对应分析

TUTU

对应分析基本概念

♣ 基本思想：

将一个列联表的行和列中各元素的比例结构以点的形式在低维空间表示出来。它最大特点是能把众多的样品和众多的变量同时作到一张图上，直观展示。

♣ 列联表：

- $p_{ij} = \dfrac{n_{ij}}{n}$
- $E(n_{ij}) = np_{ij} = np_{i\cdot}p_{\cdot j}$
- H_0：属性变量 A 与 B 相互独立；H_1：属性变量 A 与 B 不独立
- 统计量：$\chi^2 = n\sum_{i=1}^{n}\sum_{j=1}^{p}\dfrac{(p_{ij}-p_{i\cdot}p_{\cdot j})^2}{p_{i\cdot}p_{\cdot j}} \sim \chi^2[(n-1)(p-1)]$

对应分析模型

♣ 原矩阵 $X=(x_{ij})_{n\times p}$，均值化矩阵

$X^*=(\boldsymbol{x}_1-\bar{x}_1,\boldsymbol{x}_2-\bar{x}_2,\cdots,\boldsymbol{x}_p-\bar{x}_p)$

♣ 变量的积叉矩阵 $(p\times p)$：$\Sigma_R=(X^*)^{\mathrm{T}}X^*$

♣ 样品的积叉矩阵 $(n\times n)$：$\Sigma_Q=X^*(X^*)^{\mathrm{T}}$

♣ 行形象：$\left[\dfrac{p_{i1}}{p_{i\cdot}},\dfrac{p_{i2}}{p_{i\cdot}},\cdots,\dfrac{p_{ip}}{p_{i\cdot}}\right]=\left[\dfrac{x_{i1}}{x_{i\cdot}},\dfrac{x_{i2}}{x_{i\cdot}},\cdots,\dfrac{x_{ip}}{x_{i\cdot}}\right]$，$E\left(\dfrac{p_{ij}}{p_{i\cdot}}\right)=p_{\cdot j}$，$E\left(\dfrac{p_{ij}}{p_{i\cdot}\sqrt{p_{\cdot j}}}\right)=\sqrt{p_{\cdot j}}$

♣ $N(R)=\left(\dfrac{p_{ij}}{p_{i\cdot}}\right)_{n\times p}$，$D(R)=\left(\dfrac{p_{ij}}{p_{i\cdot}\sqrt{p_{\cdot j}}}\right)_{n\times p}$

♣ 令 $z_{ak}=\dfrac{p_{ak}-p_{a\cdot}p_{\cdot k}}{\sqrt{p_{a\cdot}p_{\cdot k}}}=\dfrac{x_{ak}-x_{a\cdot}x_{\cdot k}}{\sqrt{x_{a\cdot}x_{\cdot k}}}$，$Z=(z_{ak})_{n\times p}$，$A_{p\times p}=Z^{\mathrm{T}}Z$

对应分析模型

♣ 列形象：$\left[\frac{p_{1j}}{p_{\cdot j}}, \frac{p_{2j}}{p_{\cdot j}}, \cdots, \frac{p_{nj}}{p_{\cdot j}}\right] = \left[\frac{x_{1j}}{x_{\cdot j}}, \frac{x_{2j}}{x_{\cdot j}}, \cdots, \frac{x_{nj}}{x_{\cdot j}}\right]$，$E\left(\frac{p_{ij}}{p_{\cdot j}}\right) = p_{i\cdot}$, $E\left(\frac{p_{ij}}{p_{\cdot j}\sqrt{p_{i\cdot}}}\right) = \sqrt{p_{i\cdot}}$

♣ $N(Q) = \left(\frac{p_{ij}}{p_{\cdot j}}\right)_{n\times p}$，$D(R) = \left(\frac{p_{ij}}{p_{\cdot j}\sqrt{p_{i\cdot}}}\right)_{n\times p}$

♣ 令 $z_{ak} = \frac{p_{ak} - p_{a\cdot}p_{\cdot k}}{\sqrt{p_{a\cdot}p_{\cdot k}}} = \frac{x_{ak} - x_{a\cdot}x_{\cdot k}}{\sqrt{x_{a\cdot}x_{\cdot k}}}$，$Z = (z_{ak})_{n\times p}$，$B_{n\times n} = ZZ^{\mathrm{T}}$

♣ λ_k 是 $A = Z^{\mathrm{T}}Z$ 的特征根，对应的特征向量为 u_k，则 λ_k 也是 $B = ZZ^{\mathrm{T}}$ 的特征根，对应的特征向量为 Zu_k

对应分析模型

♣ 基本步骤：

1. 获取对应分析数据
2. 建立列联表
3. 对应分析
4. 对应图并解释结果的意义
    - 从中心向任意点连线 (向量)，例如从中心向“扩展能力强”作向量，然后让所有“软件”往这条向量及延长线上作垂线，垂点越靠近向量正向的表示扩展能力越强
    - 从距离中的位置看：越靠近中心，越没有特征，越远离中心，说明特征越明显。各个类别之间的距离表示相对密切关系

对应分析模型

♣ 优缺点：

- 优点：
  - 定性变量划分的类别越多，这种方法的优越性越明显
  - 揭示行变量类间与列变量类间的联系
  - 直观地表变量所含类别间的关系
- 缺点：
  - 不能用于相关关系的假设检验
  - 受极端值的影响

对应分析 SAS 代码

♣ SAS 代码:

```
/*用于列联表形式的对应分析*/
proc corresp data=yourdata out=a2 rp cp short;
var x1-x4;
id region;
run;

/*用于尚未变成列联表的原始数据的对应分析*/
proc corresp data=yourdata out=a4 rp cp short;
tables row, column;
run;
```