第七章 主成分分析

一、中学生身体四项指标的主成分分析

1、30名中学生身体四项指标数据

| **Obs** | **x1** | **x2** | **x3** | **x4** |
| --- | --- | --- | --- | --- |
| **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 | 76 |
| **7** | 153 | 43 | 76 | 83 |
| **8** | 150 | 43 | 77 | 79 |
| **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 | 30 | 65 | 75 |
| **23** | 157 | 48 | 80 | 88 |
| **24** | 151 | 36 | 74 | 80 |
| **25** | 144 | 36 | 68 | 76 |
| **26** | 141 | 30 | 67 | 76 |
| **27** | 139 | 32 | 68 | 73 |
| **28** | 148 | 38 | 70 | 78 |

2、SAS程序

data chapter7;

input x1 x2 x3 x4;

cards;

148 41 72 78

139 34 71 76

160 49 77 86

149 36 67 79

159 45 80 86

142 31 66 76

153 43 76 83

150 43 77 79

151 42 77 80

139 31 68 74

140 29 64 74

161 47 78 84

158 49 78 83

140 33 67 77

137 31 66 73

152 35 73 79

149 47 82 79

145 35 70 77

160 47 74 87

156 44 78 85

151 42 73 82

147 30 65 75

157 48 80 88

151 36 74 80

144 36 68 76

141 30 67 76

139 32 68 73

148 38 70 78

;

proc print;

run;

proc princomp;

var x1 x2 x3 x4;

run;

3、SAS结果分析

| **简单统计量** | | | | |
| --- | --- | --- | --- | --- |
|  | **x1** | **x2** | **x3** | **x4** |
| **均值** | 148.7857143 | 38.71428571 | 72.35714286 | 79.39285714 |
| **StD** | 7.4104888 | 6.69359640 | 5.27196840 | 4.41662923 |
|  |  |  |  |  |

| **相关矩阵** | | | | |
| --- | --- | --- | --- | --- |
|  | **x1** | **x2** | **x3** | **x4** |
| **x1** | 1.0000 | 0.8805 | 0.7927 | 0.9351 |
| **x2** | 0.8805 | 1.0000 | 0.9098 | 0.8834 |
| **x3** | 0.7927 | 0.9098 | 1.0000 | 0.8002 |
| **x4** | 0.9351 | 0.8834 | 0.8002 | 1.0000 |

由相关系数矩阵可以看出x1 x2 x3 x4都有很强的相关性

| **相关矩阵的特征值** | | | | |
| --- | --- | --- | --- | --- |
|  | **特征值** | **差分** | **比例** | **累积** |
| **1** | 3.60198370 | 3.33721847 | 0.9005 | 0.9005 |
| **2** | 0.26476523 | 0.19601640 | 0.0662 | 0.9667 |
| **3** | 0.06874883 | 0.00424659 | 0.0172 | 0.9839 |
| **4** | 0.06450224 |  | 0.0161 | 1.0000 |

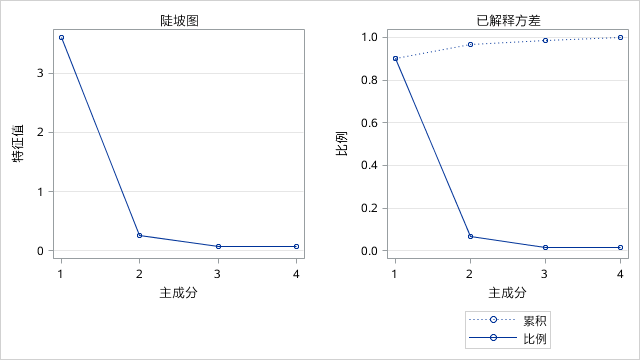
由第一主成分的特征值可以看出，第一主成分的占比已经达到0.9005，第二主成分占到0.0662.故这个可以有第一主成分和第二主成分很好的表示。

|  | **Prin1** | **Prin2** | **Prin3** | **Prin4** |
| --- | --- | --- | --- | --- |
| **x1** | 0.501236 | -.483154 | -.016691 | 0.717667 |
| **x2** | 0.510107 | 0.247982 | -.796926 | -.207856 |
| **x3** | 0.485657 | 0.707030 | 0.492209 | 0.148246 |
| **x4** | 0.502684 | -.452963 | 0.349801 | -.647899 |

由特征向量可以看出

第一主成分prin1=0.501236x1+0.510107x2+0.485657x3+0.502684x4

第二主成分prin2=-0.483154x1+0.247982x2+0.707030x3-0.452963x4



二、服装定型分类问题

1、16项身体数据指标的相关阵

三、经济分析数据的主成分回归

1、经济分析数据

| **Obs** | **x1** | **x2** | **x3** | **y** |
| --- | --- | --- | --- | --- |
| **1** | 149.3 | 4.2 | 108.1 | 15.9 |
| **2** | 161.2 | 4.1 | 114.8 | 16.4 |
| **3** | 171.5 | 3.1 | 123.2 | 19.0 |
| **4** | 175.5 | 3.1 | 126.9 | 19.1 |
| **5** | 180.8 | 1.1 | 132.1 | 18.8 |
| **6** | 190.7 | 2.2 | 137.7 | 20.4 |
| **7** | 202.1 | 2.2 | 137.7 | 20.4 |
| **8** | 212.4 | 5.6 | 154.1 | 26.5 |
| **9** | 226.1 | 5.0 | 162.3 | 27.6 |
| **10** | 239.0 | 0.7 | 167.6 | 25.3 |

2、SAS程序

data chapter7;

input x1 x2 x3 y;

cards;

149.3 4.2 108.1 15.9

161.2 4.1 114.8 16.4

171.5 3.1 123.2 19.0

175.5 3.1 126.9 19.1

180.8 1.1 132.1 18.8

190.7 2.2 137.7 20.4

202.1 2.2 137.7 20.4

212.4 5.6 154.1 26.5

226.1 5.0 162.3 27.6

239.0 0.7 167.6 25.3

;

proc print;

run;

proc princomp;

var x1 x2 x3 ;

run;

3、结果分析

|  | **特征值** | **差分** | **比例** | **累积** |
| --- | --- | --- | --- | --- |
| **1** | 2.01882654 | 1.04615744 | 0.6729 | 0.6729 |
| **2** | 0.97266910 | 0.96416473 | 0.3242 | 0.9972 |
| **3** | 0.00850436 |  | 0.0028 | 1.0000 |

| **特征向量** | | | |
| --- | --- | --- | --- |
|  | **Prin1** | **Prin2** | **Prin3** |
| **x1** | 0.699328 | 0.092945 | 0.708733 |
| **x2** | -.164764 | 0.985771 | 0.033301 |
| **x3** | 0.695553 | 0.140062 | -.704691 |

故第一主成分prin1=0.699328x1-0.164764x2+0.695553x3

第二主成分prin2=0.92945x1+0.9985771+0.140032x3

在分析可以看到y=0.68998prin1+0.1993prin3

故y=-0.130+0.0727x1+0.6091x2+0.1062x3