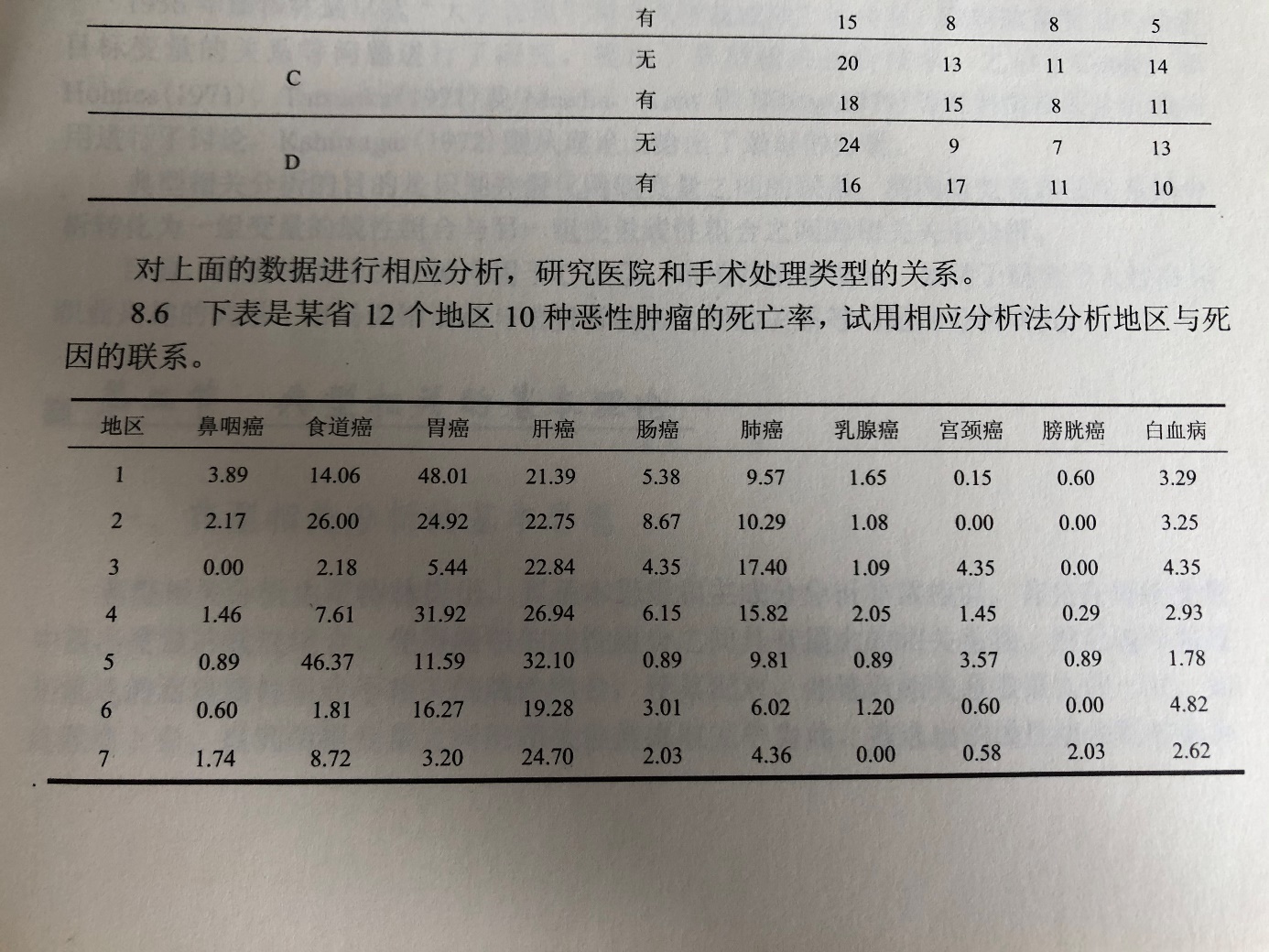
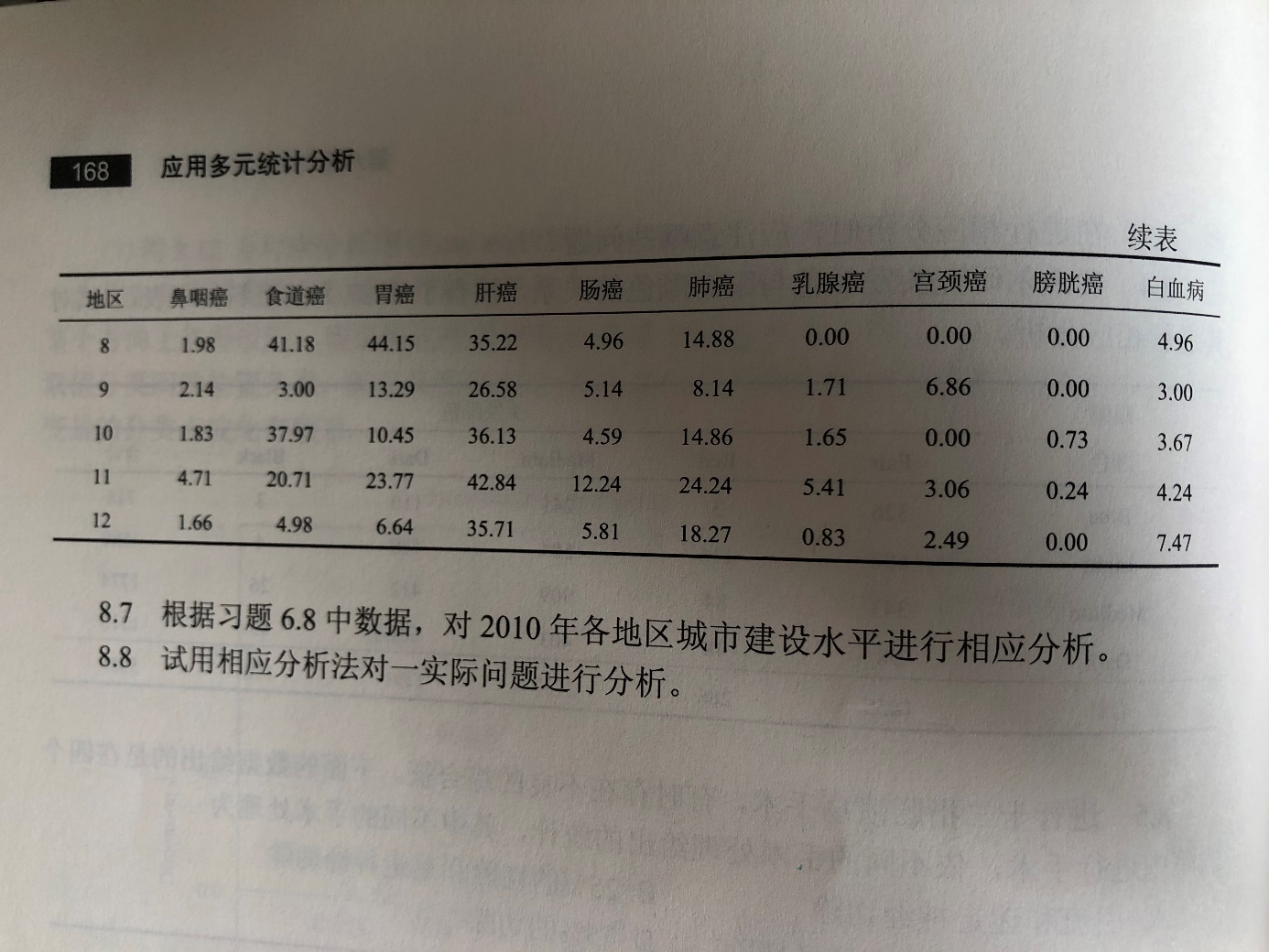
华东理工大学2019 –2020 学年第 二 学期

《多元统计学》实验报告

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 实验名称 | 实验7 因子分析 | | | | | | | | | | |
| 专 业 | | 数学与应用数学 | 姓名 | 梁天一 | 学号 | 10172911 | | 组名/组号 | |  |
| 实验报告上交时间 | | 6月7日 | | 实验地点 |  | | 指导教师 | |  | |

|  |
| --- |
| 实验目的/要求 |
| 1、掌握因子分析 |
| 实验内容 |
| 书P147 7.7题 |
| 实验总结 |
| 能够通过这次实验解了因子分析的原理以及操作步骤。  能够分清主成分分析与因子分析的区别  能够通过因子分析了解到降维处理数据并用有意义的方式命名新的因子。  实验过程中需要注意因子个数的选择。 |
| 教师批阅： 实验成绩：  教师签名: 日期： |





通过不可描述的手段,我得到了数据

不可描述的手段如下

多元统计数据高速处理方法

天若ocr把图片变成文本

13.8914.0648.0121.395.389.57 1.650.150.603.29

22.1726.0024.9222.758.6710.291.080.000.003.25

30.002.185.4422.844.3517.401.094.350.004.35

41.467.6131.9226.946.1515.822.051.450.292.93

50.8946.37 11.5932.100.899.810.893.570.891.78

60.601.8116.2719.283.016.021.200.600.004.82

1.748.723.2024.702.034.360.000.582.032.62

这些文本没有间隔,我们把所有小数点后2位提取出来,在后面加个\t分割:

粘贴到vscode上,然后正则替换(\.(\d){0,2}) 为 $1\t  
最后把空格替换为无

1 3.89 14.06 48.01 21.39 5.38 9.57 1.65 .15 .60 3.29

2 2.17 26.00 24.92 22.75 8.67 10.29 1.08 .00 .00 3.25

3 .00 2.18 5.44 22.84 4.35 17.40 1.09 4.35 .00 4.35

4 1.46 7.61 31.92 26.94 6.15 15.82 2.05 1.45 .29 2.93

5 .89 46.37 11.59 32.10 .89 9.81 .89 3.57 .89 1.78

6 .60 1.81 16.27 19.28 3.01 6.02 1.20 .60 .00 4.82

7 1.74 8.72 3.20 24.70 2.03 4.36 .00 .58 2.03 2.62

8 1.98 41.18 44.15 35.22 4.96 14.88 .00 .00 .00 4.96

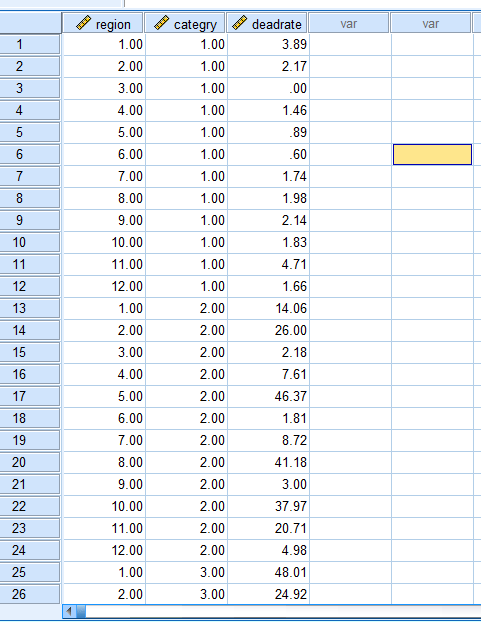
9 2.14 3.00 13.29 26.38 5.14 8.14 1.71 6.86 .00 3.00

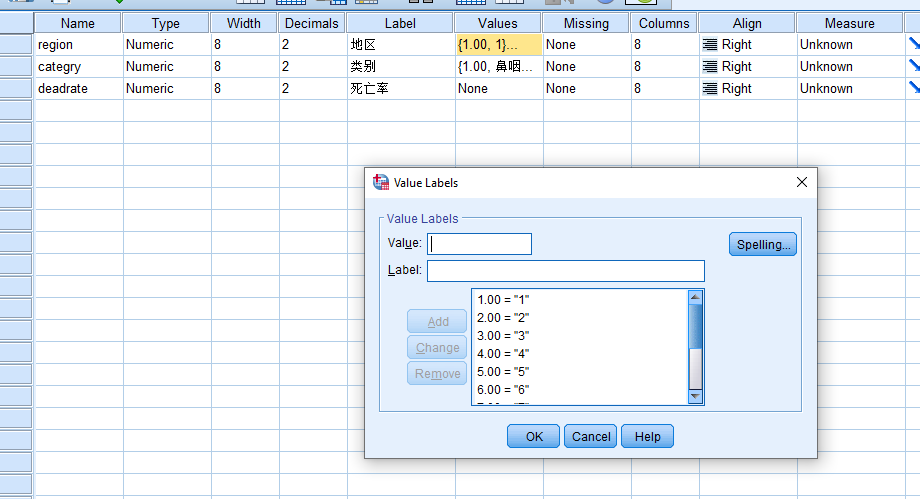
10 1.83 37.97 10.45 36.13 4.59 14.86 1.65 .00 .73 3.67

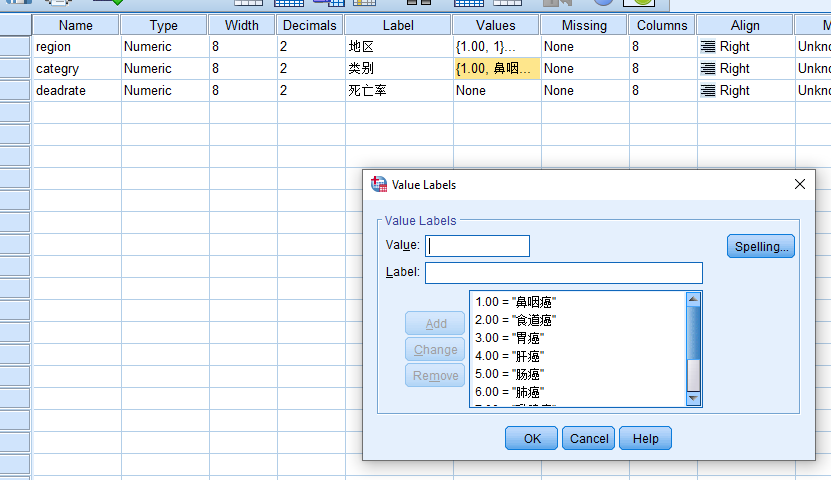
11 4.71 20.71 23.77 42.84 12.24 24.24 5.41 3.06 .24 4.24

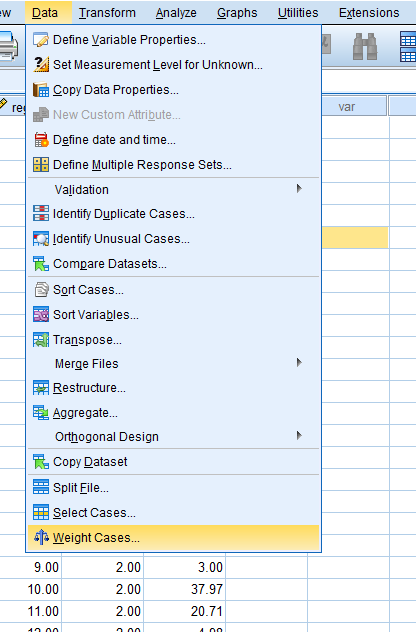
12 1.66 4.98 6.64 35.71 5.81 18.27 .83 2.49 .00 7.47

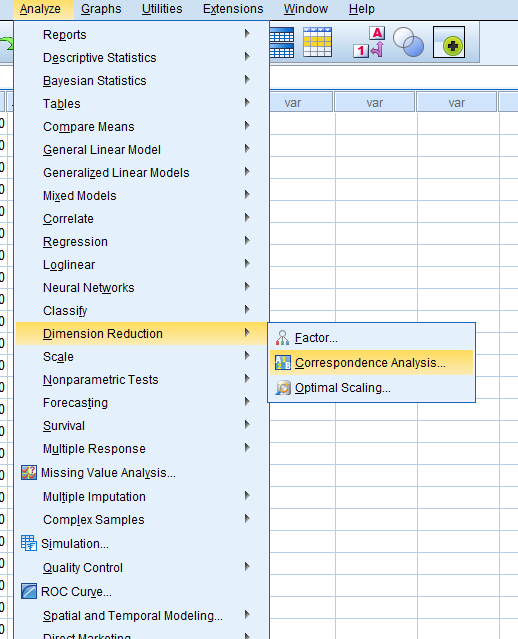
之后新建数据集,对数据做一个变换

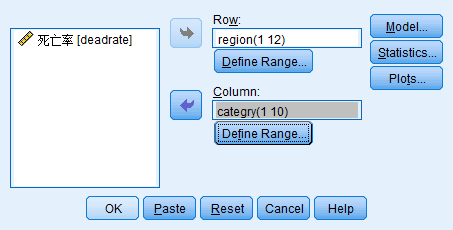


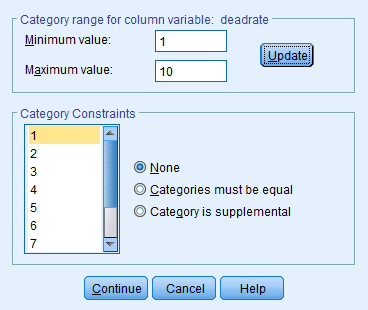


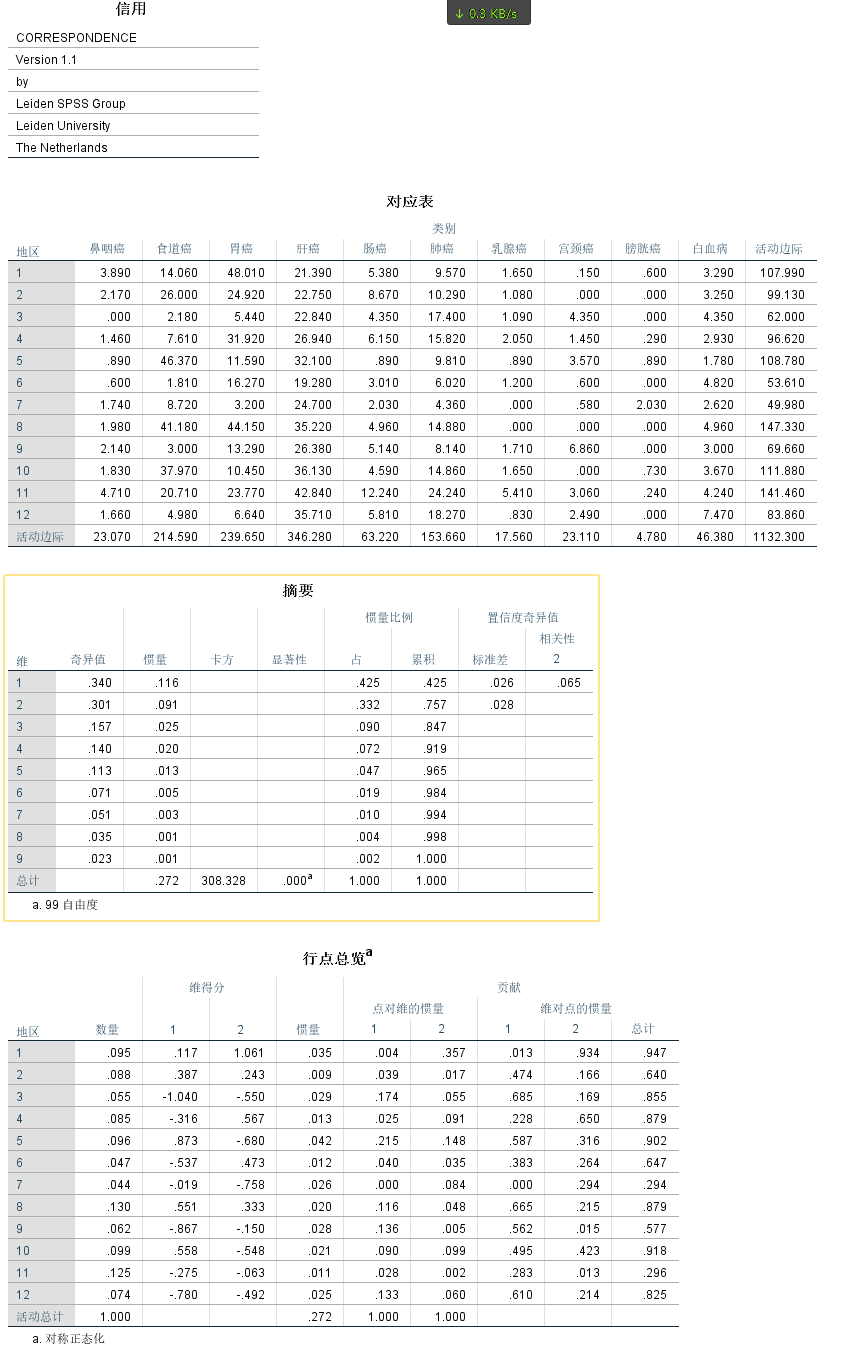


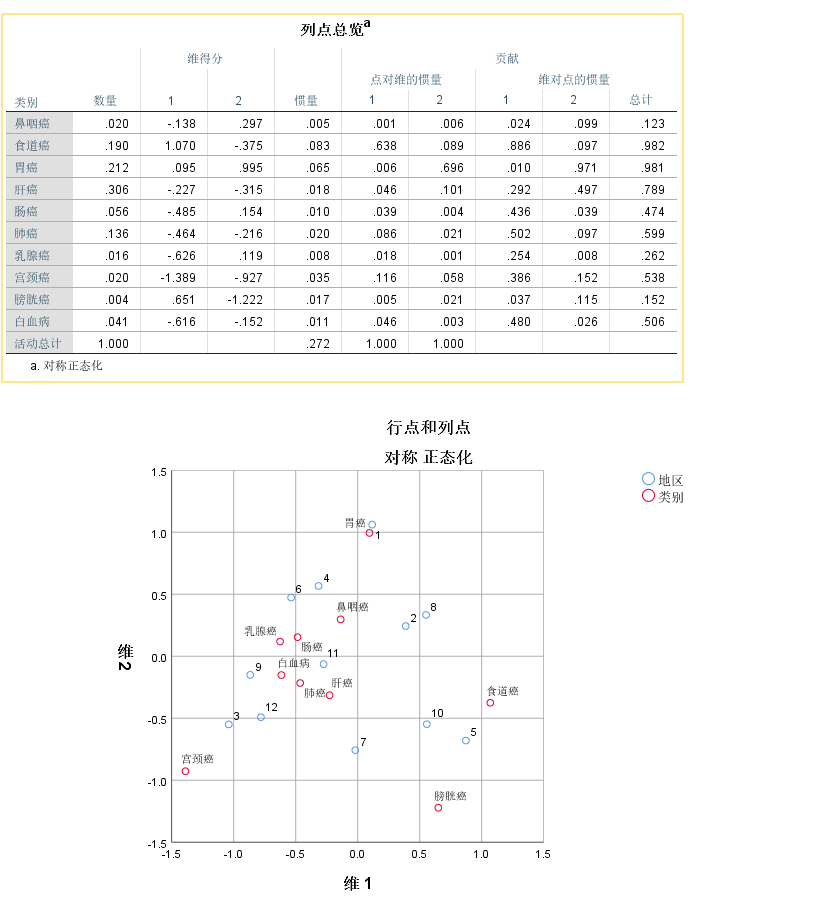












卡方是308.328066,显著性为2.4486E-23<0.05 因此两个变量不独立,地区和癌症有很强的相关性

9个惯量占比分别为



前4个惯量占比大于0.85,所以选取4个维度可以表示两个变量之间的关系

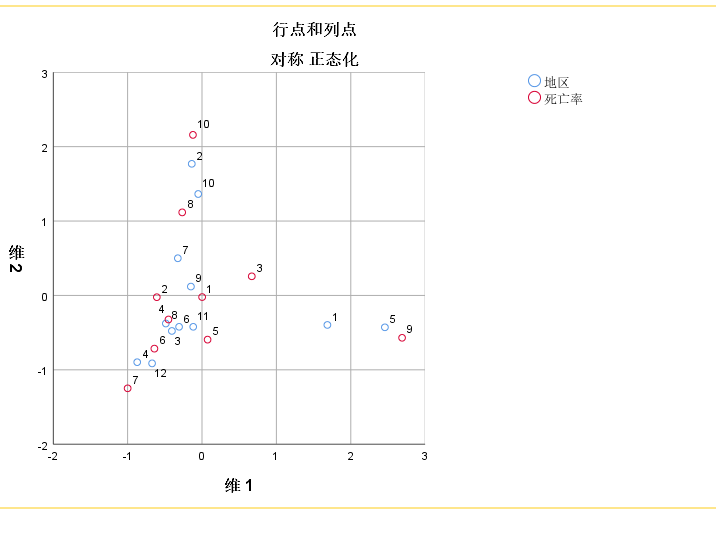
行点总览为对应分析行变量“地区”取值分布表的一部分，“第二列”是边缘概率；“维得分”是变量“地区”的各个状态在两个维度上的坐标值；“贡献”表示行变量的每一状态对每一维度特征值的贡献及每一维度对行变量的各状态的特征值的贡献，由此可以更好理解维度的来源及含义。

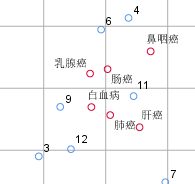
列点总览同理

最后那个图是最重要的,根据维得分得来

可以通过横轴和纵轴的区分情况看到在两个维度中,变量区别较大,两个维度都表达了更多信息

大部分地区都集中在几个主要疾病边上





而宫颈癌症膀胱癌食道癌不是特别重要,致死率相对较低地区5 10 容易受到食道癌与膀胱癌困扰

地区1与胃癌关系较大