对应分析与典型相关分析作业

- 1. 我国山区某大型化工厂,在厂区及邻近地区挑选有代表性的8个大气取样点. 每日四次同时抽取大气样品,测定其中包含的6中气体的浓度,前后共四天,每个取样点每种气体实测16次,计算每个取样点的平均浓度.
 - (1) 试用对应分析方法对取样点及大气污染气体进行分类.
- (2) 用*R* 型因子分析方法(参数估计用主成分法)分析该组数据; 并与(1) 的结果比较之.
 - (3) 用Q型因子分析方法分析该组数据;并与(1),(2)的结果比较之.

Table 1: 大气污染数据

			<u> </u>	11 7 715 75 11	-	
	氯	硫化氢	SO_2	C_4	环氧氯	环己烷
	X_1	X_2	X_3	X_4	丙烷 X_5	X_6
1	0.056	0.084	0.031	0.038	0.0081	0.0220
2	0.049	0.055	0.100	0.110	0.0220	0.0073
3	0.038	0.130	0.079	0.170	0.0580	0.0430
4	0.034	0.095	0.058	0.160	0.2000	0.0290
5	0.084	0.066	0.029	0.320	0.0120	0.0410
6	0.064	0.072	0.100	0.210	0.0280	1.3800
7	0.048	0.089	0.062	0.260	0.0380	0.0360
8	0.069	0.087	0.027	0.050	0.0890	0.0210

2. 费希尔研究头发颜色与眼睛颜色的关系,抽查了5387 人的资料如表所示,试对其进行对应分析.

Table 2· 头发颜色与眼睛颜色的数据

眼睛颜色	头发颜色							
	金黄色	红色	褐色	深红色	黑色	合计		
蓝色	326	38	241	110	3	718		
淡蓝	688	116	584	188	4	1580		
浅蓝	343	84	909	412	26	1774		
深蓝	98	48	403	681	85	1315		
合计	1455	286	2137	1391	118	5387		

3.为了了解家庭特征与其消费模式之间的关系. 家庭消费模式变量我们取每年去餐馆就餐的频率 x_1 、每年外出看电影的频率 x_2 两个指标:家庭的特征变量我们取户主的年龄 y_1 、家庭收入 y_2 、户主受教育程度 y_3 三个指标. 这两组变量的相关系数如下表所示. 试求典型变量及典型变量间的相关系数.

Table 3: 成绩

			9. 79 . 79		
	x_1	x_2	y_1	y_2	y_3
x_1	1.00	0.80	0.26	0.67	0.34
x_2	0.80	1.00	0.33	0.59	0.34
y_1	0.26	0.33	1.00	0.37	0.21
y_2	0.67	0.59	0.37	1.00	0.35
y_3	0.34	0.34	0.21	0.35	1.00

4.在某年级44 名学生的期末考试中,有的课程用闭卷,有的课程用开卷. 试对闭卷 (X_1,X_2) 和开卷 (X_3,X_4,X_5) 两组变量进行典型相关分析.

Table 4: 成绩

				Table 4	, ,,,,,	1			
力学	物理	代数	分析	统计	力学	物理	代数	分析	统计
(闭)	(闭)	(开)	(开)	(开)	(闭)	(闭)	(开)	(开)	(开)
X_1	X_2	X_3	X_4	X_5	X_1	X_2	X_3	X_4	X_5
77	82	67	67	81	63	78	80	70	81
75	73	71	66	81	55	72	63	70	68
63	63	65	70	63	53	61	72	64	73
51	67	65	65	68	59	70	68	62	56
62	60	58	62	70	64	72	60	62	45
52	64	60	63	54	55	67	59	62	44
50	50	64	55	63	65	63	58	56	37
31	55	60	57	73	60	64	56	54	40
44	69	53	53	53	42	69	61	55	45
62	46	61	57	45	31	49	62	63	62
44	61	52	62	46	49	41	61	49	64
12	58	61	63	67	49	53	49	62	47
54	49	56	47	53	54	53	46	59	44
44	56	55	61	36	18	44	50	57	81
46	52	65	50	35	32	45	49	57	64
30	69	50	52	45	46	49	53	59	37
40	27	54	61	61	31	42	48	54	68
36	59	51	45	51	56	40	56	54	35
46	56	57	49	32	45	42	55	56	40
42	60	54	49	33	40	63	53	54	25
23	55	59	53	44	48	48	49	51	37
41	53	49	46	34	46	52	53	41	40

5.12. 下表是从25 个家庭中测到的成年长子和次子的头宽、头长的数据. 试用典型相关分析的方法分析长子和次子的头宽、头长的相关情况.

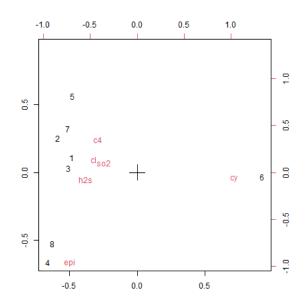
Table 5: 成年长子和次子的头宽、头长的数据

		ibic o.	/*/\	1 1111/	. J HJ/	· • > L • >	V MH13	~~ */H	
样品号	长子头长	长子头宽	次子头长	次子头宽	样品号	长子头长	长子头宽	次子头长	次子头宽
17111 3	X_1	X_2	X_3	X_4	17111 7	X_1	X_2	X_3	X_4
1	191	155	179	145	14	190	159	195	157
2	195	149	201	152	15	188	151	187	158
3	181	148	185	149	16	163	137	161	130
4	183	153	188	149	17	195	155	183	158
5	176	144	171	142	18	186	153	173	148
6	208	157	192	152	19	181	145	182	146
7	189	150	190	149	20	175	140	165	137
8	197	159	189	152	21	192	154	185	152
9	188	152	197	159	22	174	143	178	147
10	192	150	187	151	23	176	139	176	143
11	179	158	186	148	24	197	167	200	158
12	183	147	174	147	25	190	163	187	150
13	174	150	185	152					

对应分析与典型相关分析解答

1.解:

1 (1) 对应分析结果如下:



可以看出4和8与环氧氯丙烷相关性较大,6和环己烷相关性较大,其他与氯、硫化氢、二氧化硫,碳四的相关性较大

(2) 运用R型因子分析,选择三个因子,结果为氯、硫化氢、环氧氯乙烷 为一类,二氧化硫,环己烷为一类,碳4为一类

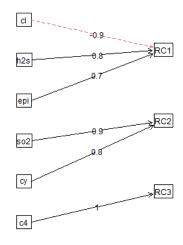
Loadings:

RC1 RC2 RC3
cl -0.868 -0.265 0.270
h2s 0.796 -0.170
so2 0.119 0.942
c4 0.149 0.970
epi 0.736 -0.147
cy -0.286 0.759 0.148

RC1 RC2 RC3

SS loadings 2.032 1.606 1.049 Proportion Var 0.339 0.268 0.175 Cumulative Var 0.339 0.606 0.781

Components Analysis



(3) Q型因子分析的结果如下:

Loadings:

RC1 RC2 RC3

1 0.187 0.274 0.902

2 0.831 0.223

3 0.897 0.249

4 0.335 0.671 -0.646

5 0.936

6 -0.232 -0.821 -0.201

7 0.980

8 -0.143 0.949

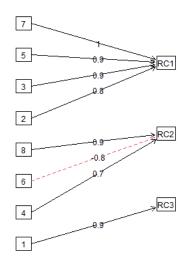
RC1 RC2 RC3

SS loadings 3.552 2.169 1.346

Proportion Var 0.444 0.271 0.168

Cumulative Var 0.444 0.715 0.883

Components Analysis



可以看出2,3,5,7为一类,4,6,8为一类,1为单独一类,与对应分析的结果不同

代码如下:

library("MASS")

```
cl < -c(0.056, 0.049, 0.038, 0.034, 0.084, 0.064, 0.048, 0.069)
h2s < -c(0.084, 0.055, 0.130, 0.095, 0.066, 0.072, 0.089, 0.087)
so2 < -c(0.031, 0.100, 0.079, 0.058, 0.039, 0.100, 0.062, 0.027)
c4 < -c(0.038, 0.110, 0.170, 0.160, 0.320, 0.210, 0.260, 0.050)
epi < -c(0.0081, 0.0220, 0.0580, 0.2000, 0.0120, 0.0280, 0.0380, 0.0890)
cy < -c(0.0220, 0.0073, 0.0430, 0.0290, 0.0410, 1.3800, 0.0360, 0.0210)
x < -data.frame(cl, h2s, so2, c4, epi, cy, row.names = c(1:8))
ca1 < -corresp(x, nf = 2)
ca1
biplot(ca1)
library(psych)
m < -fa.parallel(x)
x_f a < -principal(x, nfactors = 3, rotate = 'varimax')
x_f a loadings
fa.diagram(x_fa)
x_fa1 < -principal(t(x), nfactors = 3, rotate = 'varimax')
x_f a1loadings
fa.diagram(x_fa1)
2.解:
```

根据数据首先得到协方差矩阵,求解协方差矩阵的额特征值和特征向量,得到得特征值向量是B:

$$B = \begin{bmatrix} -0.0000 & 0 & 0 & 0 & 0 \\ 0 & 0.0000 & 0 & 0 & 0 \\ 0 & 0 & 0.0227 & 0 & 0 \\ 0 & 0 & 0 & 0.1023 & 0 \\ 0 & 0 & 0 & 0 & 0.2399 \end{bmatrix}$$

前两个特征值得累计贡献量为(.1023+0.2399)/(.1023+0.2399+0.0227)*100% = 93.78%,因此,选前两个特征值分别计算R型与Q型的因子载荷阵如下:

$$F = \begin{bmatrix} -0.2178 & -0.1217 \\ -0.2765 & 0.2306 \\ 0.0517 & -0.1305 \\ 0.3005 & 0.1102 \\ 0.1518 & 0.0715 \end{bmatrix} G = \begin{bmatrix} -0.1495 & 0.0655 \\ -0.0750 & -0.0553 \\ 0.0431 & -0.0356 \\ 0.1665 & 0.0431 \\ 0 & 0 \end{bmatrix}$$

最后在因子轴平面上以变量点和样品点画图,得到得结果如图1.

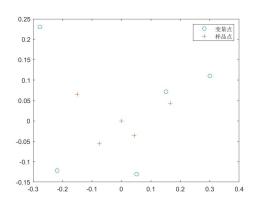


Figure 1: 因子轴

代码如下:

```
x = [190.33\ 43.77\ 9.73\ 60.54\ 49.01\ 9.04; 135.20\ 36.40\ 10.47\ 44.16\ 13.49\ 3.94; \dots
95.21\ 22.83\ 9.30\ 22.44\ 22.81\ 2.80; 104.78\ 25.11\ 6.40\ 9.89\ 18.17\ 3.25; \dots
128.41\ 27.63\ 8.94\ 12.58\ 23.99\ 3.27; 145.68\ 32.83\ 17.79\ 27.29\ 39.09\ 3.47; \dots
159.37\ 33.38\ 18.37\ 11.81\ 25.29\ 5.22; \\ 116.22\ 29.57\ 13.24\ 13.76\ 21.75\ 6.04; \\ \dots
221.11\ 38.64\ 12.53\ 115.65\ 50.82\ 5.89; 144.98\ 29.12\ 11.67\ 42.60\ 27.30\ 5.74; \dots
169.92\ 32.75\ 12.72\ 47.12\ 34.35\ 5.00; 153.11\ 23.09\ 15.62\ 23.54\ 18.18\ 6.39; \dots
144.92\ 21.26\ 16.96\ 19.52\ 21.75\ 6.73; \\ 140.54\ 21.5\ 17.64\ 19.19\ 15.97\ 4.94; \\ \dots
115.84 30.26 12.20 33.61 33.77 3.85;101.18 23.26 8.46 20.2 20.5 4.3];
[n p] = size(x);
T=sum(sum(x));
xliehe=zeros(1,p);
xhanghe=zeros(1,n);
for k=1:p
xliehe(k) = sum(x(:,k));
end
for l=1:n
xhanghe(l) = sum(x(l,:));
end
Z=zeros(p,p);
for i=1:1:n
```

```
for j=1:1:p
    Z(i,j) = (x(i,j) - xhanghe(i) * xliehe(j)/T)/((xhanghe(i) * xliehe(j))(1/2));
    end
    end
    A=Z^{*}Z;
    [XB] = eig(A);
    F = zeros(p,2);
    for t=1:1:p
    F(t,1) = (0.0357^{0}.5) * (X(t,6));
    F(t,2) = (0.0069^{0}.5) * (X(t,5));
    end
    \mathbf{F}
    G=Z*F
    h1=F(:,1);
    g1=F(:,2);
    h2=G(:,1);
    g2=G(:,2);
    plot(h1,g1,'o')
    hold on
    plot(h2,g2,'+')
    3.解:
    计算可得X组的典型变量为
                         U_1 = 0.7689X_1 + 0.2721X_2
                         U_2 = -1.4787X_1 + 1.644X_2
Y组的典型变量为
                     V_1 = 0.049Y1 + 0.8975Y_2 + 0.1900Y_3
                     V_2 = 1.000Y_1 - 0.5837Y_2 + 0.2956Y_3
典型相关系数为
                           \lambda_1 = 0.6879, \lambda_2 = 0.1869
    代码如下:
    clc, clear
    r = [1.00, 0.80, 0.26, 0.67, 0.34;
```

```
0.80, 1.00, 0.33, 0.59, 0.34;
   0.26, 0.33, 1.00, 0.37, 0.21;
   0.67, 0.59, 0.37, 1.00, 0.35;
   0.34, 0.34, 0.21, 0.35, 1.00;
   n1=2;n2=3;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'); %'接照从大到小排列'
   a = vec1(:, ind1(1:num))%'取出X组的系数阵'
   dcoef1 = val1(1:num)\%'J;'X'
   [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'
   mu = sum(xur.^2)/n1%'x组原始变量被u_i解释的方差比例'
   mv = sum(xvr.^2)/n1%'x组原始变量被v_i解释的方差比例'
```

 $nu = sum(yur.^2)/n2$ %'y组原始变量被 u_i 解释的方差比例' $nv = sum(yvr.^2)/n2$ %'y组原始变量被 v_i 解释的方差比例' $fprintf('`X| ext{ex} Cx u1 u$ %d解释的比例为'%f/n', num, sum(mu)); $fprintf('`Y| ext{ex} Cx u1 v$ %d解释的比例为'%f/n', num, sum(nv)); 4.解:

将数据输入SPSS中,得到结果

典型相关性

	相关性	特征值	威尔克统计	F	分子自由度	分母自由度	显著性
1	.643	.705	.578	4.094	6.000	78.000	.001
2	.117	.014	.986				

H0 for Wilks 检验是指当前行和后续行中的相关性均为零

集合 1 标准化典型相关 系数

变量	1	2
力学	584	947
物理	596	.940

集合 2 标准化典型相关 系数

变量	1	2
代数	844	418
分析	514	.183
统计	.481	1.048

集合 1 非标准化典型相 关系数

变量	1	2
力学	042	068
物理	052	.082

集合 2 非标准化典型相 关系数

变量	1	2
代数	117	058
分析	074	.026
统计	.031	.067

此图给出了典型相关系数及其检验,结果表明第一个典型相关系数是显著的,因此我们选择第一个典型相关变量进行解释。

具体来说,第一对典型相关变量的相关系数是0.643, p= 0.001; 上图分别是两组变量的标准化相关系数和未标准化的相关系数。 根据此图,可以写出各典型变量的表达式,如对于第一对典型变量u1和v1: 其标准化的表达式为(Z外向倾向表示将该变量标准化后的值)

 $u_1 = -0.584 * Z$ 力学-0.596 * Z物理

 $v_1 = -0.844 * Z$ 代数-0.514 * Z分析+0.481*统计

非标准化的表达式为

 $u_1 = -0.042 * Z力学 - 0.052*Z物理$

 $v_1 = -0.117 * Z$ 代数-0.074 * Z分析+0.031*统计

5.解:

由典型分析结果为

\$cor

[1] 0.7885079 0.0537397

\$xcoef

[,1] [,2]

x1 0.1127152 -0.2789099

x2 0.1064583 0.2813576

\$ycoef

[,1] [,2]

x3 0.1029701 -0.3610078

x4 0.1098775 0.3589657

\$xcenter

x1 x2

1.243450e-16 -6.049328e-16

\$ycenter

τ3

-3.380629e-16 -1.359746e-15

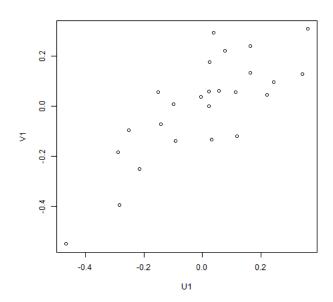
x4

进行相关系数的 χ^2 检验得到p值分别为0.0003,0.8,因此第二个典型相关系数可能为0,第一对典型相关变量为

$$U_1 = 0.1127X_1^* + 0.1065X_2^*$$

$$V_1 = 0.1030X_3^* + 0.1099X_4^*$$

其中 X_i^* 是 X_i 经过标准化后的结果。从散点图中可以看出 U_1 和 V_1 有较强的相关性



代码如下:

```
x1 < -c(191, 195, 181, 183, 176, 208, 189, 197, 188, 192, 179, 183, 174,
190, 188, 163, 195, 186, 181, 175, 192, 174, 176, 197, 190
x^{2} < -c(155, 149, 148, 153, 144, 157, 150, 159, 152, 150, 158, 147, 150,
159, 151, 137, 155, 153, 145, 140, 154, 143, 139, 167, 163
x3 < -c(179, 201, 185, 188, 171, 192, 190, 189, 197, 187, 186, 174, 185,
195, 187, 161, 183, 173, 182, 165, 185, 178, 176, 200, 187
x4 < -c(145, 152, 149, 149, 142, 152, 149, 152, 159, 151, 148, 147, 152,
157, 158, 130, 158, 148, 146, 137, 152, 147, 143, 158, 150)
x < -data.frame(x1, x2, x3, x4)
x < -scale(x)
ca1 < -cancor(x[, 1:2], x[, 3:4])
ca1
\%'Xu\Phi RS
corcoef.test;-function(r, n, p, q, alpha=0.05){
    %'r为相关系数n为样本个数且nip+q'
    m < -length(r); Q < -rep(0,m); lambda < -1
    for(kinm:1){
    %检验统计量
```

```
lambda < -lambda * (1 - r[k]^2);
    Q[k] < --\log(lambda)
    s < -0; i < -m
    for(kin1:m){
    %统计量
        Q[k] < -(n-k+1-1/2*(p+q+3)+s)*Q[k]
       chi < -1 - pchisq(Q[k], (p - k + 1) * (q - k + 1))
       print(chi)
        if(chi > alpha){
           i < -k-1; break
       s < -s + 1/r[k]^2
    %显示输出结果选用前几对典型变量
   print(i)
corcoef.test(r=ca1\$cor,n=25,p=2,q=2)
U < -as.matrix(x[, 1:2])
V < -as.matrix(x[, 3:4])
plot(U[, 1], V[, 1], xlab = "U1", ylab = "V1")
```