

2018 年全国大学生数学建模竞赛暨美赛培训

对应分析与典型相关分析

厦门大学2016 级各学院

数学建模团队：谭忠教授；助教：陈小伟，姜小蒙，姚瑶，余娇妍

要求：(1) 必须用TEX输入编辑后将TEXPDF以及图表一并发邮件提交给ztan85@163.com及sxjm004@163.com，压缩包及邮件主题名为“编号+姓名+专业+第\*次作业”；

(2) 必须抄题，以免判错。

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.

	氯 $X_1$	硫化氢 $X_2$	$SO_2$ $X_3$	$C_4$ $X_4$	环氧氯 丙烷 $X_5$	环己烷 $X_6$
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. 我国16 个地区农民1982 年支出情况的抽样汇总资料，每个地区都调查了反映每人平均生活消费支出情况的6 个指标.

(1) 使用对应分析方法对所考察的6 项指标和16 个地区进行分类.

(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) 用聚类分析方法分析该组数据；并与(1)，(2) 的结果比较之.

地区	食品 $X_1$	衣着 $X_2$	燃料 $X_3$	住房 $X_3$	生活用品 及其他 $X_5$	文化生活 服务支出 $X_6$
北京	190.33	43.77	9.73	60.54	49.01	9.04
天津	135.20	36.40	10.47	44.16	36.49	3.94
河北	95.21	22.83	9.30	22.44	22.81	2.80
山西	104.78	25.11	6.40	9.89	18.17	3.25
内蒙古	128.41	27.63	8.94	12.58	23.99	3.27
辽宁	145.68	32.83	17.79	27.29	39.09	3.47
吉林	159.37	33.38	18.37	11.81	25.29	5.22
黑龙江	116.22	29.57	13.24	13.76	21.75	6.04
上海	221.11	38.64	12.53	115.65	50.82	5.89
江苏	144.98	29.12	11.67	42.60	27.30	5.74
浙江	169.92	32.75	12.72	47.12	34.35	5.00
安徽	153.11	23.09	15.62	23.54	18.18	6.39
福建	144.92	21.26	16.96	19.52	21.75	6.73
江西	140.54	21.50	17.64	19.19	15.97	4.94
山东	115.84	30.26	12.20	33.61	33.77	3.85
河南	101.18	23.26	8.46	20.20	20.50	4.30

3. 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.

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

4. 下表是美国十个城市之间的飞行距离(英里). 试对其进行对应分析.

	亚特兰大	芝加哥	丹佛	休斯顿	洛杉矶	迈阿密	纽约	旧金山	西雅图
芝加哥	587	0							
丹佛	1212	920	0						
休斯顿	701	940	879	0					
洛杉矶	1936	1745	831	1374	0				
迈阿密	604	1188	1726	968	2339	0			
纽约	748	713	1631	1420	2451	1092	0		
旧金山	2139	1858	949	1645	347	2594	2571	0	
西雅图	2182	1737	1021	1891	959	2734	2048	678	0
华盛顿	543	579	1494	1220	2300	923	205	2442	2329

5. 进行十二指肠溃疡手术，有时存在不良的综合征. 下面的数据给出的是在四个医院中进行手术，依不同的手术处理给出的统计，其中不同的手术处理为

$A$ ：引流和迷走神经切除

$B$ ：25% 的切除和迷走神经

切除

头发颜色与眼睛颜色的数据

$C$  : 50% 的切除和迷走神经切除       $D$ : 75% 的切除

手术处理	综合症	医院			
		1	2	3	4
$A$	无	23	18	8	12
	有	9	7	9	10
$B$	无	23	18	12	15
	有	15	8	8	5
$C$	无	20	13	11	14
	有	18	15	8	11
$D$	无	24	9	7	13
	有	16	17	11	10

对上面的数据进行对应分析，研究医院和手术处理类型的关系.

6. 下表是某省12 个地区10 种恶性肿瘤的死亡率，试用对应分析法分析地区和死因的联系.

地区	鼻咽癌	食道癌	胃癌	肝癌	肠癌	肺癌	乳腺癌	宫颈癌	膀胱癌	白血病
1	3.89	14.06	48.01	21.39	5.38	9.57	1.65	0.15	0.60	3.29
2	2.17	26.00	24.92	22.75	8.67	10.29	1.08	0.00	0.00	3.25
3	0.00	2.18	5.44	22.84	4.35	17.40	1.09	4.35	0.00	4.35
4	1.46	7.61	31.92	26.94	6.15	15.82	2.05	1.45	0.29	2.93
5	0.89	46.37	11.59	32.10	0.89	9.81	0.89	3.57	0.89	1.78
6	0.60	1.81	16.27	19.28	3.01	6.02	1.20	0.60	0.00	4.82
7	1.74	8.72	3.20	24.70	2.03	4.36	0.00	0.58	2.03	2.62
8	1.98	41.18	44.15	35.22	4.96	14.88	0.00	0.00	0.00	4.96
9	2.14	3.00	13.29	26.58	5.14	8.14	1.71	6.86	0.00	3.00
10	1.83	37.97	10.45	36.13	4.59	14.86	1.65	0.00	0.73	3.67
11	4.71	20.71	23.77	42.84	12.24	24.24	5.41	3.06	0.24	4.24
12	1.66	4.98	6.64	35.71	5.81	18.27	0.83	2.49	0.00	7.47

7. 下表是2003 年13 个地区的城市房屋建筑和住房情况的几项指标数据，其中；

$A$ . 建成区面积(平方公里)

$B$ . 征用土地面积(平方公里)

$C$ . 城市人口密度(人/平方公里)

$D$ . 年末全市实有房屋建筑面积(万平方米)

$E$ . 年末全市实有住宅建筑面积(万平方米)

地区	$A$	$B$	$C$	$D$	$E$
全国	14321.4	719.9	847	1409091.4	891114.6
北京	1180.1	56.6	2128	43121.5	23847.5
河北	1171.0	50.7	2057	66359.1	42375.3
辽宁	1694.6	36.6	1244	67315.8	40488.0
上海	549.6	88.1	1971	51375.0	30560.0
浙江	1397.0	209.5	1117	76756.9	50449.6
山东	2195.4	111.4	901	99112.1	60591.8
湖北	1415.6	23.5	619	60716.8	39255.5
广东	2546.9	84.2	1779	123771.5	80755.3
海南	176.7	2.1	851	7695.5	4747.8
四川	1357.4	43.1	237	76077.6	49999.8
西藏	72.4	0.4	1137	977.3	781.9
新疆	564.8	13.7	211	21275.2	12797.5

试对这些地区城市房屋建筑及住房情况进行对应分析.

8. 在140 个学生中进行阅读训练速度 $X_1$ ，阅读能力，运算速度和运算能力供4 种测验，由所得测验成绩算出相关系数阵为

$$R = \begin{pmatrix} 1.00 & 0.63 & 0.24 & 0.59 \\ 0.63 & 1.00 & -0.06 & 0.07 \\ 0.24 & -0.06 & 1.00 & 0.42 \\ 0.59 & 0.07 & 0.42 & 1.00 \end{pmatrix}$$

试分析学生的阅读能力和运算能力之间的相关程度.

9. 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.

	$x_1$	$x_2$	$y_1$	$y_2$	$y_3$
$x_1$	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

10. 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.

力学 (闭)	物理 (闭)	代数 (开)	分析 (开)	统计 (开)	力学 (闭)	物理 (闭)	代数 (开)	分析 (开)	统计 (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

11. 某调查公司对某一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.

$X$  组:  $X_1$  — 用户反馈;  $X_2$  — 任务重要性;  $X_3$  — 任务多样性;  $X_4$  — 任务特殊性;  $X_5$  — 自主权.

$Y$  组:  $y_1$  — 主管满意度;  $y_2$  — 事业前景满意度;  $y_3$  — 财

政满意度； $y_4$  —— 工作强度满意度； $y_5$  —— 公司地位满意度； $y_6$  —— 工作满意度； $y_7$  —— 总体满意度.

它们之间的相关矩阵如下，试对这两组指标进行典型相关分析.

变量间的相关系数矩阵

	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$Y_1$	$Y_2$	$Y_3$	$Y_4$	$Y_5$	$Y_6$	$Y_7$
$X_1$	1.00	0.49	0.53	0.49	0.51	0.33	0.32	0.20	0.19	0.30	0.37	0.21
$X_2$	0.49	1.00	0.57	0.46	0.53	0.30	0.21	0.16	0.08	0.27	0.35	0.20
$X_3$	0.53	0.57	1.00	0.48	0.57	0.31	0.23	0.14	0.07	0.24	0.37	0.18
$X_4$	0.49	0.46	0.48	1.00	0.57	0.24	0.22	0.12	0.19	0.21	0.29	0.16
$X_5$	0.51	0.53	0.57	0.57	1.00	0.38	0.32	0.17	0.23	0.32	0.36	0.27
$Y_1$	0.33	0.30	0.31	0.24	0.38	1.00	0.43	0.27	0.24	0.34	0.37	0.40
$Y_2$	0.32	0.21	0.23	0.22	0.32	0.43	1.00	0.33	0.26	0.54	0.32	0.58
$Y_3$	0.20	0.16	0.14	0.12	0.17	0.27	0.33	1.00	0.25	0.46	0.29	0.45
$Y_4$	0.19	0.08	0.07	0.19	0.23	0.24	0.26	0.25	1.00	0.28	0.30	0.27
$Y_5$	0.30	0.27	0.24	0.21	0.32	0.34	0.54	0.46	0.28	1.00	0.35	0.59
$Y_6$	0.37	0.35	0.37	0.29	0.36	0.37	0.32	0.29	0.30	0.35	1.00	0.31
$Y_7$	0.21	0.20	0.18	0.16	0.27	0.40	0.58	0.45	0.27	0.59	0.31	1.00

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.

成年长子和次子的头宽、头长的数据

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
$X_1$	$X_2$	$X_3$	$X_4$		$X_1$	$X_2$	$X_3$	$X_4$	
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					

13. 某学校为研究学生的体质与运动能力的关系，对38 名学生的体质情况，每人测试了7 项指标：

$X_1$ ：反复横荡次数的次数       $X_2$ ：纵跳高度       $X_3$ ：背力  
 $X_4$ ：握力  
 $X_5$ ：踏台升降指数       $X_6$ ：立姿体前屈       $X_7$ ：卧姿上体后仰

对运动能力情况每人测试了5 项指标：

$X_8$ ：50米跑       $X_9$ ：1000米长跑       $X_{10}$ ：投掷  
 $X_{11}$ ：悬垂次数       $X_{12}$ ：持久走

七项体质数据和5 项运动数据见下表. 试对这两组数据进行典型相关分析.

学生体质与运动能力数据												
学生 序号	体质情况							运动能力				
	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$	$X_8$	$X_9$	$X_{10}$	$X_{11}$	$X_{12}$
1	46	55	126	51	75.0	25	72	6.8	489	27	8	360
2	52	55	95	42	81.2	18	50	7.2	464	30	5	348
3	46	69	107	38	98.0	18	74	6.8	430	32	9	386
4	49	50	105	48	97.6	16	60	6.8	362	26	6	331
5	42	55	90	46	66.5	2	68	7.2	453	23	11	391
6	48	61	106	43	78.0	25	58	7.0	405	29	7	389
7	49	60	100	49	90.6	15	60	7.0	420	21	10	379
8	48	63	122	52	56.0	17	68	7.0	466	28	2	362
9	45	55	105	48	76.0	15	61	6.8	415	24	6	386
10	48	64	120	38	60.2	20	62	7.0	413	28	7	398
11	49	52	100	42	53.4	6	42	7.4	404	23	6	400
12	47	62	100	34	61.2	10	62	7.2	427	25	7	407
13	41	51	101	53	62.4	5	60	8.0	372	25	3	409
14	52	55	125	43	86.3	5	62	6.8	496	30	10	350
15	45	52	94	50	51.4	20	65	7.6	394	24	3	399
16	49	57	110	47	72.3	19	45	7.0	446	30	11	337
17	53	65	112	47	90.4	15	75	6.6	420	30	12	357
18	47	57	95	47	72.3	9	64	6.6	447	25	4	447
19	48	60	120	47	86.4	12	62	6.8	398	28	11	181
20	49	55	113	41	84.1	15	60	7.0	398	27	4	387
21	48	69	128	42	47.9	20	63	7.0	485	30	7	350
22	42	57	122	46	54.2	15	63	7.2	400	28	6	388
23	54	64	155	51	71.4	19	61	6.9	511	33	12	298
24	53	63	120	42	56.6	8	53	7.5	430	29	4	353
25	42	71	138	44	65.2	17	55	7.0	487	29	9	370
26	46	66	120	45	62.2	22	68	7.4	470	28	7	360
27	45	56	91	29	66.2	18	51	7.9	380	26	5	358
28	50	60	120	42	56.6	8	57	6.8	460	32	5	348
29	42	51	126	50	50.0	13	57	7.7	398	27	2	383
30	48	50	115	41	52.9	6	39	7.4	415	28	6	314
31	42	52	140	48	56.3	15	60	6.9	470	27	11	348
32	48	67	105	39	69.2	23	60	7.6	450	28	10	326
33	49	74	151	49	54.2	20	58	7.0	500	30	12	330
34	47	55	113	40	71.4	19	64	7.6	410	29	7	331
35	49	74	120	53	54.5	22	59	6.9	500	33	21	348
36	44	52	110	37	54.9	14	57	7.5	400	29	2	421
37	52	66	130	47	45.9	14	45	6.8	505	28	11	355
38	48	68	100	45	53.6	23	70	7.2	522	28	9	352

year：年份；

Dec：12 月份平均气温；

Jan：次年一月份平均气温；

Feb：次年二月份平均气温；

high7：7 月500hpa 图上13° ~ 14°E，40° ~ 50°N范围内6 点高度距平和；

high4：4 月500hpa 图上(110°E,40°N)(100°W,40°N) 和(100° E,50° N) 3 点高度距平和；

high8：7 月500hpa 图上150°E,35° ~ 45°N，100° E 40° ~ 50°N 范围内5 点高度距平和.

14. 现有北京地区1951 ~ 1976 年的资料见下表. 其中：

北京地区冬季气温

year	Dec	Jan	Feb	high7	high4	high8
1951	1.0	-2.7	-4.3	4	-7	12
1952	-5.3	-5.9	-3.5	0	21	5
1953	-2.0	-3.4	-0.8	6	-9	5
1954	-5.7	-4.7	-1.1	10	17	6
1955	-0.9	-3.8	-3.1	1	5	11
1956	-5.7	-5.3	-5.9	-3	1	-12
1957	-2.1	-5.0	-1.6	-15	3	13
1958	0.6	-4.3	-0.2	10	-3	0
1959	-1.7	-5.7	2.0	-9	-5	-14
1960	-3.6	-3.6	1.3	11	-3	18
1961	-3.0	-3.1	-0.8	5	-15	4
1962	0.1	-3.9	-1.1	8	12	1
1963	-2.6	-3.0	-5.2	11	3	-3
1964	-1.4	-4.9	-1.7	-11	-8	7
1965	-3.9	-5.7	-2.5	-18	6	-6
1966	-4.7	-4.8	-3.3	-9	-6	15
1967	-6.0	-5.6	-4.9	4	0	-20
1968	-1.7	-6.4	-5.1	-7	-2	-15
1969	-3.4	-5.6	-2.0	4	17	-23
1970	-3.1	-4.2	-2.9	9	-16	23
1971	-3.8	-4.9	-3.9	-13	5	-2
1972	-2.0	-4.1	-2.4	7	0	10
1973	-1.7	-4.2	-2.0	27	-11	4
1974	-3.6	-3.3	-2.0	17	-2	0
1975	-2.7	-3.7	0.1	-1	-13	10
1976	-2.4	-7.6	-2.2	5	9	-30