

CA/MCAの数理(3)

resultの解釈で一番重要な寄与率/Contribution

寄与率/Contribution

- 分析対象の系（システム）は、ある分散値をもっている。
 - 全体の平均位置の座標はゼロ。
 - そのゼロからの偏差に注目。（残差行列）
 - その重み付き二乗和 -> 分散
 - この「重み」weightを「質量」massと呼ぶ。
 - イメージとしては原点Oから、偏差（n次元）を持ったポイントが、質量mで散らばっている様子。

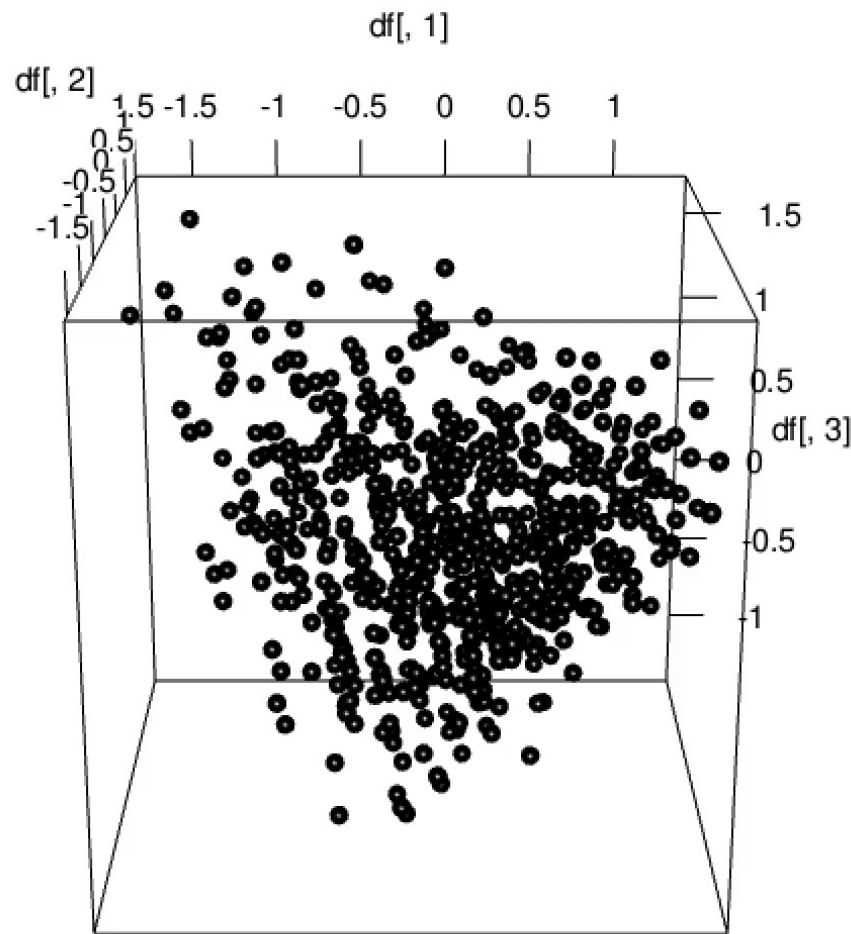
注意点

- 注意点
 - 質量は、元表から計算される周辺度数（比率）
 - 座標は、SVDで計算される値。
- 全体の分散値は、SVDをやっても保存されている。
- CA/MCAは、分散の分解技法と呼ばれる
- この分散を、ベンゼクリは、「慣性」inertiaと呼ぶ。
 - 物理学での慣性モーメントとの同型の式であるから。
 - もう一つh、系を記述するにあたって、分析者が選択した「慣性系」という含みもある。（ニュートンの第一法則）

動かすところな感じ。

```
` `{r}  
install.packages("rgl")  
library(rgl)  
  
plot3d(df[,1], df[,2], df[,3],  
       pch = 16, size = 1, type = "s")  
...
```

MCAの結果の行（個体）座標1~3を
選んで、plot3dで描画。
見る向きをマウスで変えられます



主軸技法

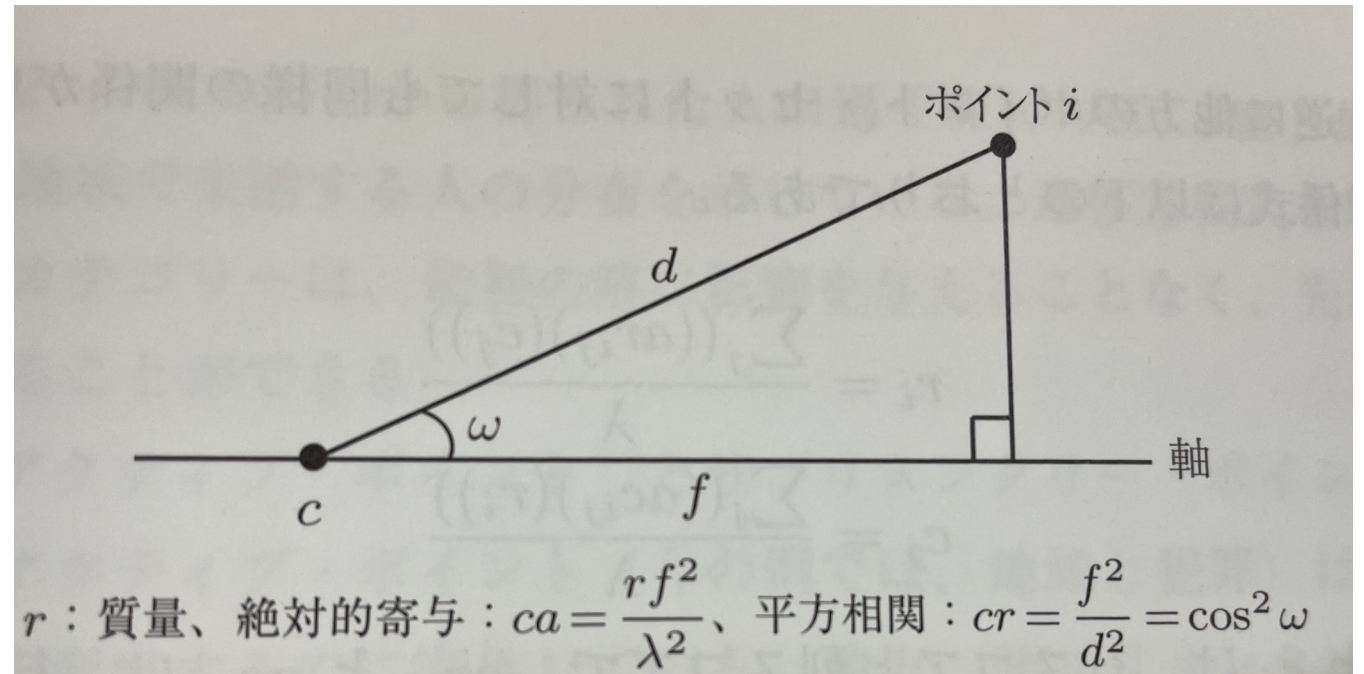
- 元の系がもっている分散値=情報は、SVDでも保存されている。
- それが、「主軸」 principal axis
- 主軸技法とは
 - 最初に系の分散をもっとも大きく体現できる軸を選ぶ（第1軸 dim1）。
 - 次に、それと直行し次にもっとも大きく体現できる軸を選ぶ（第2軸 dim2）
 - 以下、繰り返し。
 - どこまでやれるかは、 $m \times n$ 行列の場合、小さい方-1次元の軸を計算できる。
- 体現する分散（慣性）の大きさの順に並べて「累積寄与率」を計算する。
 - これを確認すると何軸を評価すれば、分散/慣性の何%が分析の対象になるか、もしくは、何%を誤差として捨てているか、ということをはっきりできる。
 - -> 分析を何軸までにするかを判断。
 - 表と棒グラフ、累積寄与率は折れ線グラフで、表現するのが定番。

SVDで取得されるのは座標値、 質量（周辺度数比率）は元表

- 計算式

- r を質量、 f を軸上の座標値、 d を原点 C からポイント i までの距離、軸と i の方向角を ω
- $\text{CONTR} = r f^2 / \lambda^2$
(ca と書くのは、absolute contribution)
- $\text{COS2} = f^2 / d^2 = \cos^2 \omega$ (cr と書くのは、relative contribution)

- 『対応分析入門』 p21



用語さまざま

- 絶対的寄与率/相対的寄与率
 - 絶対的寄与率は、あるポイントがある軸に寄与している値の軸全体の慣性への割合。CONTR
 - つまり、そのポイントのその軸への影響度合い。
 - 座標軸の意味を解釈するヒント。
 - 相対的寄与率は、ある軸にそのポイントの何%が表現されているか、と表している。平方相関。cos2。
 - 座標軸によって、そのポイントがどの程度説明されているかを示している。
- CONTRがたかければ、高いcos2を示すが、逆ではにあい (Greenacer1984, Clausen1998=2015:21)

分散の分解技法としてのCA/MCA

- 分散の分解という視点で、resultを解釈していく。
- その際に前提にしているのが、分解する際のグルーピングによる部分集合。
- 平方和の分解の式（回帰分析や分散分析でも登場するやつです）
 - 全分散 $V_{total} = \text{群内分散 } V_{within} + \text{群間分散 } V_{between}$
 - 「群」は、グループ化されたポイント。
 - それは、群の重心（平均位置）を持つ。
 - その群の平均位置と重心Oの間の分散が、「群間」分散。
 - 大隅他訳の『多重対応分析』p30～35
 - 相関比2乗 $\eta^2 = V_{between} / (V_{within} + V_{between})$
 - 注目している群/グループ（大抵はある変数）の分散の中で、その群の平均位置に表現された群重心の分散が占める割合の2乗。

計算するのか…

- ここにあげた数値は、GDAtoolsやFactoMineRのパッケージが自動で計算してくれます。
 - 一度は、自分で（Excelでも使って）手計算しておくといいかもしれません。
 - 毎回やる必要はありません。
- 以上を念頭に実際の例をみていきます。

CA/MCAの数理のまとめ

- 入力は 2 元表
 - 期待値からの残差行列をつくる
 - $m \times n$ 次元の入力行列を主軸で分解 (SVD)
 - これで、行座標、列座標、軸の分散 (特異値²、固有値)
- 各軸ごとの分散への寄与率を計算する