

2018 年全国大学生数学建模竞赛暨美赛培训

主成分分析和因子分析

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 $sxjm004@163.com$ ，压缩包及邮件主题为“编号+姓名+专业+第*次作业”；

（2）必须抄题，以免判错。

一、主成分分析练习

1. 用主成分分析法探讨城市工业主体结构，下表是某市工业部门13个行业8个指标的数据.

（1）利用主成分分析确定8个指标的几个主分量，并对主分量进行解释；

（2）利用主成分得分对13个行业进行排序和分类.

指标	年末固定 资产净值	职工人数	工业总产值	全员劳动 生产率	百元固定 资产原值	资金利 税率	标准燃料 消费量	能源利 用效果
指标值	X_1	X_2	X_3	$X_4/\text{元}\cdot$	实现产值 X_5	X_6	X_7	$X_8(/ \text{万元}$
序号	/万元	/人	/万元	(人·年 $^{-1}$)	/元	/%	/吨	·吨 $^{-1}$)
1 冶金	90432	52455	101091	19272	82.000	16.100	197435	0.172
2 电力	4903	1973	2035	10313	34.200	7.100	592077	0.003
3 煤炭	6735	21139	3767	1780	36.100	8.200	726396	0.003
4 化学	49454	36241	81557	22504	98.100	25.900	348226	0.985
5 机械	139190	203505	215898	10609	93.200	12.600	139572	0.628
6 建材	12215	16219	10351	6328	62.500	8.700	145818	0.066
7 森工	2372	6572	8103	12329	184.400	22.200	20921	0.152
8 食品	11062	23078	54935	23804	370.400	41.000	65486	0.263
9 纺织	17111	23907	52108	21796	221.500	21.500	63806	0.276
10 缝纫	1206	3930	6126	15586	330.400	29.500	1840	0.437
11 皮革	2150	5704	6200	10870	184.200	12.000	8913	0.274
12 造纸	5251	6155	10383	16875	146.400	27.500	78796	0.151
13 文教艺 术用品	14341	13203	19396	14691	94.600	17.800	6354	1.574

2. 上海某县有19个乡，各乡的经济发展水平不尽相同，由县统计局得到下列数据：1986年乡人均收入（ X_1 ）；1986年乡净收入（ X_2 ）；1980年至1986年每年各乡的村办企业的总收入为（ Y_1 ）至（ Y_7 ）. 数据见下表.

（1）用主成分分析方法对变量进行综合；并解释综合指标的含义；

（2）对19个乡的经济发展水平按综合变量进行排序；

序号	X_1	X_2	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7
1	843	4000	920	1083	1166	1536	1692	2743	2741
2	942	5215	1216	1280	1549	1924	2717	3743	4160
3	895	6379	1836	2287	2715	3218	3240	4572	5439
4	884	3770	978	1010	1227	1558	1755	2552	2632
5	1002	3329	757	684	718	931	1422	1827	2222
6	792	3039	759	765	1253	1296	1902	1822	2233
7	930	2476	559	616	611	863	891	1539	1768
8	758	2127	633	661	762	769	864	1241	1482
9	886	2721	703	836	813	914	895	1272	1431
10	859	1664	427	451	507	551	581	839	1107
11	808	2435	772	790	839	990	771	1132	1316
12	845	2661	557	599	670	855	1128	1639	1861
13	877	2242	549	471	480	643	664	1173	1530
14	787	3088	1061	1021	978	1102	1301	2373	2452
15	943	3322	567	557	630	836	984	1994	2704
16	866	3507	675	678	847	1187	1476	2107	2415
17	924	3412	545	729	821	1017	1079	1878	2325
18	1806	5869	1268	1227	1607	1765	2318	2959	3157
19	1027	3605	651	835	924	1125	1194	2176	2411

行业	资产总计	固定资产净值	从业人员 人数/万人	工业产值	主营业务收入	利润总额
煤炭开采和 洗选业	19457.74	6477.02	502.38	14625.92	15315.15	2348.45
石油和天然 气开采业	12806.58	6867.23	112.76	10615.96	11052.97	4601.23
黑色金属矿 采选业	3179.97	819.91	61.52	3760.65	3635.66	700.37
有色金属矿 采选业	2290.30	689.75	53.53	2727.84	2705.99	407.31
非金属矿 采选业	1330.02	500.63	54.23	1869.49	1822.50	168.92
其他采矿业	4.07	2.19	0.28	10.35	9.74	0.57

3. 利用主成分分析法，对我国2008年6个工业行业的经济效益指标进行综合评价.

4. 对中国大陆31个省（市、区）2008年第三产业综合发展水平的主成分分析与评估. 选取了人均地区生产总值(元)，人均第三产业增加值（元），第二产业占GDP的比重，第三产业占GDP的比重，第三产业就业人员比重，城镇化水平(%），第三产业固定资产投资比重八项指标，具体数据如下表.

地区	人均地区 生产总 值/元	人均第三 产业增加 值/元	第二产业 占GDP的 比重/%	第三产业 占GDP的 比重/%	第三产业 就业人员 比重/%	城镇化 水平/%	第三产业 固定资产 投资比重
北京	63029	5186.03	25.7	73.2	72.5	84.9	0.364
天津	55473	2383.56	60.1	37.9	44	77.23	0.344
河北	23239	853.68	54.2	33.2	26.5	41.9	0.223
山西	20.398	770.16	61.5	34.2	33	45.11	0.2583
内蒙古	32214	1184.36	55	33.3	32.7	51.71	0.2713
辽宁	31259	1192.02	55.8	34.5	41.2	60.05	0.255
吉林	23514	986.40	47.7	38	35.4	53.21	0.2545
黑龙江	21727	823.053	52.5	34.4	33	55.4	0.336
上海	73124	4361.24	45.5	53.7	55.2	88.6	0.374
江苏	39622	1669.67	55	38.1	34.7	54.3	0.199
浙江	42214	1919.64	53.9	41	35.3	57.6	0.261
安徽	14485	597.99	46.6	37.4	28.3	40.5	0.248
福建	30123	1308.21	50	39.3	33.3	49.9	0.336
江西	14781	506.04	52.7	30.9	31.9	41.6	0.254
山东	33083	1220.10	57	33.4	31	47.6	0.254
河南	19593	620.81	56.9	28.6	24.4	36.03	0.201
湖北	19860	887.24	43.8	40.5	40.1	45.2	0.343
湖南	17521	731.36	44.2	37.8	30.4	42.15	0.317
广东	37589	1787.76	51.6	42.9	38.2	63.37	0.295
广西	14966	619.61	42.4	37.4	24.8	38.16	0.324

地区	人均地区 生产总 值/元	人均第三 产业增加 值/元	第二产业 占GDP的 比重/%	第三产业 占GDP的 比重/%	第三产业 就业人员 比重/%	城镇化 水平/%	第三产业 固定资产 投资比重
海南	17175	766.11	29.8	40.2	34.9	48	0.484
重庆	18025	817.38	47.7	41	36.3	49.99	0.321
四川	15378	590.05	46.3	34.8	33.1	37.4	0.307
贵州	8824	403.46	42.3	41.3	36.3	29.11	0.315
云南	12587	544.14	43	39.1	25.1	33	0.348
西藏	13861	852.64	29.2	55.5	33.9	22.61	0.541
陕西	18246	663.41	56.1	32.9	32.5	42.1	0.407
甘肃	12110	523.051	46.3	39.1	32.8	32.15	0.299
青海	17389	652.19	55.1	34	34.2	40.86	0.330
宁夏	17892	718.00	52.9	36.2	30.1	44.98	0.233
新疆	19893	750.14	49.6	33.9	34.7	39.64	0.273

5. 对我国31省2008年城镇居民家庭平均每人全年消费性支出进行主成分分析. 其8个指标分别为: X_1 , 食品; X_2 , 衣着; X_3 , 居住; X_4 , 家庭设备用品及服务; X_5 , 医疗保健; X_6 , 交通和通信; X_7 , 文教、娱乐及服务; X_8 , 其他商品和服务. 具体数据如下表: 数据单位(元).

地区	食品	衣着	居住	家庭设备				
				用品及 服务	医疗保健	交通和 通信	文教、娱乐 用品及服务	其他商 品和服务
北京	5561.54	1571.74	1286.32	1096.57	1563.10	2293.23	2383.52	704.24
天津	5005.09	1153.66	1528.28	817.18	1220.92	1567.87	1608.97	520.49
河北	3155.40	1137.22	1097.41	574.84	808.88	1062.31	946.38	304.28
山西	2974.76	1137.11	1250.87	471.65	769.79	931.33	1041.91	228.53
内蒙古	3553.48	1616.56	1028.19	672.64	869.71	1191.70	1383.53	512.81
辽宁	4378.14	1187.41	1270.95	507.40	913.13	1295.70	1145.46	533.29
吉林	3307.14	1259.62	1285.28	510.49	914.47	954.96	1071.80	425.30
黑龙江	3128.10	1217.04	941.25	494.49	864.89	749.05	906.19	321.95
上海	7108.62	1520.61	1646.19	1182.24	755.29	3373.19	2874.54	937.21
江苏	4544.64	1166.91	1042.10	813.45	794.63	1357.96	1799.75	458.10
浙江	5522.56	1546.46	1333.69	713.31	933.11	2392.63	2195.58	520.95
安徽	3905.05	1010.61	988.12	579.59	633.93	920.77	1160.14	325.82
福建	5078.85	1105.31	1300.10	722.17	540.63	1777.06	1453.18	523.83
江西	3633.05	969.58	851.15	623.17	483.96	872.57	945.99	337.91
山东	3699.42	1394.11	1247.04	806.35	799.79	1410.45	1277.43	372.01
河南	3079.82	1141.76	963.59	633.32	790.87	915.12	988.95	324.03
湖北	3996.27	1099.16	914.26	604.40	675.32	890.12	1037.24	260.74
湖南	3970.42	1090.72	960.82	674.84	790.95	971.05	1110.11	376.62
广东	5866.91	975.06	1748.16	947.54	836.39	262308	1936.38	594.45
广西	4082.99	772.28	891.33	603.84	529.36	1376.03	1081.54	290.04
海南	4226.90	491.84	1106.39	565.51	536.40	1303.50	930.84	247.08
重庆	4418.34	1294.30	1096.82	842.09	878.25	1044.36	1267.03	305.60

地区	食品	衣着	居住	家庭设备				
				用品及 服务	医疗保健	交通和 通信	文教、娱乐 用品及服务	其他商 品和服务
四川	4255.48	1042.45	819.28	590.51	564.93	1121.45	947.01	338.03
贵州	3597.94	851.50	836.54	525.70	471.39	871.15	934.73	260.27
云南	4272.29	1026.50	739.20	331.94	606.86	1216.46	732.95	150.42
西藏	4262.77	1011.82	634.94	310.22	317.08	966.74	419.59	400.38
陕西	3586.13	1047.61	1007.68	618.16	862.70	967.52	1281.58	400.68
甘肃	3183.79	1022.62	846.26	546.23	654.82	817.17	936.33	301.40
青海	3315.94	945.14	802.73	538.54	610.02	787.63	880.86	311.72
宁夏	3352.83	1178.88	1069.15	596.81	816.87	1096.32	1043.72	403.71
新疆	3235.77	1245.02	781.30	535.31	643.48	1003.89	812.36	411.63

二、因子分析练习

1. 下表是全国医药行业20个企业1980-1982年三年平均经济效益的几个主要数据.

(1)试用因子分析法找出这4个变量的公因子，并进行合理的解释；

(2)利用因子得分对20个企业进行分类，并与其它分类方法比较分类的结果.

企业号	总产值/消耗 X_1	资产值/工资 X_2	盈利/资金占用 X_3	销售收入/成本 X_4
1	1.611	10.59	0.69	1.67
2	1.429	9.44	0.61	1.50
3	1.447	5.97	0.24	1.25
4	1.572	10.72	0.75	1.71
5	1.483	10.99	0.75	1.44
6	1.371	6.46	0.41	1.31
7	1.665	10.51	0.53	1.52
8	1.403	6.11	0.17	1.32
9	2.620	21.51	1.40	2.59
10	2.033	24.15	1.80	1.89
11	2.015	26.86	1.93	2.02
12	1.501	9.74	0.87	1.48
13	1.578	14.52	1.12	1.47
14	1.735	14.64	1.21	1.91
15	1.453	12.88	0.87	1.52
16	1.765	17.94	0.89	1.40
17	1.532	29.42	2.52	1.80
18	1.488	9.23	0.81	1.45
19	2.586	16.07	0.82	1.83
20	1.992	21.63	1.01	1.89

序号	数学	物理	化学	语文	历史	英语
1	65	61	72	84	81	79
2	77	77	76	64	70	55
3	67	63	49	65	67	57
4	80	69	75	74	74	63
5	74	70	80	84	81	74
6	78	84	75	62	71	64
7	66	71	67	52	65	57
8	77	71	57	72	86	71
9	83	100	79	41	67	50
10	86	94	97	51	63	55
11	74	80	88	64	73	66
12	67	84	53	58	66	56
13	81	62	69	56	66	52
14	71	64	94	52	61	52
15	78	96	81	80	89	76
16	69	56	67	75	94	80
17	77	90	80	68	66	60
18	84	67	75	60	70	63
19	62	67	83	71	85	77
20	74	65	75	72	90	73
21	91	74	97	62	71	66
22	72	87	72	79	83	76
23	82	70	83	68	77	85
24	63	70	60	91	85	82
25	74	79	95	59	74	59
26	66	61	77	62	73	64
27	90	82	98	47	71	60
28	77	90	85	68	73	76
29	91	82	84	54	62	60
30	78	84	100	51	60	60

2. 利用因子分析方法分析下列30个学生成绩的因子构成，并分析各个学生较适合学文科还是理科.

3. 在某中学随机抽取某年级30名学生，测量其身高(X_1)、体

重(X_2)、胸围(X_3)和坐高(X_4), 见下表. 试对这30名中学生身体4项指标数据做主成分分析.

序号	X_1	X_2	X_3	X_4	序号	X_1	X_2	X_3	X_4
1	148	41	72	78	2	139	34	71	76
3	160	49	77	86	4	149	36	67	79
5	159	45	80	86	6	142	31	66	67
7	153	43	76	83	8	150	43	77	68
9	151	42	77	80	10	139	31	68	74
11	140	29	64	74	12	161	47	78	84
13	158	49	78	83	14	140	33	67	77
15	137	31	66	73	16	152	35	73	79
17	149	47	82	79	18	145	35	70	77
19	160	47	74	87	20	156	44	78	85
21	151	42	73	82	22	147	38	73	78
23	157	39	68	80	24	147	30	65	75
25	157	39	68	80	26	151	36	74	80
27	144	36	68	76	28	141	30	67	76
29	139	32	68	73	30	148	38	70	78

4. 对中国大陆31省（市、区）2008年第三产业综合发展水平进行因子分析，找出公因子，并进行合理解释.（数据即主成分分析第4题数据.）

5. 对我国31省2008年城镇居民家庭平均每人全年消费性支出进行因子分析，找出公因子，并进行合理解释.（数据即主成分分析第5题数据.）

6. 问题背景：我国31个省(市、区)2008年经济发展基本情况的7项指标为：

X_1 ：地区生产总值/亿元； X_2 ：居民消费水平/元； X_3 ：固定资

产投资/亿元； X_4 ：职工平均工资/元； X_5 ：居民消费价格指数； X_6 ：商品零售价格指数； X_7 ：工业总产值/亿元. 找出公因子，并进行合理解释.

地区	地区生产总值/亿元	居民消费水平/元	固定资产投资/亿元	职工平均工资/元	居民消费价格指数	商品零售价格指数	工业总产值/亿元
北京	9353.20	20346.00	3814.70	56328.00	105.10	104.40	10413.09
天津	5050.40	14000.00	3389.80	41748.00	105.40	105.10	12503.25
河北	13709.50	6570.00	8866.60	24756.00	106.20	106.70	23030.73
山西	5733.35	6187.00	3531.20	25828.00	107.20	107.20	10023.87
内蒙古	6091.12	8108.00	5475.40	26114.00	105.70	104.70	8740.18
辽宁	11023.49	9625.00	10019.10	27729.00	104.60	105.30	24769.09
吉林	5284.69	7591.00	5038.90	23486.00	105.10	106.20	8406.85
黑龙江	7065.00	7039.00	3656.00	23046.00	105.60	105.80	7624.54
上海	12188.85	27343.00	4823.10	56565.00	105.80	105.30	25120.92
江苏	25741.15	11013.00	15300.60	31667.00	105.40	104.90	67798.68
浙江	18780.44	13893.00	9323.00	34146.00	105.00	106.30	40832.10
安徽	7364.18	6377.00	6747.00	26363.00	106.20	106.30	11162.16
福建	9249.13	10361.00	5207.70	25702.00	104.60	105.70	15212.81
江西	5500.25	5753.00	4745.40	21000.00	106.00	106.10	8499.58
山东	25965.91	9573.00	15435.90	26404.00	105.30	104.90	62958.53
河南	15012.46	5877.00	10490.60	24816.00	107.00	107.50	26028.41
湖北	9230.68	7406.00	5647.00	22739.00	106.30	106.30	13454.94
湖南	9200.00	7145.00	5534.00	24870.00	106.00	105.60	11553.31
广东	31084.40	14390.00	10868.70	33110.00	105.60	106.00	65424.61
广西	5955.65	6103.00	3756.40	25660.00	107.80	107.60	6071.98

地区	地区生产 总值/亿元	居民消费 水平/元	固定资产 投资/亿元	职工平均 工资/元	居民消费 价格指数	商品零售 价格指数	工业总产 值/亿元
海南	1223.28	6550.00	705.40	21864.00	106.90	106.70	1103.07
重庆	4122.51	9835.00	3979.60	26985.00	105.60	105.00	5755.90
四川	10505.30	6072.00	7127.80	25038.00	105.10	105.30	14761.86
贵州	2741.90	4426.00	1864.50	24602.00	107.60	107.20	3111.13
云南	4741.31	4553.00	3435.90	24030.00	105.70	106.10	5144.58
西藏	342.19	3504.00	309.90	47280.00	105.70	103.90	48.19
陕西	5465.79	6290.00	4614.40	25942.00	106.40	106.90	7480.79
甘肃	2702.40	4869.00	1712.80	24017.00	108.20	107.90	3667.52
青海	783.61	5830.00	583.20	30983.00	110.10	110.60	1103.10
宁夏	889.20	7193.00	828.90	30719.00	108.50	108.50	1366.45
新疆	3523.16	5542.00	2260.00	24687.00	108.10	108.50	4276.05