLAB的程序，想看的可以看一下

例 满意度典型相关分析

某调查公司从一个大型零售公司随机调查了 784 人，测量了 5 个职业特性指标和 7个职业满意变量，有关的变量见表 1讨论两组指标之间是否相联系。

表1 指标变量表

|  |  |
| --- | --- |
| 组 | —用户反馈，—任务重要性，—任务多样性—任务特殊性  —自主性 |
| 组 | —主管满意度，—事业前景满意度，—财政满意度，—工作强度满意度，—公司地位满意度， —工作满意度，—总体满意度 |

相关系数矩阵数据见表 2

表2 相关系数矩阵数据

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 1.00 | 0.49 | 0.53 | 0.49 | 0.51 | 0.33 | 0.32 | 0.20 | 0.19 | 0.30 | 0.37 | 0.21 |
|  | 0.49 | 1.00 | 0.57 | 0.46 | 0.53 | 0.30 | 0.21 | 0.16 | 0.08 | 0.27 | 0.35 | 0.20 |
|  | 0.53 | 0.57 | 1.00 | 0.48 | 0.57 | 0.31 | 0.23 | 0.14 | 0.07 | 0.24 | 0.37 | 0.18 |
|  | 0.49 | 0.46 | 0.48 | 1.00 | 0.57 | 0.24 | 0.22 | 0.12 | 0.19 | 0.21 | 0.29 | 0.16 |
|  | 0.51 | 0.53 | 0.57 | 0.57 | 1.00 | 0.38 | 0.32 | 0.17 | 0.23 | 0.32 | 0.36 | 0.27 |
|  | 0.33 | 0.30 | 0.31 | 0.24 | 0.38 | 1.00 | 0.43 | 0.27 | 0.24 | 0.34 | 0.37 | 0.40 |
|  | 0.32 | 0.21 | 0.23 | 0.22 | 0.32 | 0.43 | 1.00 | 0.33 | 0.26 | 0.54 | 0.32 | 0.58 |
|  | 0.20 | 0.16 | 0.14 | 0.12 | 0.17 | 0.27 | 0.33 | 1.00 | 0.25 | 0.46 | 0.29 | 0.45 |
|  | 0.19 | 0.08 | 0.07 | 0.19 | 0.23 | 0.24 | 0.26 | 0.25 | 1.00 | 0.28 | 0.30 | 0.27 |
|  | 0.30 | 0.27 | 0.24 | 0.21 | 0.32 | 0.34 | 0.54 | 0.46 | 0.28 | 1.00 | 0.35 | 0.59 |
|  | 0.37 | 0.35 | 0.37 | 0.29 | 0.36 | 0.37 | 0.32 | 0.29 | 0.30 | 0.35 | 1.00 | 0.31 |
|  | 0.21 | 0.20 | 0.18 | 0.16 | 0.27 | 0.40 | 0.58 | 0.45 | 0.27 | 0.59 | 0.31 | 1.00 |

一些计算结果的数据见下面的表格。

表3 的典型变量

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
|  | 0.421704 | -0.34285 | 0.857665 | -0.78841 | 0.030843 |
|  | 0.195106 | 0.668299 | -0.44343 | -0.26913 | 0.983229 |
|  | 0.167613 | 0.853156 | 0.259213 | 0.468757 | -0.91414 |
|  | -0.02289 | -0.35607 | 0.423106 | 1.042324 | 0.524367 |
|  | 0.459656 | -0.72872 | -0.97991 | -0.16817 | -0.43924 |

表 4原始变量与本组典型变量之间的相关系数

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
|  | 0.829349 | -0.10934 | 0.48534 | -0.24687 | 0.061056 |
|  | 0.730368 | 0.436584 | -0.20014 | 0.002084 | 0.485692 |
|  | 0.753343 | 0.466088 | 0.105568 | 0.301958 | -0.33603 |
|  | 0.615952 | -0.22251 | 0.205263 | 0.661353 | 0.302609 |
|  | 0.860623 | -0.26604 | -0.38859 | 0.148424 | -0.12457 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
|  | 0.756411 | 0.044607 | 0.339474 | 0.129367 | -0.33702 |
|  | 0.643884 | 0.358163 | -0.17172 | 0.352983 | -0.33353 |
|  | 0.387242 | 0.037277 | -0.17673 | 0.53477 | 0.414847 |
|  | 0.377162 | 0.791935 | -0.00536 | -0.28865 | 0.334077 |
|  | 0.653234 | 0.108391 | 0.209182 | 0.437648 | 0.434613 |
|  | 0.803986 | -0.2416 | -0.23477 | -0.40522 | 0.196419 |
|  | 0.502422 | 0.162848 | 0.4933 | 0.188958 | 0.067761 |

表 5原始变量与对应组典型变量之间的相关系数

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
|  | 0.459216 | 0.025848 | -0.05785 | 0.017831 | 0.003497 |
|  | 0.404409 | -0.10321 | 0.023854 | -0.00015 | 0.027816 |
|  | 0.417131 | -0.11019 | -0.01258 | -0.02181 | -0.01924 |
|  | 0.341056 | 0.052602 | -0.02446 | -0.04777 | 0.01733 |
|  | 0.476532 | 0.062893 | 0.046315 | -0.01072 | -0.00713 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
|  | 0.41883 | -0.01055 | -0.04046 | -0.00934 | -0.0193 |
|  | 0.356523 | -0.08467 | 0.020466 | -0.0255 | -0.0191 |
|  | 0.214418 | -0.00881 | 0.021064 | -0.03863 | 0.023758 |
|  | 0.208837 | -0.18722 | 0.000639 | 0.020849 | 0.019133 |
|  | 0.3617 | -0.02562 | -0.02493 | -0.03161 | 0.02489 |
|  | 0.445172 | 0.057116 | 0.027981 | 0.029268 | 0.011249 |
|  | 0.278194 | -0.0385 | -0.05879 | -0.01365 | 0.003881 |

表6 典型相关系数

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1 | 2 | 3 | 4 | 5 |
| 0.5537 | 0.2364 | 0.1192 | 0.0722 | 0.0573 |

MATLAB源代码：

clc,clear

load r.txt %原始的相关系数矩阵保存在纯文本文件r.txt中

n1=5;n2=7;num=min(n1,n2);

s1=r(1:n1,1:n1); %提出X与X的相关系数

s12=r(1:n1,n1+1:end); %提出X与Y的相关系数

s21=s12'; %提出Y与X的相关系数

s2=r(n1+1:end,n1+1:end); %提出Y与Y的相关系数

m1=inv(s1)\*s12\*inv(s2)\*s21; %计算矩阵M1

m2=inv(s2)\*s21\*inv(s1)\*s12; %计算矩阵M2

[vec1,val1]=eig(m1); %求M1的特征向量和特征值

for i=1:n1

vec1(:,i)=vec1(:,i)/sqrt(vec1(:,i)'\*s1\*vec1(:,i)); %特征向量归一化，满足a's1a=1

vec1(:,i)=vec1(:,i)/sign(sum(vec1(:,i))); %特征向量乘以1或－1，保证所有分量和为正

end

val1=sqrt(diag(val1)); %计算特征值的平方根

[val1,ind1]=sort(val1,'descend'); %按照从大到小排列

a=vec1(:,ind1(1:num)) %取出X组的系数阵

dcoef1=val1(1:num) %提出典型相关系数

[vec2,val2]=eig(m2);

for i=1:n2

vec2(:,i)=vec2(:,i)/sqrt(vec2(:,i)'\*s2\*vec2(:,i)); %特征向量归一化，满足b's2b=1

vec2(:,i)=vec2(:,i)/sign(sum(vec2(:,i))); %特征向量乘以1或－1，保证所有分量和为正

end

val2=sqrt(diag(val2)); %计算特征值的平方根

[val2,ind2]=sort(val2,'descend'); %按照从大到小排列

b=vec2(:,ind2(1:num)) %取出Y组的系数阵

dcoef2=val2(1:num) %提出典型相关系数

x\_u\_r=s1\*a %x,u的相关系数

y\_v\_r=s2\*b %y,v的相关系数

x\_v\_r=s12\*b %x,v的相关系数

y\_u\_r=s21\*a %y,u的相关系数

mu=sum(x\_u\_r.^2)/n1 %x组原始变量被u\_i解释的方差比例

mv=sum(x\_v\_r.^2)/n1 %x组原始变量被v\_i解释的方差比例

nu=sum(y\_u\_r.^2)/n2 %y组原始变量被u\_i解释的方差比例

nv=sum(y\_v\_r.^2)/n2 %y组原始变量被v\_i解释的方差比例

fprintf('X组的原始变量被u1~u%d解释的比例为%f\n',num,sum(mu));

fprintf('Y组的原始变量被v1~v%d解释的比例为%f\n',num,sum(nv));

可以看出，所有五个表示职业特性的变量与有大致相同的相关系数，视为形容职业特性的指标。第一对典型变量的第二个成员与，，，有较大的相关系数，说明主要代表了主管满意度，事业前景满意度，公司地位满意度和工种满意度。而和之间的相关系数0.5537。

和解释的本组原始变量的比率：

，

X组的原始变量被到解释了100％， Y 组的原始变量被到解释了80.3％。