第十七讲 对应分析(续)

2024.5.8

对应分析用2维欧氏向量表示属性变量：

对应分析求解列联表行属性和列属性的二维欧氏坐标表示，使得联系紧密的属性在欧氏空间中接近。

对应分析（续）

原文：Developed by the French, correspondence analysis is a graphical procedure for representing associations in a table of frequencies or counts.

中译版P557：对应分析是一种将频数或计数表中的各种联系用图来表示的方法，由弗伦奇所提出

French应译作“法国人”而不是“弗伦奇”！

对应分析

对列联表应用SVD和双标图可视化方法，研究行标和列标之间的对应或关联性的方法称为对应分析（correspondence analysis，课本12.7），由法国学者Jean-Paul Benzécri（1973）提出。对应分析是法国统计学派的代表性方法。

有一种观点认为，统计学作为一个学科是从法国人Pierre-Simon Laplace（1749-1827）和比利时-法国人Adolphe Quetelet(凯特勒1796-1896)开始的，后者提出了BMI。

上一讲对于列联表准备知识的介绍过于复杂，可忽略，
我们只需知道以下准备知识即可：

列联表的中心标准化

列联表: $X_{I\times J} = (x_{ij})$, 两个属性变量x, y的交叉分类计数表
x_{ij}： x属性为i，y属性为j的样本个数。

X也可以是任何其它非负强度、丰度矩阵：
x_{ij} = 行标i与列标j之间的联系强度

- 行和: $\mathbf{r} = X\mathbf{1}_J = (r_1, \dots, r_I)^\top$, $D_r = diag(\mathbf{r})$;
  列和: $\mathbf{c} = X^\top\mathbf{1}_I = (c_1, \dots, c_J)^\top$, $D_c = diag(\mathbf{c})$.
- 行归一化: $P = (x_{ij}/r_i) = D_r^{-1}X$;
  列归一化: $Q = (x_{ij}/c_j) = XD_c^{-1}$.

双向中心化: $X_c = X - \mathbf{r}\mathbf{c}^\top/n = (x_{ij} - r_ic_j/n)$

中心标准化： $R_{I\times J} = \left(\frac{x_{ij}-r_ic_j/n}{\sqrt{r_ic_j}}\right) = D_r^{-1/2}X_cD_c^{-1/2}$

Pearson卡方： $\chi^2 = n\|R\|^2 = \sum_{i,j}\frac{(x_{ij}-r_ic_j/n)^2}{r_ic_j/n}$

Contingency度量： $\frac{\chi^2}{n} = \|R\|^2$

对应分析求解列联表的行属性(I类)和列属性(J类)的二维欧氏向量表示：

$$(f_{i1}, f_{i2}), i = 1, \ldots., I;\ (g_{j1}, g_{j2}), j = 1, \ldots, J$$

使得联系紧密的属性(i, j)的f, g坐标在欧氏空间中接近。

具体地，矩阵 $P = \left(x_{ij}/r_i\right) = D_r^{-1}X$ 的每行都是一个概率分布，其第i行$\mathbf{p}_i = (x_{i1}, \ldots x_{iJ})/r_i$ 是行属性变量第 i 个值(类别)的欧氏向量刻画。

同样，矩阵$Q = \left(x_{ij}/c_j\right) = XD_c^{-1}$的第$j$列$\mathbf{q}_j = (x_{1j}, \ldots x_{Ij})/c_j$是列属性的第 j 个值的数值向量刻画。

对应分析将这些向量压缩为2维欧氏向量用于可视化。

对应分析的直观导出

中心标准化列联表R的奇异值分解：

$$R = D_r^{-1/2} X_c D_c^{-1/2} = \underbrace{U\,D}_{\text{行}}\overbrace{D\,V^{\top}}^{\text{列}}$$

UD, VD分别代表R的行和列

左乘$D_r^{1/2}$ ⇩ 右乘$D_c^{1/2}$

X的表示:

$$X = \underbrace{D_r^{1/2} U D}_{\text{行}} \overbrace{V^{\top} D_c^{1/2}}^{\text{列}} + \mathbf{r}\mathbf{c}^{\top}/n$$

$D_r^{1/2}UD$, $D_c^{1/2}VD$分别代表X的行和列

左乘D_r^{-1} ⇩ ⇩ 右乘D_c^{-1}

$P = D_r^{-1}X$的表示:

$$P = \underbrace{D_r^{-1/2} U D}_{\text{行}} V^{\top} D_c^{1/2} + \frac{\mathbf{1}\mathbf{c}^{\top}}{n}$$

$Q = XD_c^{-1}$的表示:

$$Q = D_r^{1/2} U \overbrace{D V^{\top} D_c^{-1/2}}^{\text{列}} + \frac{\mathbf{1}\mathbf{c}^{\top}}{n}$$

⇩ ⇩

$F = D_r^{-1/2}UD$表示P的行

$G = D_c^{-1/2}VD$表示Q的列

归一化矩阵的表示

假设奇异值分解，$R = D_r^{-1/2} X_c D_c^{-1/2} = UDV^\top, rank(R) = k,$

$U = (\mathbf{u}_1,\dots,\mathbf{u}_k), V = (\mathbf{v}_1,\dots,\mathbf{v}_k), D = diag(\sqrt{\lambda_1},\dots,\sqrt{\lambda_k}).$

- $P = D_r^{-1}X$ 行的表示(行属性的主坐标，Principal coordinate of row):

$$F_{I\times k} = D_r^{-1/2}UD = D_r^{-1/2}(\mathbf{u}_1\sqrt{\lambda_1},\dots,\mathbf{u}_k\sqrt{\lambda_k})$$

- $Q = X\ D_c^{-1}$列的表示(列属性的主坐标Principal coordinate of col):

$$G_{J\times k} = D_c^{-1/2}VD = D_c^{-1/2}(\mathbf{v}_1\sqrt{\lambda_1},\dots,\mathbf{v}_k\sqrt{\lambda_k})$$

对应分析/双标图

对于$I\times J$列联表X，对应分析的双标图将$F=(F_{ij})$的前两列和$G=(G_{ij})$的前两列画在同一个图上($biplot$):

$$(F_{i1},F_{i2}) = (u_{i1}\sqrt{\lambda_1}/\sqrt{r_i}, u_{i2}\sqrt{\lambda_2}/\sqrt{r_i}), i=1,\dots,I \text{表示行属性},$$

$$(G_{j1},G_{j2}) = (v_{j1}\sqrt{\lambda_1}/\sqrt{c_j}, v_{j2}\sqrt{\lambda_2}/\sqrt{c_j}), j=1,\dots,J \text{ 表示列属性}$$

上述$biplot$具备如下性质：

(1) 互相靠近的F坐标点说明对应的行属性相似，

(2) 互相靠近的G坐标点说明对应的列属性相似，

(3) 互相靠近的F,G坐标点说明对应的行、列属性关联性/对应密切。

P12命题1将证明性质(3)。尚缺少(1),(2)的证明或近似说明。

SVD是正交对角化

为了证明上述对应分析方法的合理性，我们从正交对角化的角度考察奇异值分解：

SVD：假设A为任何$q\times p$矩阵，秩为k，假设奇异值分解

$$A = UDV^{\top},$$

其中 $D_{k\times k} = \mathrm{diag}(\sqrt{\lambda_1},\ldots,\sqrt{\lambda_k})$，$\lambda_1 \ge \ldots \ge \lambda_k > 0$为$A^{\top}A$的非0特征根，

$U^{\top}U = I_k, V^{\top}V = I_k, U_{q\times k} = (\mathbf{u}_1,\ldots,\mathbf{u}_k), V_{p\times k} = (\mathbf{v}_1,\ldots,\mathbf{v}_k).$

SVD $\Leftrightarrow$ 正交对角化：

$A = UDV^{\top} \Rightarrow U, V$ 将A对角化：$U^{\top}AV = D$，即

$$\mathbf{u}_i^{\top}A\mathbf{v}_i = \sqrt{\lambda_i},\ \mathbf{u}_i^{\top}A\mathbf{v}_j = 0\ \ (i \ne j)$$

若将U, V补全为正交方阵$\tilde{U}, \tilde{V}$，D补全为与A大小相同的矩阵$\tilde{D}_{q\times p} = \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix}$，则奇异值分解可改写为$A$的正交对角化

$$A_{q\times p} = \tilde{U}\tilde{D}_{q\times p}\tilde{V}^{\top} \Leftrightarrow \tilde{U}^{\top}A_{q\times p}\tilde{V}^{\top} = \tilde{D}_{q\times p}$$

下页引理1说明奇异值分解中的U，V列向量是优化问题

$\max \mathbf{u}^\top A_{q\times p}\mathbf{v}$ 的最优解。

为什么极大化$\mathbf{u}^\top A\mathbf{v}$？

假设我们求解能最好地刻画A的行特征的向量$\mathbf{v}$，刻画A的列特征的$\mathbf{u}$。

因为A的行列之间存在对偶性，$A\mathbf{v}\in C(A)$,我们希望$A\mathbf{v}$是A的列的好的特征刻画，即$\mathbf{u}\propto A\mathbf{v}$。

由Cauchy – Schwartz不等式，当$\mathbf{u}\propto A\mathbf{v}$时，$\mathbf{u}^\top A\mathbf{v}$最大。

后面会看到（P11），如果A是双向中心化的，那么该极大化问题有更好的解释。

SVD是最优的正交对角化

引理1. 假设矩阵$A_{q\times p}$的秩为k,其奇异值分解$A=UDV^{\top}$如上页所述,
$U_{q\times k}=(\mathbf{u}_1,\ldots,\mathbf{u}_k),V_{p\times k}=(\mathbf{v}_1,\ldots,\mathbf{v}_k)$,考虑双线性函数的极大化问题

$$\max_{\substack{\mathbf{u}\in R^q,\|\mathbf{u}\|=1\\ \mathbf{v}\in R^p,\|\mathbf{v}\|=1}} \mathbf{u}^{\top}A\mathbf{v} \qquad (*)$$

(1) 极大化问题(*)在$\mathbf{u}=\mathbf{u}_1,\mathbf{v}=\mathbf{v}_1$时达到最大，最大值等于最大奇异值:

$$\max_{\mathbf{u},\mathbf{v}}\mathbf{u}^{\top}A\mathbf{v}=\sqrt{\lambda_1}。$$

(2) 对任何$t\le k$,限制$\mathbf{u}\perp\mathbf{u}_1,\ldots,\mathbf{u}_{t-1}$, $\mathbf{v}\perp\mathbf{v}_1,\ldots,\mathbf{v}_{t-1}$,则(*)在$\mathbf{u}=\mathbf{u}_t,\mathbf{v}=\mathbf{v}_t$时达到最大,最大值等于第$t$大奇异值，即$\max\limits_{\mathbf{u}\perp\mathbf{u}_1,\ldots,\mathbf{u}_{t-1},\ \mathbf{v}\perp\mathbf{v}_1,\ldots,\mathbf{v}_{t-1}}\mathbf{u}^{\top}A\mathbf{v}=\sqrt{\lambda_t}$。

证明1: 由Cauchy-Schwarz不等式,

$$\mathbf{u}^{\top}A\mathbf{v}\le\sqrt{\mathbf{u}^{\top}\mathbf{u}}\sqrt{\mathbf{v}^{\top}A^{\top}A\mathbf{v}}=\sqrt{\mathbf{v}^{\top}A^{\top}A\mathbf{v}}$$

等号成立当且仅当 $\mathbf{u}\propto A\mathbf{v}$。进一步,对任何$\mathbf{v}\in R^p$ $(\mathbf{v}^{\top}\mathbf{v}=1)$

$$\sqrt{\mathbf{v}^{\top}A^{\top}A\mathbf{v}}\le\sqrt{\lambda_1(A^{\top}A)},$$

等号成立当且仅当$\mathbf{v}=\mathbf{v}_1$($A^{\top}A$的最大特征根λ_1对应的特征向量),
此时，$\mathbf{u}\propto A\mathbf{v}_1=\mathbf{u}_1\sqrt{\lambda_1}$,所以$\mathbf{u}=\mathbf{u}_1$。依次类推，其它证明类似。

证明2: $A=UDV^\top$, 令$\mathbf{a}=U^\top\mathbf{u},\mathbf{b}=V^\top\mathbf{v}$, 因为$UU^\top\le I_p$, 则

$\|\mathbf{a}\|=\mathbf{u}^\top UU^\top\mathbf{u}\le\mathbf{u}^\top\mathbf{u}=1, \|\mathbf{b}\|\le 1$, 所以

$$\mathbf{u}^\top A\mathbf{v}=\mathbf{u}^\top UDV^\top\mathbf{v}=\mathbf{a}^\top D\mathbf{b}=\sum_{i=1}^{k}a_ib_i\sqrt{\lambda_i}\le\sqrt{\lambda_1}\sum_{i=1}^{k}|a_ib_i|\le\sqrt{\lambda_1},$$

若约束$\mathbf{u}\perp\mathbf{u}_1$，$\mathbf{v}\perp\mathbf{v}_1$，则$\mathbf{u}^\top U=(0,\mathbf{u}^\top U_{-1}), V^\top\mathbf{v}=\begin{pmatrix}0\\ V_{-1}{}^\top\mathbf{v}\end{pmatrix}$,

其中$U_{-1}=(\mathbf{u}_2,\dots,\mathbf{u}_k), V_{-1}=(\mathbf{v}_2,\dots,\mathbf{v}_k)$, 则$\mathbf{u}^\top A\mathbf{v}=\sum_{i=2}^{k}a_ib_i\sqrt{\lambda_i}\le\sqrt{\lambda_2}$，等等。

由引理1可知SVD是最优的正交对角化：

定理1. 假设矩阵$A_{q\times p}$的秩为k, 其奇异值分解$A=UDV^\top$,

$D=\mathrm{diag}(\sqrt{\lambda_1},\dots,\sqrt{\lambda_k})$, $\lambda_1\ge\dots\ge\lambda_k>0$. 则对任何$t\le k$

$$\max_{\substack{P\in R^{p\times t},\ Q\in R^{q\times t}\\ P^\top P=Q^\top Q=I_t}} trP^\top AQ=trU_t^\top AV_t,$$

其中$U_t=(\mathbf{u}_1,\dots,\mathbf{u}_t), V_t=(\mathbf{v}_1,\dots,\mathbf{v}_t)$分别为$U,V$的前$t$列。

证明：记$P=(\mathbf{p}_1,\dots,\mathbf{p}_t), Q=(\mathbf{q}_1,\dots,\mathbf{q}_t), trP^\top AQ=\sum_{i=1}^{t}\mathbf{p}_i^\top A\mathbf{q}_i$.

对应分析的最优性

如前所述，对应分析的欧氏表示如下：

- 行属性的主坐标表示：$F = D_r^{-1/2}UD = D_r^{-1/2}(\mathbf{u}_1\sqrt{\lambda_1},\ldots,\mathbf{u}_k\sqrt{\lambda_k}) \overset{\Delta}{=} (\mathbf{f}_1,\ldots,\mathbf{f}_k)$
- 列属性的主坐标表示：$G = D_c^{-1/2}VD = D_c^{-1/2}(\mathbf{v}_1\sqrt{\lambda_1},\ldots,\mathbf{v}_k\sqrt{\lambda_k}) \overset{\Delta}{=} (\mathbf{g}_1,\ldots,\mathbf{g}_k)$

我们下面试图说明为什么用上述方式表示属性变量：

假设$X_{I\times J}$是列联表，其中心化: $X_c = X - \frac{\mathbf{r}\mathbf{c}^\top}{n} = \left(x_{ij} - \frac{r_i c_j}{n}\right)$

X的中心标准化: $R = D_r^{-1/2}X_c D_c^{-1/2}$

假设R有奇异值分解：$R = D_r^{-1/2}X_c D_c^{-1/2} = UDV^\top$

由引理1，U,V的第一列$\mathbf{u}_1$，$\mathbf{v}_1$最大化

$$\mathbf{u}^\top R\mathbf{v} = \mathbf{u}^\top D_r^{-1/2}X_c D_c^{-1/2}\mathbf{v}$$

其中$\|\mathbf{u}\| = \|\mathbf{v}\| = 1$，记$\mathbf{f} = D_r^{-1/2}\mathbf{u}, \mathbf{g} = D_c^{-1/2}\mathbf{v}$,

则第一主坐标 $\mathbf{f}_1, \mathbf{g}_1$(principal coordinate F, G的第一列）最大化

$$\mathbf{f}^\top X_c\mathbf{g} = \sum_{i,j}(x_{ij} - r_i c_j/n)\, f_i g_j = -\frac{1}{2}\sum_{i,j}\left(x_{ij} - \frac{r_i c_j}{n}\right)(f_i - g_j)^2$$

即第一主坐标 $\mathbf{f}_1, \mathbf{g}_1$极小化

$$\sum_{i,j}\left(x_{ij} - \frac{r_i c_j}{n}\right)(f_i - g_j)^2\text{ , s.t. } ||D_r^{1/2}\mathbf{f}|| = ||D_c^{1/2}\mathbf{g}|| = 1$$

因为 X_c 是双向中心化的，这里对双线性函数 $\mathbf{f}^\top X_c \mathbf{g}$ 有更好的解释

基于上述讨论，我们有结论：

命题1：对应分析的第一主坐标 $\mathbf{f}_1, \mathbf{g}_1$ 是下述极小化问题的最优解：

$$\min_{\mathbf{f},\mathbf{g}} \mathbf{f}^\top X_c \mathbf{g} = \min_{\mathbf{f},\mathbf{g}} \sum_{i,j} \left(x_{ij} - \frac{r_i c_j}{n}\right) (f_i - g_j)^2 \ ,$$

$$\text{s.t. } ||D_r^{1/2}\mathbf{f}|| = ||D_c^{1/2}\mathbf{g}|| = 1$$

类似地，第二主坐标$\mathbf{f}_2, \mathbf{g}_2$是上述优化问题在进一步约束 $\mathbf{f}^\top D_r \mathbf{f}_1 = 0$, $\mathbf{g}^\top D_c \mathbf{g}_1 = 0$下的最优解。

上述优化问题理解为：求解行属性的表示$(f_{i1}, f_{i2}), i = 1, \dots, I$和列属性的表示$(g_{j1}, g_{j2}), j = 1, \dots, J$, 使得

当$x_{ij} > \frac{r_i c_j}{n}$时，$|f_{i\alpha} - g_{j\alpha}|$尽量接近于0，$\alpha = 1,2$

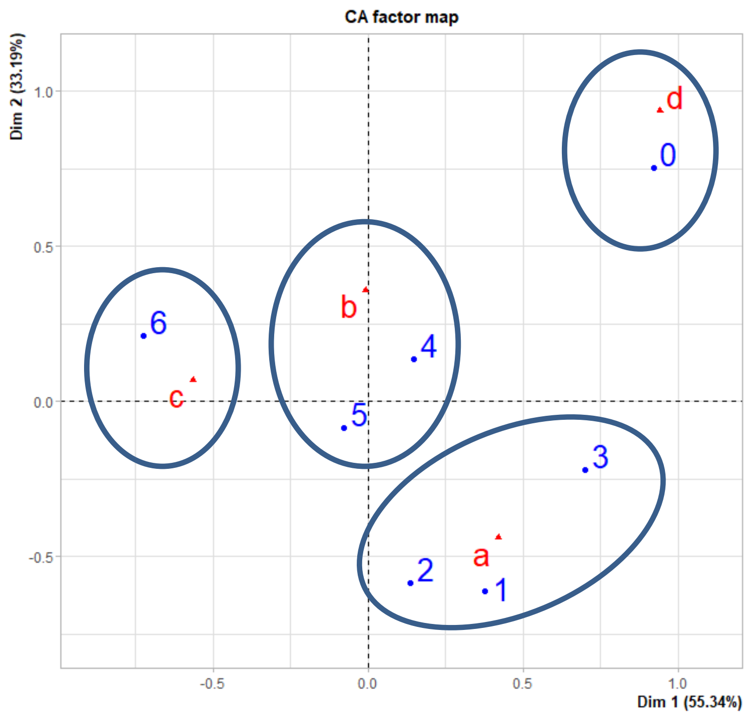
当$x_{ij} < \frac{r_i c_j}{n}$时， $|f_{i\alpha} - g_{j\alpha}|$尽量大,$\alpha = 1,2$

这正符合对应分析的目标 .

在某种意义上，命题1证明了P6所述对应分析的性质（3）：

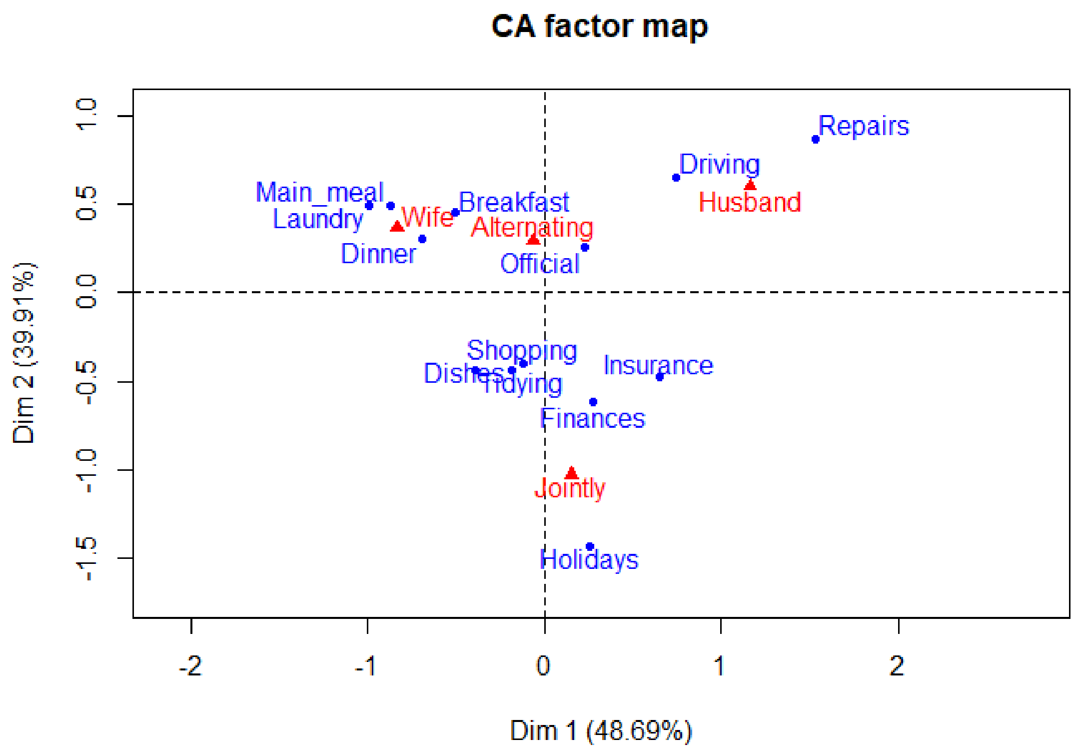
(3) 互相靠近的F, G坐标点说明对应的行、列属性关联性/对应密切。

例2（续）对应分析。双标图显示四种陶器差别较大（b，c比较接近），地点1，2，3很相似，它们出土较多的a类型陶器。（d，0）联系密切，与其他遗址和陶器差别较大。




```
library(FactoMineR)
CA(contingency_table)
```

例3（续） 13项家务活，共有4种分配方案：wife， alternating， husband， jointly.



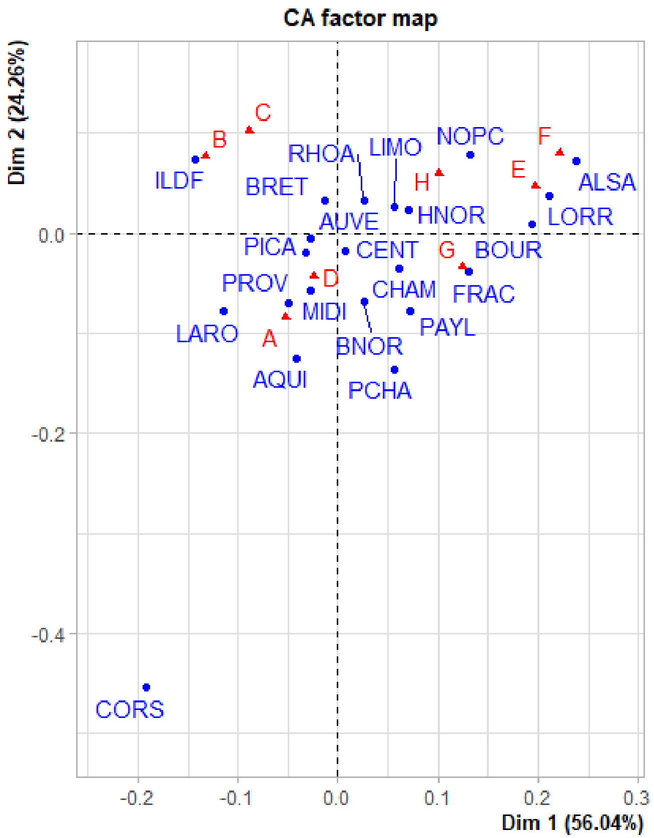
双标图表明： 妻子主要负责做饭洗衣，丈夫负责驾驶和修理房子，一起做的家务主要是购物、财务、保险和节日活动。

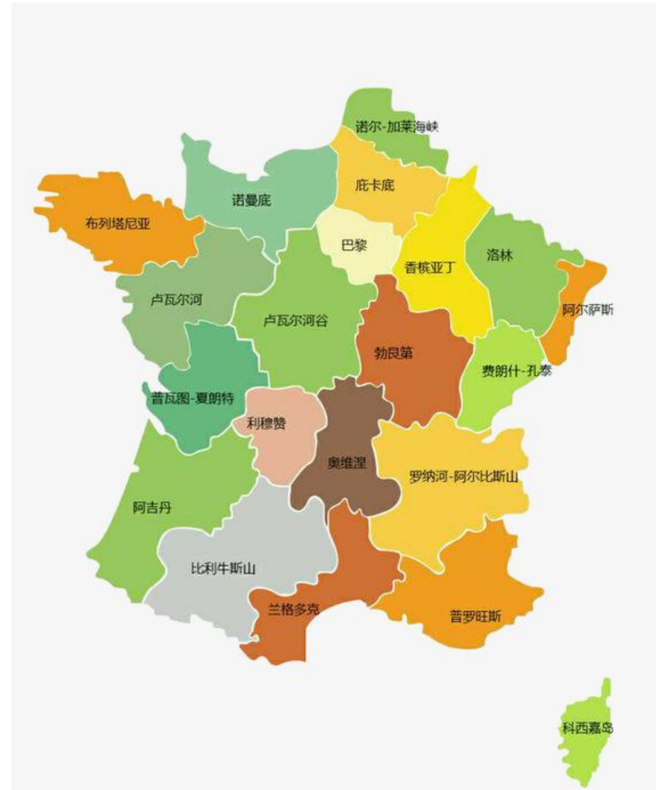
例4. 1976年法国本土22个大区的202100位大学毕业生的专业统计如下，8个专业(A-H)如下：

	A	B	C	D	E	F	G	H
ILDF	9724	5650	8679	9432	839	3353	5355	83
CHAM	924	464	567	984	132	423	736	12
PICA	1081	490	830	1222	118	410	743	13
HNOR	1135	587	686	904	83	629	813	13
CENT	1482	667	1020	1535	173	629	989	26
BNOR	1033	509	553	1063	100	433	742	13
BOUR	1272	527	861	1116	219	769	1232	13
NOPC	2549	1141	2164	2752	587	1660	1951	41
LORR	1828	681	1364	1741	302	1289	1683	15
ALSA	1076	443	880	1121	145	917	1091	15
FRAC	827	333	481	892	137	451	618	18
PAYL	2213	809	1439	2623	269	990	1783	14
BRET	2158	1271	1633	2352	350	950	1509	22
PCHA	1358	503	639	1377	164	495	959	10
AQUI	2757	873	1466	2296	215	789	1459	17
MIDI	2493	1120	1494	2329	254	855	1565	28
LIMO	551	297	386	663	67	334	378	12
RHOA	3951	2127	3218	4743	545	2072	3018	36
AUVE	1066	579	724	1239	126	476	649	12
LARO	1844	816	1154	1839	156	469	993	16
PROV	3944	1645	2415	3616	343	1236	2404	22
CORS	327	31	85	178	9	27	79	0

A: Philosophy-Letters,
B: Economics and Social Sciences,
C: Mathematics and Physics,
D: Mathematics and Natural Sciences,
E: Mathematics and Techniques,
F: Industrial Techniques,
G: Economic Techniques,
H: Computer Techniques.

NOPC: Nord Pas-de-Calais：北部-加莱海峡
PICA: Picardie：皮卡第
HNOR: Normandie：诺曼底
ILDF: ILE de France：法兰西岛(大巴黎)
CHAM: Champagne Ardenne：香槟-阿登
LORR: Lorraine：洛林
ALSA: Alsace：阿尔萨斯
BRET: Bretagne：布列塔尼
PAYL: Pays de loire:卢瓦尔河地区
CENT: Centre Val de loire:中央 - 卢瓦河山谷
BOUR: Bourgogne:勃艮第
FRAC: Franche comté:弗朗什孔泰
PCHA: Poitou Charentes：普瓦图-夏朗德
LIMO: Limousin：利穆赞
AUVE: Auvergne：奥弗涅
RHOA: Rhone Alpes：罗讷-阿尔卑斯
Alpes Pays de savoie：阿尔卑斯 - 萨瓦地区
AQUI: Aquitaine：阿基坦
MIDI: Midi pyrénées：南部-比利牛斯
PROV: Provence Alpes Cote d'azur：普罗旺斯 - 蓝色海岸
LARO: Languedoc Roussilon：朗格多克-鲁西永
CORS: corse：科西嘉





科西嘉(CORS)远离其它地区。大巴黎区也较为特殊，与B,C关系密切，有较多的经济学、数理专业学生。

阿尔萨斯、洛林地区工业技术（E，F）学生较多。普罗旺斯地区较多文学艺术(A)专业学生。

香槟亚丁？农业区

普氏分析

古希腊神话人物Procrustes普罗克路斯忒斯是一个强盗，客栈老板，他把每个入住的客人拉伸或将腿截断，以使其与他的铁床对齐。

有一种说法：他实际上有两个大小不同的铁床。。

Procrustean bed：不合理但需要严格遵守的规则或标准.
Procrustean solution：先定目标模型和方法，再寻找数据或删减数据以符合模型的欺骗做法。

普氏分析（Procustes analysis）将两个数据集对齐，假设有两个$n\times p$数据集X,Y,普氏分析寻找旋转、平移变换使得变换之后两者尽量相似。

目标函数

设$X_{n\times p}=(\mathbf{x}_1,\ldots,\mathbf{x}_n)^{\mathsf{T}},Y_{n\times p}=(\mathbf{y}_1,\ldots,\mathbf{y}_n)^{\mathsf{T}}$,普氏分析求解$p\times p$正交矩阵$Q$和$\mathbf{b}\in R^p$使得

$$\sum_{i=1}^{n}\|\mathbf{x}_i-(Q\mathbf{y}_i+\mathbf{b})\|^2=\|X-(YQ^{\mathsf{T}}+\mathbf{1}\mathbf{b}^{\mathsf{T}})\|_F^2=\min!$$

对于给定的正交矩阵Q，$\mathbf{b}$的最优解为$\hat{\mathbf{b}} = \bar{\mathbf{x}} - Q\bar{\mathbf{y}}$,此时目标函数

$\| X - (YQ^\top + \mathbf{1}\mathbf{b}^\top) \|_F^2 = \| X - \mathbf{1}\bar{\mathbf{x}}^\top - (Y - \mathbf{1}\bar{\mathbf{y}}^\top)Q^\top \|_F^2$ 所以我们不妨

假设X, Y都是中心化的。因为

$$\| X - YQ^\top \|_F^2 = tr(X^\top X) + tr(Y^\top Y) - 2tr(X^\top YQ^\top)$$

所以原极小化问题等价于极大化问题：

$$\max_{Q^\top Q = QQ^\top = I_p} tr(X^\top YQ^\top)$$

其中$X^\top Y/(n-1)$为样本协方差矩阵。

> 命题1：若$A_{p\times p} = X^\top Y$的SVD分解为$A = UDV^\top$,其中U, V是$p\times p$正交矩阵，
> 则上述优化问题最优的正交变换为A的“方向”矩阵$Q = UV^\top$,
> 即$YQ = YVU^\top$与X最接近。

证明：记矩阵$A = X^\top Y$的SVD分解为$A = U_{p\times p} D_{p\times p} V^\top{}_{p\times p} = \sum_{i=1}^{p} \sqrt{\lambda_i}\mathbf{u}_i \mathbf{v}_i^\top$,

$$\Rightarrow tr(X^\top YQ^\top) = tr\left(\sum_{i=1}^{p} \sqrt{\lambda_i}\mathbf{u}_i \mathbf{v}_i^\top Q^\top\right) = \sum_{i=1}^{p} \sqrt{\lambda_i}\, tr\left(\mathbf{u}_i \mathbf{v}_i^\top Q^\top\right) = \sum_{i=1}^{p} \sqrt{\lambda_i}\mathbf{v}_i^\top Q^\top \mathbf{u}_i,$$

由Cauchy不等式，$\mathbf{v}_i{}^\top Q^\top \mathbf{u}_i \le \| Q\mathbf{v}_i \| \cdot \| \mathbf{u}_i \| = 1$, 当$\mathbf{u}_i = Q\mathbf{v}_i$时达到最大，

此时$U = QV \Rightarrow \quad Q = UV^\top$,

$$\max tr(X^\top YQ^\top) = tr(X^\top YVU^\top) = tr(U^\top X^\top YV) = tr(D) = \sqrt{\lambda_1} + \ldots + \sqrt{\lambda_p}$$

任何复数有极分解（polarvdecomposition）：$z = re^{i\theta}$。

任何向量$\mathbf{x} \in R^p$有极分解$: \mathbf{x} = \boldsymbol{\theta} r$, 其中$\boldsymbol{\theta} = \mathbf{x} / \| \mathbf{x} \|, r = \| \mathbf{x} \|$。

任何矩阵也有极分解$: A = \Theta R_1 = R_2 \Theta$

命题2. 假设矩阵A的SVD为$A = UDV^\top$, 则A有极分解表示

$$A = \Theta R_1 = R_2 \Theta,$$

其中$R_1 = (A^\top A)^{1/2}$和$R_2 = (AA^\top)^{1/2}$代表A的模长,

$\Theta = UV^\top$代表A的"方向",

证: 因为$A^\top A = VDU^\top UDV^\top = VD^2 V^\top$，所以

$$A = UDV^\top = UV^\top VDV^\top = UV^\top \left(VD^2 V^\top \right)^{1/2} = UV^\top (A^\top A)^{1/2} \triangleq \Theta (A^\top A)^{1/2}$$

其中$\Theta = UV^\top$. 同理也有 $A = (AA^\top)^{1/2} \Theta$