题目一

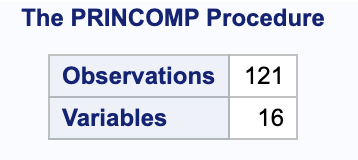
Pain数据集包含了121位颞下颌关节疾病患者抱怨面部和下颌疼痛的数据。受试者用数字1到9对痛感评级，对14个问题作出回答。数据集中包含如下变量：

* 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?
* Stiffjaw Stiff 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?

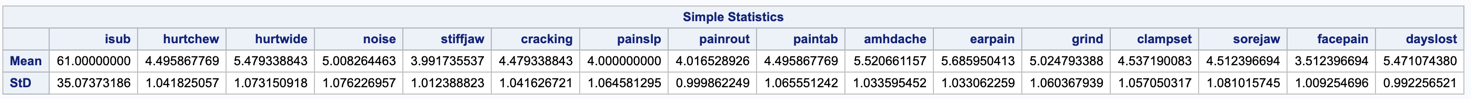
1. 对Pain数据做主成分分析，解释主成分分析的结果。
2. 使用最大似然法(Maximum likelihood, ML)对Pain数据做因子分析，你得到几个因子？得出该结论的依据是？
3. 对因子做正交旋转(Varimax)，解释旋转后得到的因子。

【答】本次作业所有源代码见附件solution.sas

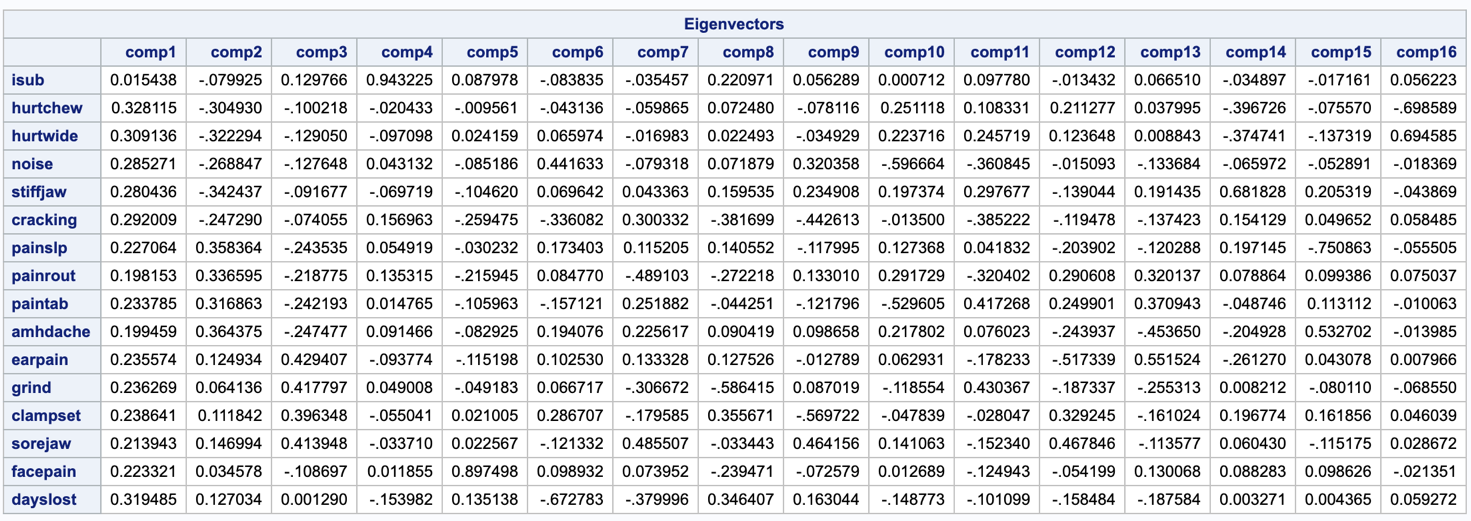
1） 使用PCA对pain数据分析的结果如下：



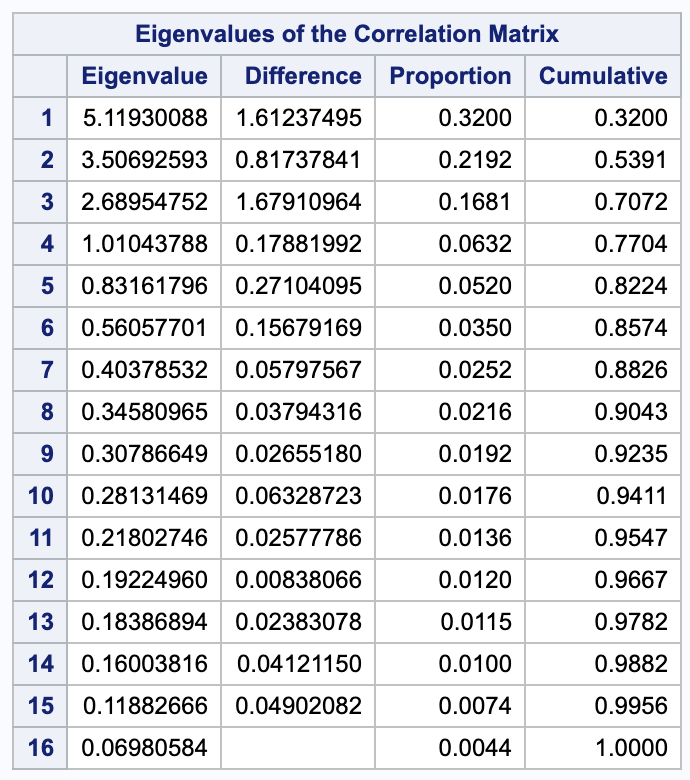
上图表明我们对16个变量成功观测到了121个观测值。下图为各变量均值与方差。

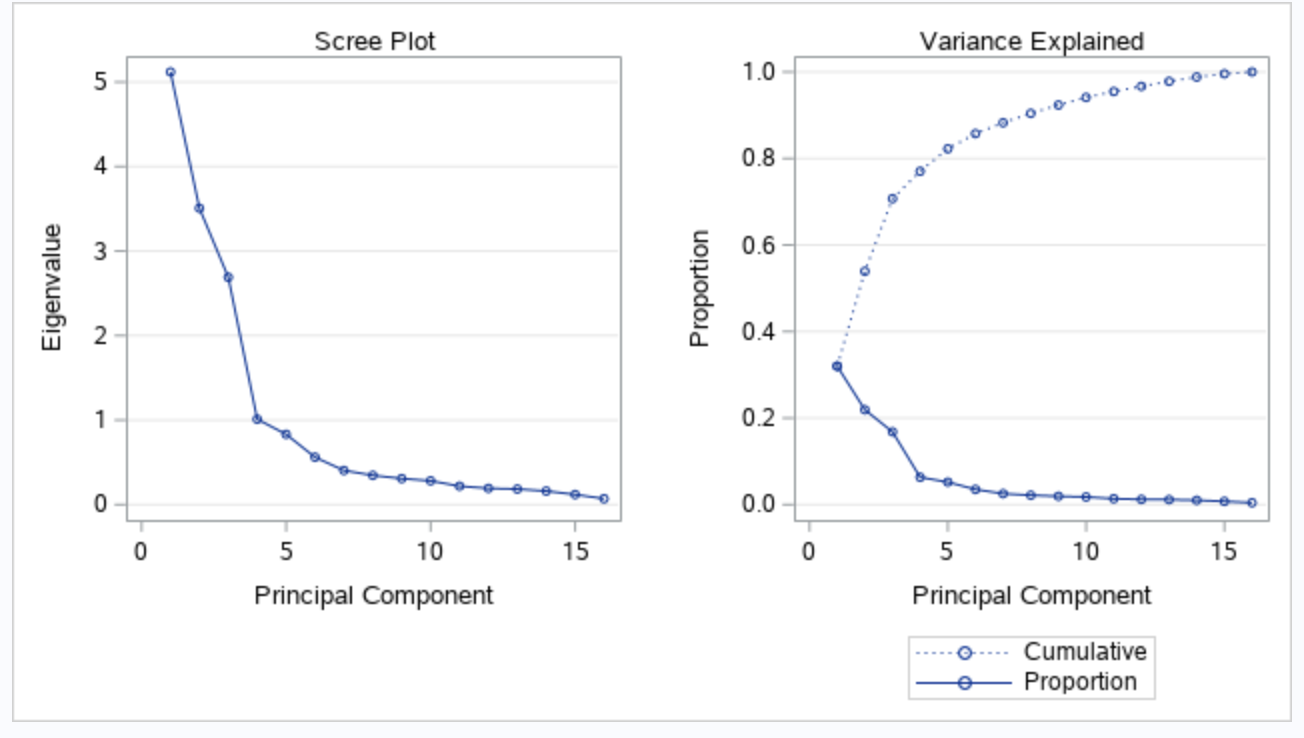


下图的每一列是第compi个主成分变量对16个原始自变量的解释力度。

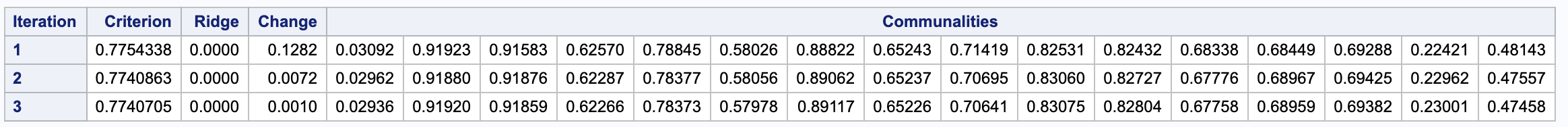


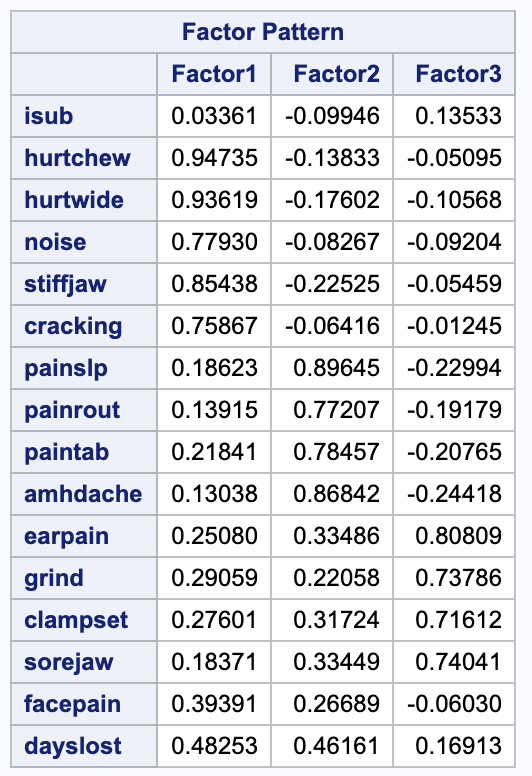
下图中EIgvalue列表示按大小顺序排列的变量标准化变量的相关系数矩阵按从大到小顺序排列的特征值，cumulative→表示前n个特征值的累计方差贡献率，可理解为对矩阵的代表程度，代表程度越高，表明对矩阵的解释程度越好，一般经验值85%以上，本示例达到85%择选前四个特征值，故需要选取4个主成分。

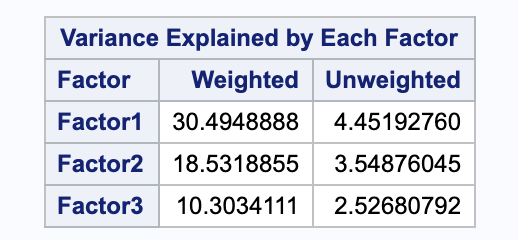


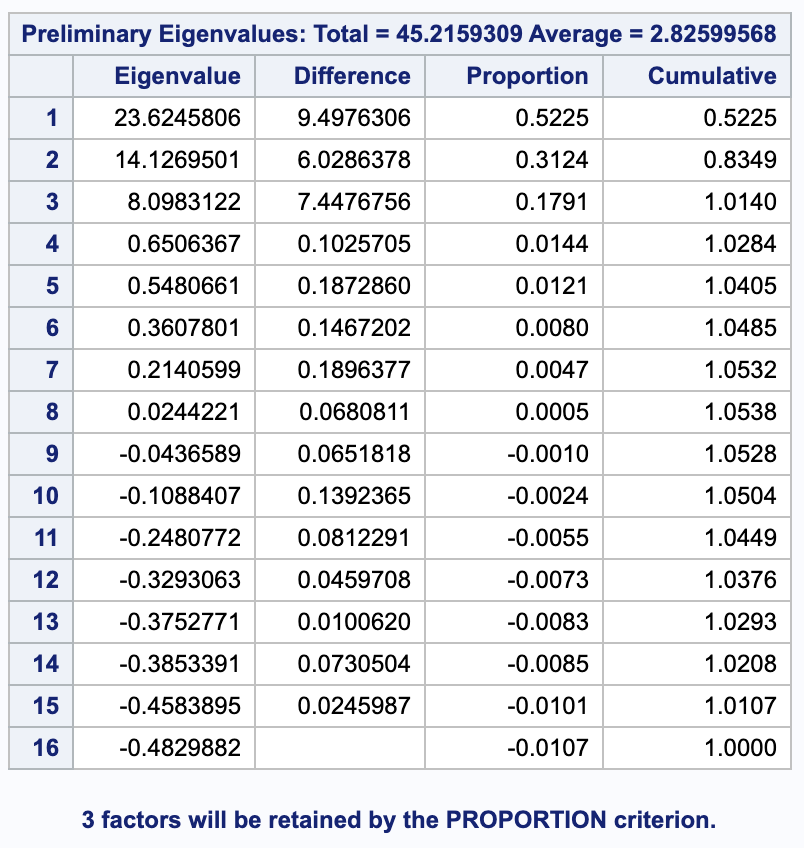


2）使用最大似然法(Maximum likelihood, ML)对Pain数据做因子分析的结果如下：



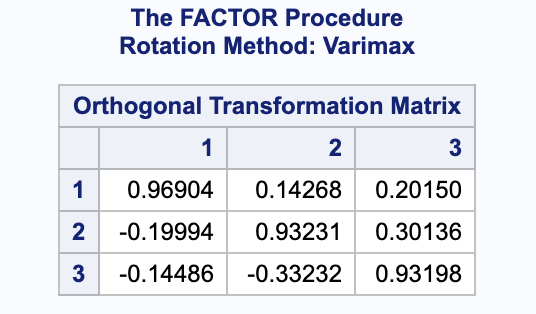


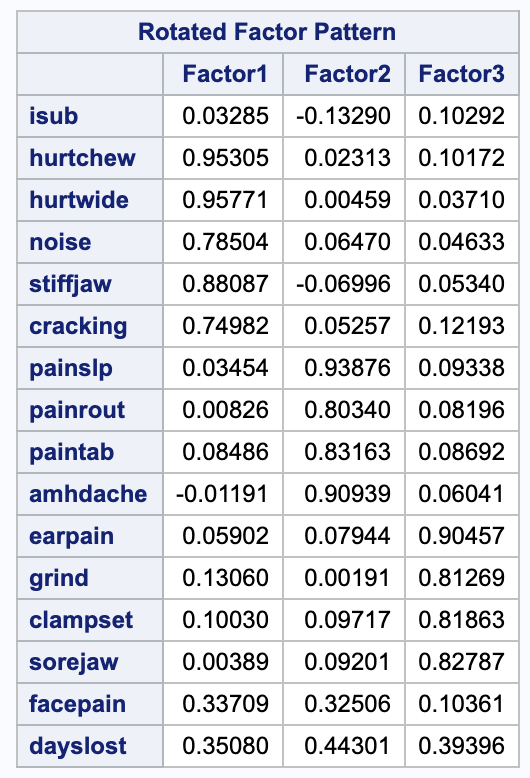


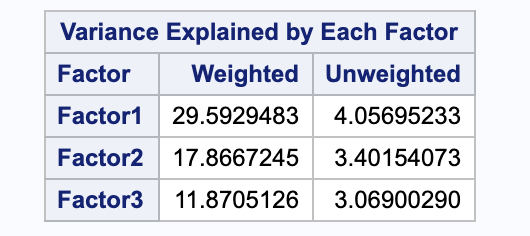


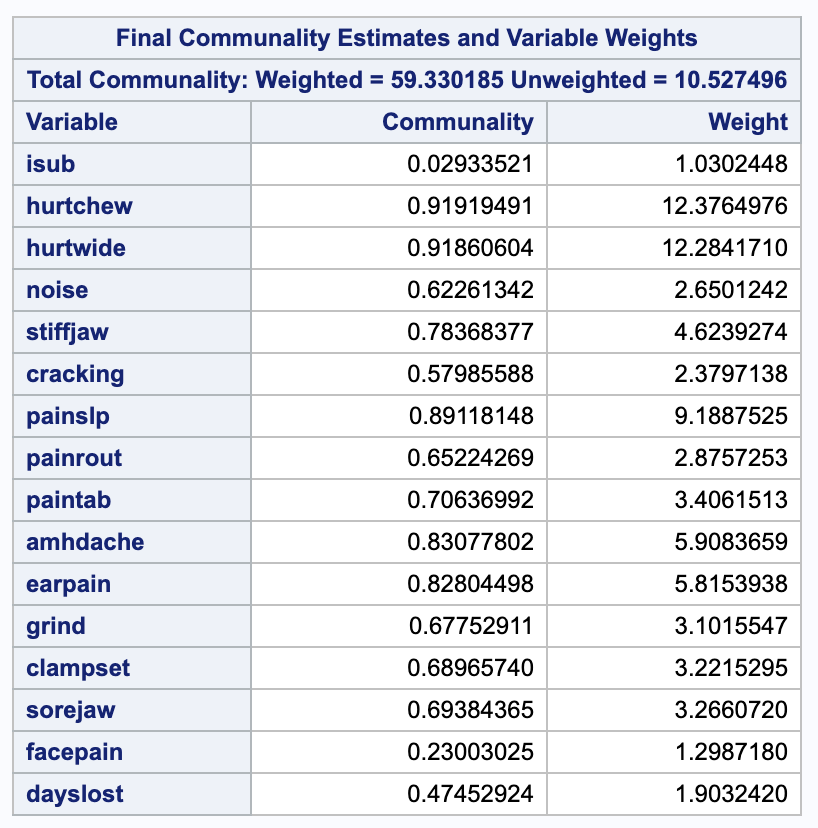
使用最大似然估计法后得到如上图所示的结果，结果表明，Eigenvalue值大于1的因子，应该是我们最终得到的，因此在此数据集中应该选取3个因子。

3）对因子做正交旋转(Varimax)，得到的结果如下：









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**