主成分分析法

主成分分析（Principal Component Analysis，PCA）是将多个变量通过线性变换以选出较少个数重要变量，并尽可能多地反映原来变量信息的一种多元统计分析方法，又称主分量分析。也是数学上处理降维的一种方法。主成分分析是设法将原来众多具有一定相关性的指标（比如 P 个指标），重新组合成一组新的互相无关的综合指标来代替原来的指标。通常数学上的处理就是将原来 P 个指标作线性组合，作为新的综合指标。最经典的做法就是用 F1（选取的第一个线性组合，即第一个综合指标）的方差来表达，即 Var（F1）越大，表示 F1 包含的信息越多。因此在所有的线性组合中选取的 F1 应该是方差最大的，故称 F1 为第一主成分。如果第一主成分不足以代表原来 P 个指标的信息，再考虑选取 F2 即选第二个线性组合，为了有效地反映原来信息，F1 已有的信息就不需要再出现在 F2 中，用数学语言表达就是要求 cov（F1，F2）=0，则称 F2 为第二主成分，依此类推可以构造出第三、第四，……，第 P 个主成分。

主要作用

1. 主成分分析能降低所研究的数据空间的维数。即用研究 m 维的 Y 空间代替 p 维的 X 空间(m＜p)，而低维的 Y 空间代替高维的 X 空间所损失的信息很少。即使只有一个主成分 Y1（即 m＝1）时，这个 Y1 仍是使用全部 X 变量（p 个）得到的，例如要计算 Y1 的均值也得使用全部 x 的均值。在所选的前 m 个主成分中，如果某个 Xi 的系数全部近似于零的话，就可以把这个 Xi 删除，这也是一种删除多余变量的方法。

2. 有时可通过因子负荷 aij 的结论，弄清 X 变量间的某些关系。

3. 多维数据的一种图形表示方法。当维数大于 3 时不能画出几何图形，多元统计研究的问题大都多于 3 个变量。要把研究的问题用图形表示出来是不可能的。然而，经过主成分分析后，我们可以选取前两个主成分或其中某两个主成分，根据主成分的得分，画出 n 个样品在二维平面上的分布，由图形可直观地看出各样品在主分量中的地位，进而还可以对样本进行分类处理，可以由图形发现远离大多数样本点的离群点。

4. 由主成分分析法构造回归模型。即把各主成分作为新自变量代替原来自变量 x 做回归分析。

5. 用主成分分析筛选回归变量。用主成分分析筛选变量，从原始变量所构成的子集合中选择最佳变量，使模型本身易于做结构分析、控制和预报。

例：练习项目 attitude，主成分分析输入界面如下：

输出结果：

主成分分析

```
Call:
princomp(x = tmp.xx, scores = TRUE)

Standard deviations:
   Comp.1    Comp.2    Comp.3    Comp.4    Comp.5    Comp.6    Comp.7
22.415761 11.387252  9.686490  9.077332  6.296622  4.988263  4.589976

 7  variables and  30 observations.
Importance of components:
                          Comp.1     Comp.2    Comp.3     Comp.4     Comp.5
Standard deviation     22.415761 11.3872521 9.6864899 9.07733239 6.29662167
Proportion of Variance  0.562068  0.1450507 0.1049578 0.09217187 0.04435036
Cumulative Proportion   0.562068  0.7071187 0.8120765 0.90424841 0.94859877
                           Comp.6     Comp.7
Standard deviation     4.98826289 4.58997585
Proportion of Variance 0.02783432 0.02356691
Cumulative Proportion  0.97643309 1.00000000
```

Loadings:

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7
RATING	-0.447	0.422	-0.240	0.126	0.201	0.472	0.534
COMPLAINTS	-0.521	0.372	-0.143	-0.108	-0.372		-0.647
PRIVILEGES	-0.376		0.651	-0.626			0.173
LEARNING	-0.421	-0.146	0.186	0.485	0.621	-0.302	-0.235
RAISES	-0.376	-0.233	-0.224	0.104	-0.447	-0.593	0.437
CRITICAL	-0.130	-0.398	-0.633	-0.517	0.378		-0.115
ADVANCE	-0.229	-0.666	0.110	0.258	-0.295	0.577	

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7
SS loadings	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Proportion Var	0.143	0.143	0.143	0.143	0.143	0.143	0.143
Cumulative Var	0.143	0.286	0.429	0.571	0.714	0.857	1.000