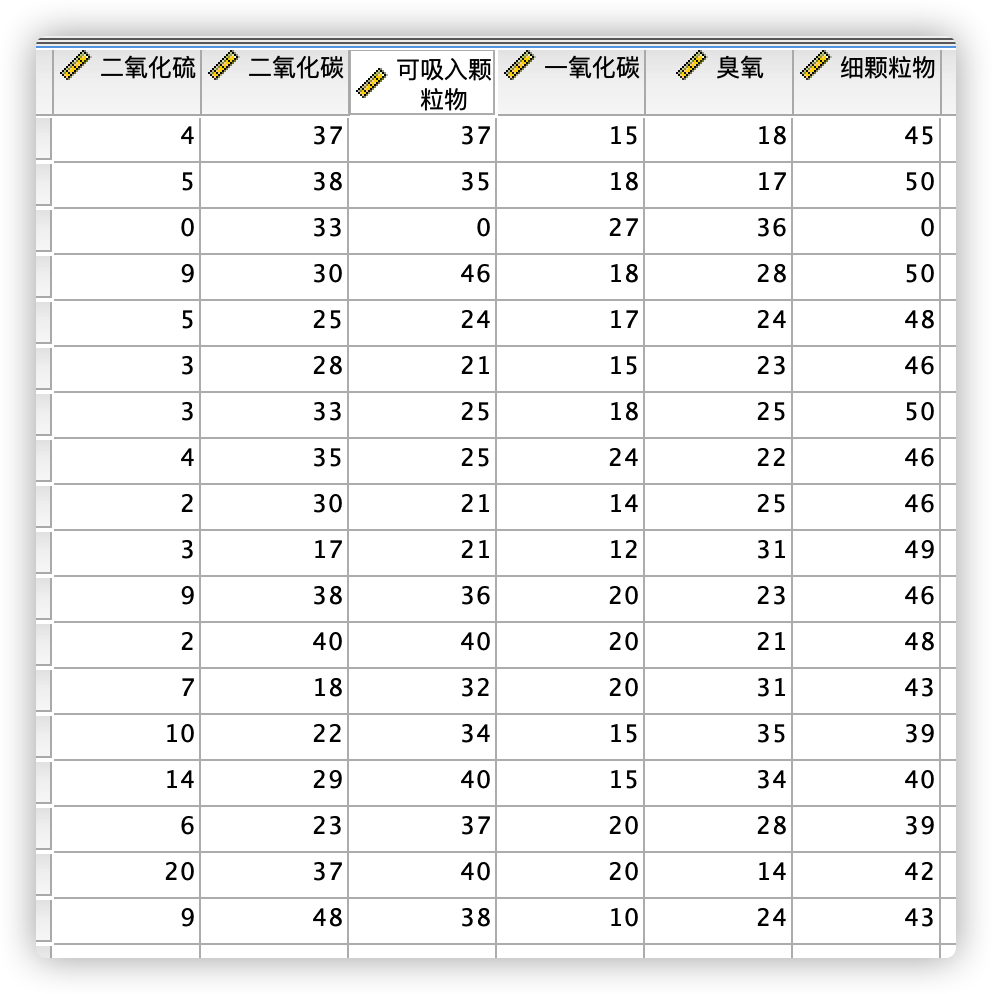
**因子分析**

**1202110887 徐鸿飞**

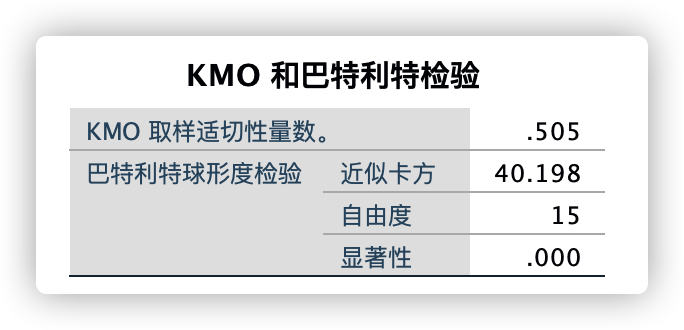
**题目**（来源网络）：下图数据是一份某城市的空气质量数据，一共6个变量，分别是：二氧化硫、二氧化氮、可吸入颗粒物、一氧化碳、臭氧、细颗粒物。在SPSS中打开数据如下：



**解答：**

主成分分析是试图寻找原有变量的一个线性组合。这个线性组合方差越大，那么该组合所携带的信息就越多。也就是说，主成分分析就是将原始数据的主要成分放大。

因子分析，它是假设原有变量的背后存在着一个个隐藏的因子，这个因子可以可以包括原有变量中的一个或者几个，因子分析并不是原有变量的线性组合。



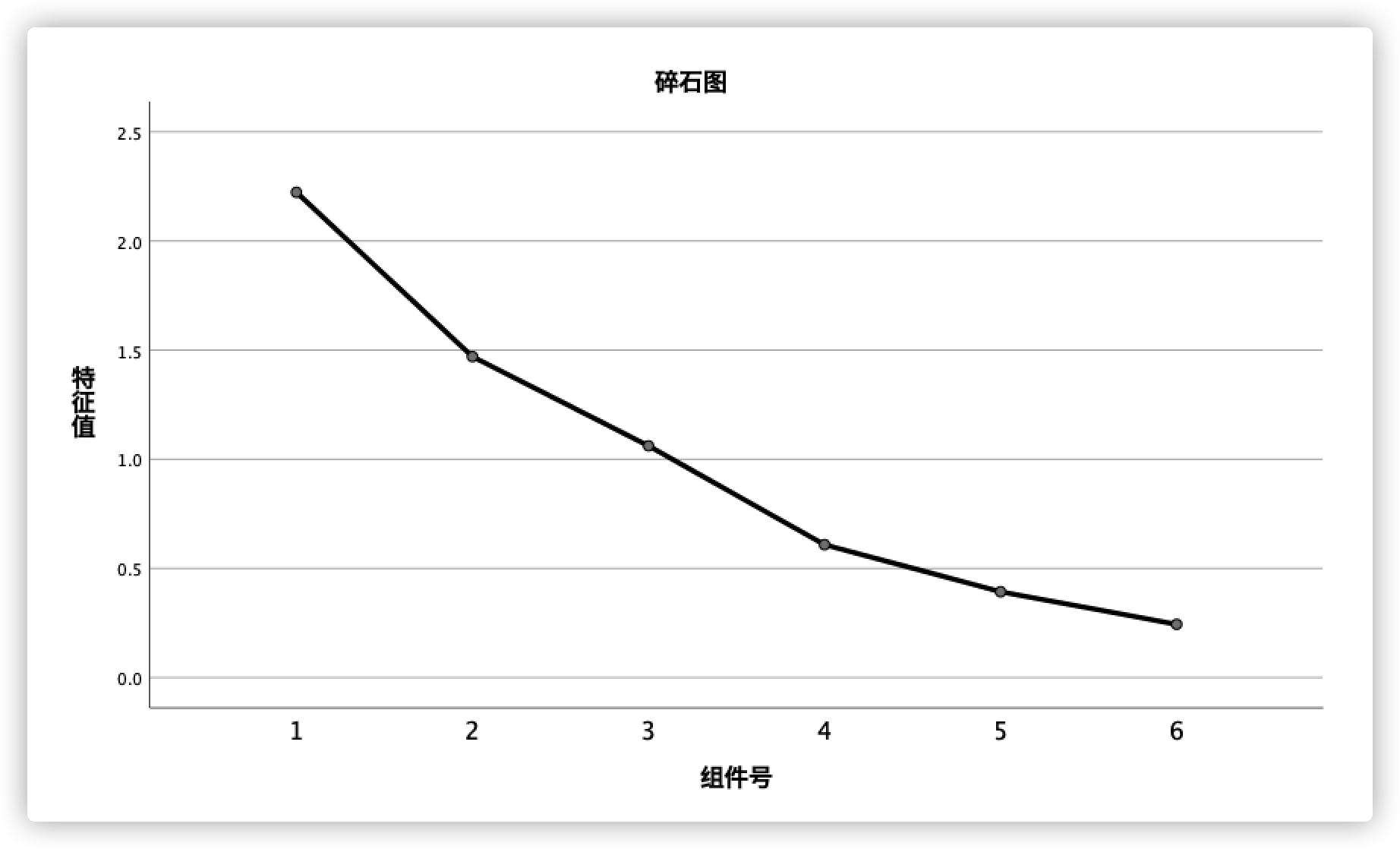
首先是KMO的值为0.505，大于阈值0.5，所以说明了变量之间是存在相关性的，符合要求；然后是Bartlett球形检验的结果，在这里只需要看显著性这一项，其值为0.000，所以小于0.05。那么也就是说，这份数据是可以进行因子分析的。



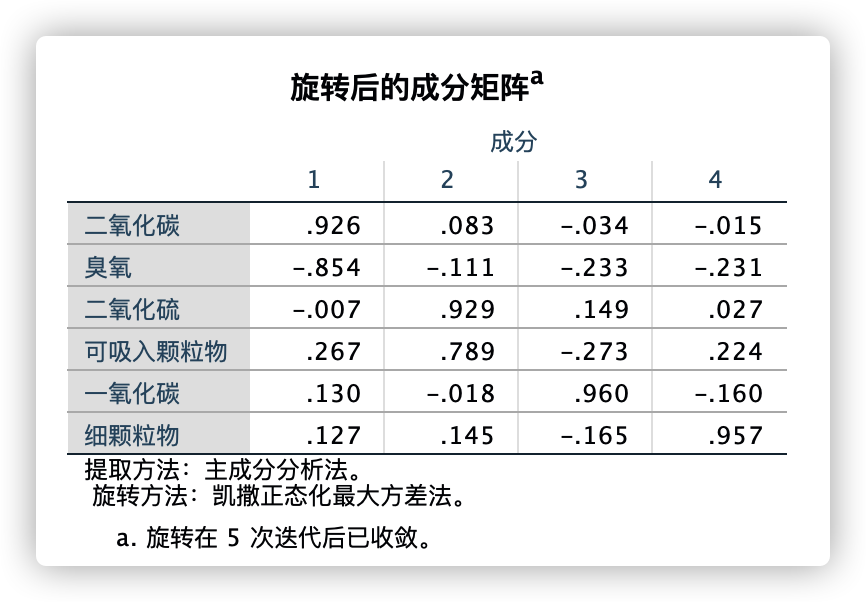
公因子方差表的意思就是，每一个变量都可以用公因子表示，而公因子究竟能表达多少呢，其表达的大小就是公因子方差表中的“提取”，“提取”的值越大说明变量可以被公因子表达的越好，一般大于0.5即可以说是可以被表达，但是更好的是要求大于0.7才足以说明变量能被公因子表的很合理。在本例中可以看到，“提取”的值都是大于0.7的，所以变量可以被表达的很不错。



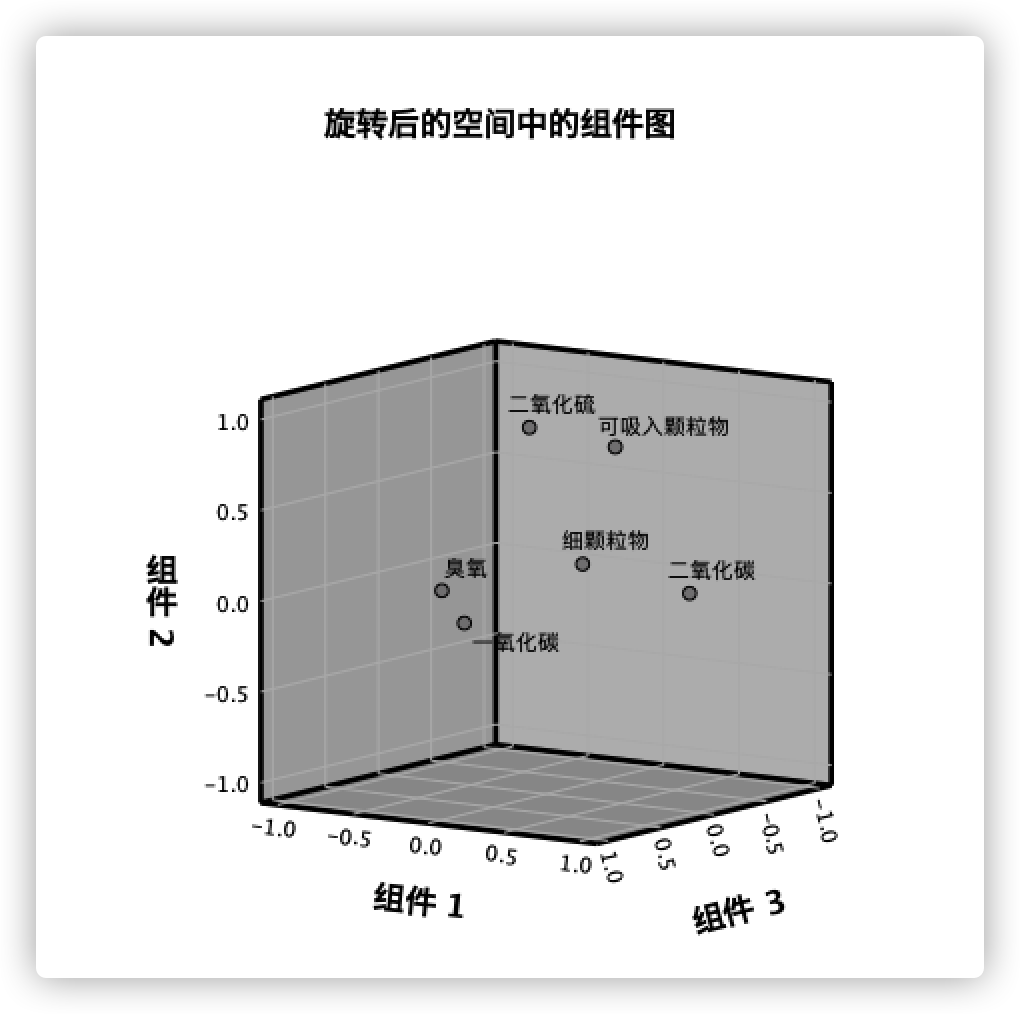
解释地总方差就是看因子对于变量解释的贡献率（可以理解为究竟需要多少因子才能把变量表达为100%）。这张表只需要看图中红框的一列，表示的就是贡献率，蓝框则代表四个因子就可以将变量表达到了89.382%，说明表达的还是不错的，一般都要表达到85%以上才可以，否则就要调整因子数据。



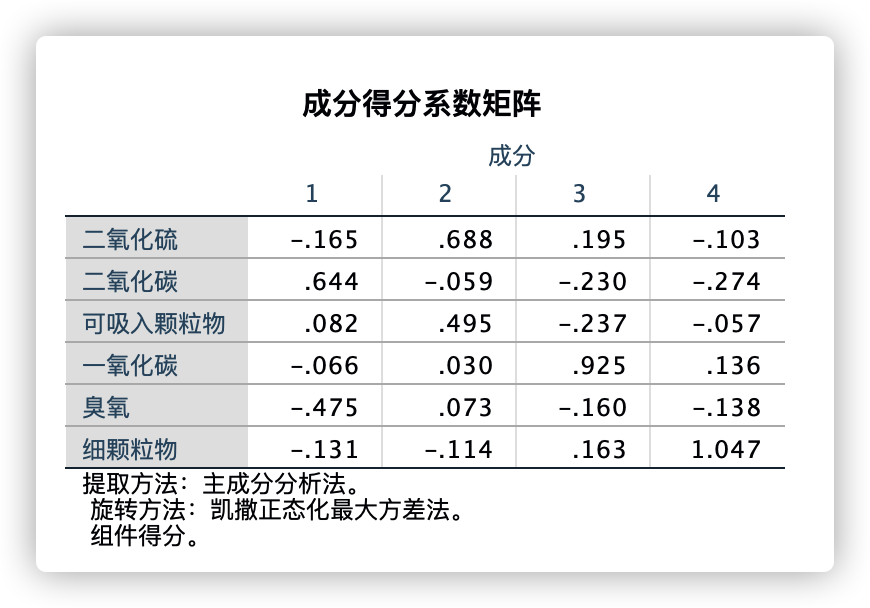
碎石图变得平缓说明提取的因子数量比较合适，再看碎石图，也确实就是四个因子之后折线就变得平缓了。



这一张表是用来看哪些变量可以包含在哪些因子里，一列一列地看：第一列，最大的值为0.926，分别对应的是二氧化碳。第二列，最大值为0.929、0.789对应着二氧化硫和可吸入颗粒物。第三列，最大值为0.960，对应着一氧化碳。第四列，最大值为0.957，分别对应着细颗粒物。



如果是有2个主成分，组件图就是二维的。有3个主成分的话，组件图就是三维的。如果提取了4个主成分，那么最多也就显示三维图。



因子得分矩阵表示各项指标变量与提取的公因子之间的关系，在某一公因子上得分高，表明该指标与该公因子之间关系越密切。假设X1为指标变量1，a11、a12、a13分别为与变量X1在同一行的因子载荷，F1、F2、F3分别为提取的公因子；通过因子得分矩阵可以得到公因子的线性组合，如F1=a11\*X1+a21\*X2+a31\*X3。

所以此处就是表示提取出的四个因子与原数据的表达式。