



题目一

Pain 数据集包含了 121 位颞下颌关节疾病患者抱怨面部和下颌疼痛的数据。受试者用数字 1 到 9 对痛感评级，对 14 个问题作出回答。数据集中包含如下变量：

- Hurtchew does it hurt when you chew?
- Hurtwide does it hurt when you open wide or take a big bite?
- Noise does your jaw make a noise so that it bothers you or others?
- StiffjawStiff jaw (in the morning)?
- Cracking locking/cracking of jaw joint (in the morning)?
- Painslp does the pain or discomfort disturb your sleep?
- Painrout does the pain or discomfort interfere with your daily routine or other activities?
- Paintab do you take tablets for pain or discomfort?
- Amhdache headache (in the morning)?
- Earpain do you have ear pain or pain in front of the ears?
- Grind has anyone ever heard you grinding your teeth in your sleep, or are you aware of it yourself?
- Clampset are you aware that you clamp or set your jaw?
- Sorejaw sore jaw or teeth (in the morning)?
- Facepain do you have pain in the face, jaw, eyes, throat, neck, or temples?

- 1) 对 Pain 数据做主成分分析，解释主成分分析的结果。
- 2) 使用最大似然法(Maximum likelihood, ML)对 Pain 数据做因子分析，你得到几个因子？
得出该结论的依据是？
- 3) 对因子做正交旋转(Varimax)，解释旋转后得到的因子。

【答】 本次作业所有源代码见附件 solution.sas

- 1) 使用 PCA 对 pain 数据分析的结果如下：

The PRINCOMP Procedure

Observations	121
Variables	16

上图表明我们对 16 个变量成功观测到了 121 个观测值。下图为各变量均值与方差。

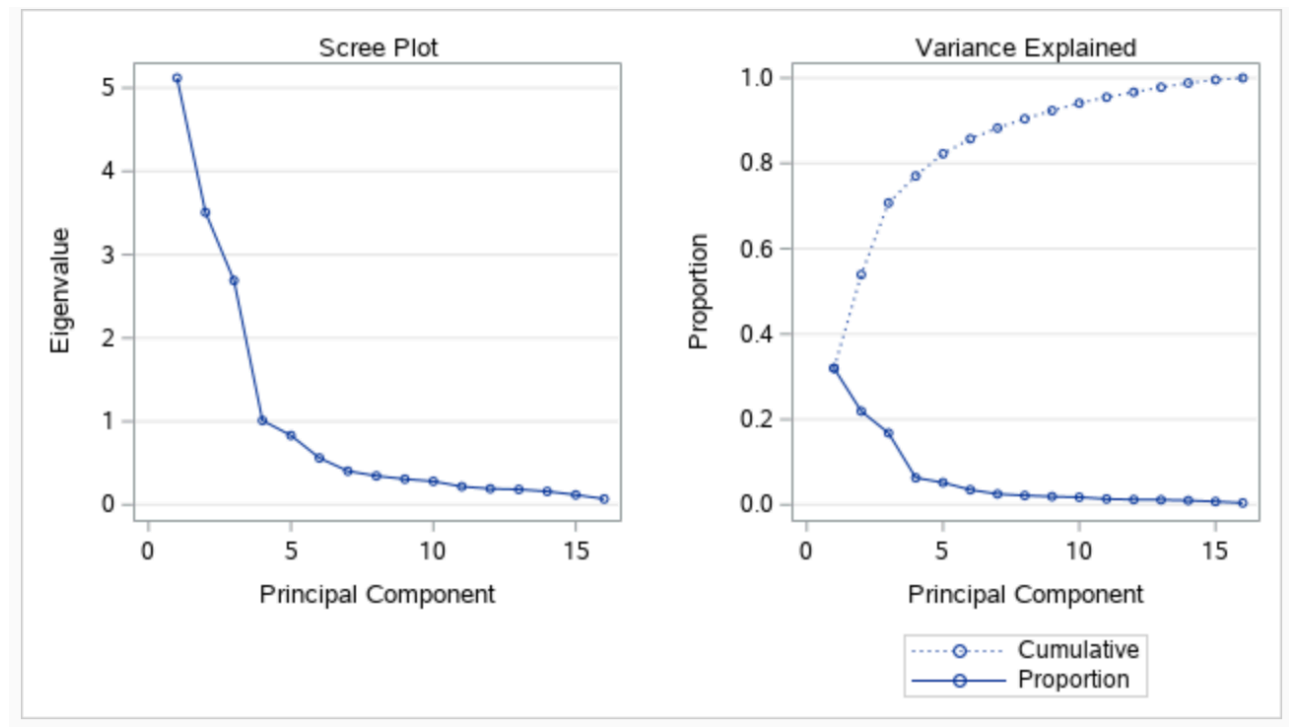
	Simple Statistics															
	isub	hurtchew	hurtwide	noise	stiffjaw	cracking	painslp	painrout	paintab	amhdache	earpain	grind	clampset	sorejaw	facepain	dayslost
Mean	61.00000000	4.495867769	5.479338843	5.008264463	3.991735537	4.479338843	4.000000000	4.016528926	4.495867769	5.520661157	5.685950413	5.024793388	4.537190083	4.512396694	3.512396694	5.471074380
Std	35.07373186	1.041825057	1.073150918	1.076226957	1.012388823	1.041626721	1.064581295	0.999862249	1.065551242	1.033595452	1.033062259	1.060367939	1.057050317	1.081015745	1.009254696	0.992256521

下图的每一列是第 compi 个主成分变量对 16 个原始自变量的解释力度。

	Eigenvectors															
	comp1	comp2	comp3	comp4	comp5	comp6	comp7	comp8	comp9	comp10	comp11	comp12	comp13	comp14	comp15	comp16
isub	0.015438	-0.079925	0.129766	0.943225	0.087978	-0.083835	-0.035457	0.220971	0.056289	0.000712	0.097780	-0.013432	0.066510	-0.034897	-0.017161	0.056223
hurtchew	0.328115	-0.304930	-0.100218	-0.020433	-0.009561	-0.043136	-0.059865	0.072480	-0.078116	0.251118	0.108331	0.211277	0.037995	-0.396726	-0.075570	-0.698589
hurtwide	0.309136	-0.322294	-0.129050	-0.097098	0.024159	0.065974	-0.016983	0.022493	-0.034929	0.223716	0.245719	0.123648	0.008843	-0.374741	-0.137319	0.694585
noise	0.285271	-0.268847	-0.127648	0.043132	-0.085186	0.441633	-0.079318	0.071879	0.320358	-0.596664	-0.360845	-0.015093	-0.133684	-0.065972	-0.052891	-0.018369
stiffjaw	0.280436	-0.342437	-0.091677	-0.069719	-0.104620	0.069642	0.043363	0.159535	0.234908	0.197374	0.297677	-0.139044	0.191435	0.681828	0.205319	-0.043869
cracking	0.292009	-0.247290	-0.074055	0.156963	-0.259475	-0.336082	0.300332	-0.381699	-0.442613	-0.013500	-0.385222	-0.119478	-0.137423	0.154129	0.049652	0.058485
painslp	0.227064	0.358364	-0.243535	0.054919	-0.030232	0.173403	0.115205	0.140552	-0.117995	0.127368	0.041832	-0.203902	-0.120288	0.197145	-0.750863	-0.055505
painrout	0.198153	0.336595	-0.218775	0.135315	-0.215945	0.084770	-0.489103	-0.272218	0.133010	0.291729	-0.320402	0.290608	0.320137	0.078864	0.099386	0.075037
paintab	0.233785	0.316863	-0.242193	0.014765	-0.105963	-0.157121	0.251882	-0.044251	-0.121796	-0.529605	0.417268	0.249901	0.370943	-0.048746	0.113112	-0.010063
amhdache	0.199459	0.364375	-0.247477	0.091466	-0.082925	0.194076	0.225617	0.090419	0.098658	0.217802	0.076023	-0.243937	-0.453650	-0.204928	0.532702	-0.013985
earpain	0.235574	0.124934	0.429407	-0.093774	-0.115198	0.102530	0.133328	0.127526	-0.012789	0.062931	-0.178233	-0.517339	0.551524	-0.261270	0.043078	0.007966
grind	0.236269	0.064136	0.417797	0.049008	-0.049183	0.066717	-0.306672	-0.586415	0.087019	-0.118554	0.430367	-0.187337	-0.255313	0.008212	-0.080110	-0.068550
clampset	0.238641	0.111842	0.396348	-0.055041	0.021005	0.286707	-0.179585	0.355671	-0.569722	-0.047839	-0.028047	0.329245	-0.161024	0.196774	0.161856	0.046039
sorejaw	0.213943	0.146994	0.413948	-0.033710	0.022567	-0.121332	0.485507	-0.033443	0.464156	0.141063	-0.152340	0.467846	-0.113577	0.060430	-0.115175	0.028672
facepain	0.223321	0.034578	-0.108697	0.011855	0.897498	0.098932	0.073952	-0.239471	-0.072579	0.012689	-0.124943	-0.054199	0.130068	0.088283	0.098626	-0.021351
dayslost	0.319485	0.127034	0.001290	-0.153982	0.135138	-0.672783	-0.379996	0.346407	0.163044	-0.148773	-0.101099	-0.158484	-0.187584	0.003271	0.004365	0.059272

下图中 EIgvalue 列表示按大小顺序排列的变量标准化变量的相关系数矩阵按从大到小顺序排列的特征值，cumulative→表示前 n 个特征值的累计方差贡献率，可理解为对矩阵的代表程度，代表程度越高，表明对矩阵的解释程度越好，一般经验值 85%以上，本示例达到 85% 择选前四个特征值，故需要选取 4 个主成分。

Eigenvalues of the Correlation Matrix				
	Eigenvalue	Difference	Proportion	Cumulative
1	5.11930088	1.61237495	0.3200	0.3200
2	3.50692593	0.81737841	0.2192	0.5391
3	2.68954752	1.67910964	0.1681	0.7072
4	1.01043788	0.17881992	0.0632	0.7704
5	0.83161796	0.27104095	0.0520	0.8224
6	0.56057701	0.15679169	0.0350	0.8574
7	0.40378532	0.05797567	0.0252	0.8826
8	0.34580965	0.03794316	0.0216	0.9043
9	0.30786649	0.02655180	0.0192	0.9235
10	0.28131469	0.06328723	0.0176	0.9411
11	0.21802746	0.02577786	0.0136	0.9547
12	0.19224960	0.00838066	0.0120	0.9667
13	0.18386894	0.02383078	0.0115	0.9782
14	0.16003816	0.04121150	0.0100	0.9882
15	0.11882666	0.04902082	0.0074	0.9956
16	0.06980584		0.0044	1.0000



2) 使用最大似然法(Maximum likelihood, ML)对 Pain 数据做因子分析的结果如下:

Iteration	Criterion	Ridge	Change	Communalities															
1	0.7754338	0.0000	0.1282	0.03092	0.91923	0.91583	0.62570	0.78845	0.58026	0.88822	0.65243	0.71419	0.82531	0.82432	0.68338	0.68449	0.69288	0.22421	0.48143
2	0.7740863	0.0000	0.0072	0.02962	0.91880	0.91876	0.62287	0.78377	0.58056	0.89062	0.65237	0.70695	0.83060	0.82727	0.67776	0.68967	0.69425	0.22962	0.47557
3	0.7740705	0.0000	0.0010	0.02936	0.91920	0.91859	0.62266	0.78373	0.57978	0.89117	0.65226	0.70641	0.83075	0.82804	0.67758	0.68959	0.69382	0.23001	0.47458

Factor Pattern			
	Factor1	Factor2	Factor3
isub	0.03361	-0.09946	0.13533
hurtchew	0.94735	-0.13833	-0.05095
hurtwide	0.93619	-0.17602	-0.10568
noise	0.77930	-0.08267	-0.09204
stiffjaw	0.85438	-0.22525	-0.05459
cracking	0.75867	-0.06416	-0.01245
painslp	0.18623	0.89645	-0.22994
painrout	0.13915	0.77207	-0.19179
paintab	0.21841	0.78457	-0.20765
amhdache	0.13038	0.86842	-0.24418
earpain	0.25080	0.33486	0.80809
grind	0.29059	0.22058	0.73786
clampset	0.27601	0.31724	0.71612
sorejaw	0.18371	0.33449	0.74041
facepain	0.39391	0.26689	-0.06030
dayslost	0.48253	0.46161	0.16913

Variance Explained by Each Factor		
Factor	Weighted	Unweighted
Factor1	30.4948888	4.45192760
Factor2	18.5318855	3.54876045
Factor3	10.3034111	2.52680792

Preliminary Eigenvalues: Total = 45.2159309 Average = 2.82599568				
	Eigenvalue	Difference	Proportion	Cumulative
1	23.6245806	9.4976306	0.5225	0.5225
2	14.1269501	6.0286378	0.3124	0.8349
3	8.0983122	7.4476756	0.1791	1.0140
4	0.6506367	0.1025705	0.0144	1.0284
5	0.5480661	0.1872860	0.0121	1.0405
6	0.3607801	0.1467202	0.0080	1.0485
7	0.2140599	0.1896377	0.0047	1.0532
8	0.0244221	0.0680811	0.0005	1.0538
9	-0.0436589	0.0651818	-0.0010	1.0528
10	-0.1088407	0.1392365	-0.0024	1.0504
11	-0.2480772	0.0812291	-0.0055	1.0449
12	-0.3293063	0.0459708	-0.0073	1.0376
13	-0.3752771	0.0100620	-0.0083	1.0293
14	-0.3853391	0.0730504	-0.0085	1.0208
15	-0.4583895	0.0245987	-0.0101	1.0107
16	-0.4829882		-0.0107	1.0000

3 factors will be retained by the PROPORTION criterion.

使用最大似然估计法后得到如上图所示的结果，结果表明，Eigenvalue 值大于 1 的因子，应该是我们最终得到的，因此在此数据集中应该选取 3 个因子。

3) 对因子做正交旋转(Varimax)，得到的结果如下：

The FACTOR Procedure			
Rotation Method: Varimax			
Orthogonal Transformation Matrix			
	1	2	3
1	0.96904	0.14268	0.20150
2	-0.19994	0.93231	0.30136
3	-0.14486	-0.33232	0.93198

Rotated Factor Pattern			
	Factor1	Factor2	Factor3
isub	0.03285	-0.13290	0.10292
hurtchew	0.95305	0.02313	0.10172
hurtwide	0.95771	0.00459	0.03710
noise	0.78504	0.06470	0.04633
stiffjaw	0.88087	-0.06996	0.05340
cracking	0.74982	0.05257	0.12193
painslp	0.03454	0.93876	0.09338
painrout	0.00826	0.80340	0.08196
paintab	0.08486	0.83163	0.08692
amhdache	-0.01191	0.90939	0.06041
earpain	0.05902	0.07944	0.90457
grind	0.13060	0.00191	0.81269
clampset	0.10030	0.09717	0.81863
sorejaw	0.00389	0.09201	0.82787
facepain	0.33709	0.32506	0.10361
dayslost	0.35080	0.44301	0.39396

Variance Explained by Each Factor		
Factor	Weighted	Unweighted
Factor1	29.5929483	4.05695233
Factor2	17.8667245	3.40154073
Factor3	11.8705126	3.06900290

Final Communalities and Variable Weights		
Total Communality: Weighted = 59.330185 Unweighted = 10.527496		
Variable	Communality	Weight
isub	0.02933521	1.0302448
hurtchew	0.91919491	12.3764976
hurtwide	0.91860604	12.2841710
noise	0.62261342	2.6501242
stiffjaw	0.78368377	4.6239274
cracking	0.57985588	2.3797138
painslp	0.89118148	9.1887525
painrout	0.65224269	2.8757253
paintab	0.70636992	3.4061513
amhdache	0.83077802	5.9083659
earpain	0.82804498	5.8153938
grind	0.67752911	3.1015547
clampset	0.68965740	3.2215295
sorejaw	0.69384365	3.2660720
facepain	0.23003025	1.2987180
dayslost	0.47452924	1.9032420

由于我们采用的是方差最大正交旋转方法，因此旋转后的因子仍然为正交。从上图中可以看出，在旋转前，变量在各个因子上均有载荷，而且载荷明显；经过旋转后，变量则均在某一因子上有高载荷，而在另一因子上的载荷低，而这正是我们希望看到的。

郑重声明

本作业由作者独立完成。抄袭行为在任何情况下都是不能容忍的 (COPY is strictly prohibited under any circumstances)！由抄袭所产生的一切后果由抄袭者承担，勿谓言之不预也。



陈麒先