

コホート分析には何年くらいのデータ期間が必要か？

中村 隆 データ科学研究系 教授

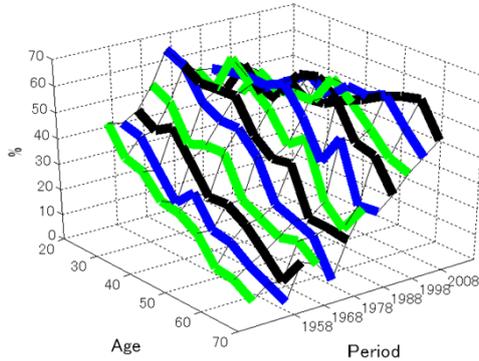
◆ 1. 日本人の国民性調査

- 統計数理研究所では、昭和28(1953)年から5年ごとに、継続的社会調査「日本人の国民性調査」を実施している。
- 平成20(2008)年秋には、第12次全国調査を実施した。(次回は2013年秋を予定。)

日本人の国民性調査のURL
<http://www.ism.ac.jp/kokuminsei/index.html>

◆ 2. コホート分析

- 社会の変動を、意見や意識の変化を通して捉えようとするとき、主にどのような要因によっているのかを見極めることが大切である。
 - 加齢の要因**・・・時代や世代によらず、人が歳をとることによって変わる。
 - 時勢の要因**・・・時代につれ人々全体の意見がある方向に変わっていく。
 - 世代の要因**・・・生まれ育った時代背景が違うことによる差。
- これらの要因の働き方によって、将来の社会の様相は大きく異なる。
- コホート分析とは
 - 社会の変動を捉え、将来の動向を探るために。
 - 継続的に調査された長期にわたるデータを分析。
 - 各要因の影響の大きさ(効果)を分離する統計的方法。
 - 年齢・時代・世代(=コホート、同時出生集団)効果の分離。



#1.3 学歴(高校卒、女性)の割合の推移(学歴は世代効果しかないはず)

参考文献
 中村 隆 (2005). コホート分析における交互作用効果モデル再考, 統計数理 53(1), 103-132.
ベイズ型コホートモデルのURL
<http://www.ism.ac.jp/~nakamura/>

◆ 4. 克服法 ~ ベイズ型コホートモデル

- コホートモデル:

$$\eta_{ij} = \log[\pi_{ij} / (1 - \pi_{ij})] = \beta^G + \beta_j^A + \beta_j^P + \beta_k^C + \beta_j^{AP} \quad (i \text{ for age, } j \text{ for period, and } k \text{ for cohort})$$
- 困難点を克服するために、年齢、時代および世代効果が緩やかに変化するという「パラメータの漸進的変化の条件」という付加条件を取り込む。
- 赤池のベイズ型情報量規準 (ABIC) を最小にするようなモデルを選ぶ「ベイズ型コホートモデル」を開発。

◆ 3. 困難点 ~ コホート分析における識別問題

- 年齢・時代・世代の効果を分けて考えるというコホート分析の考え方は魅力的。
- しかし、いざ実際に分析を行おうとすると、年齢・時代・世代の3つの効果が分離できないという困難がある。
 - たとえば、若い時の意見を、異なる世代について比較しようとするときに、時代の違いを持ち込まないようにするのはとても難しい。
- それがコホート分析における「識別問題」と呼ばれる困難点である。

◆ 5. 何年くらいのデータ期間が必要か？ 10年, 20年, 30年？

- データ期間が短いと、同一世代が通過する年齢範囲も短く、またある年齢範囲を通過する世代数も少なく、年齢によるか世代によるかが安定して得られない可能性がある。そこで、
- 日本人の国民性調査データについて、データ期間を20年、30年と増やした分析結果をみてゆき、現時点での全期間による分析結果とどのように異なるのかをみる。

いくつかの項目の例:

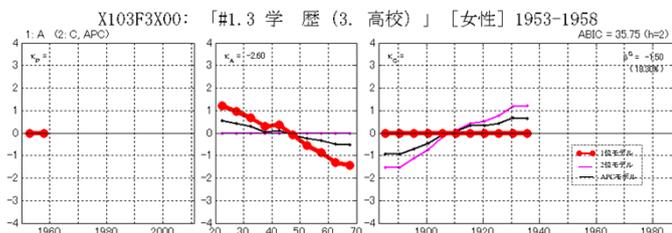
- 「#1.3 学歴(高校卒)」[女性]
 - ✓ Cモデル(15年で安定)
 - 「#6.2 男・女の生まれかわり(男)」[男性]
 - ✓ Gモデル(20年) → 現時点では[AP]モデル
 - 「#6.2 男・女の生まれかわり(女)」[女性]
 - ✓ APCモデル(25年) → [AP]APCモデル(35年) → [AP]PCモデル
 - 「#2.7 一番大切なもの(家族)」[男性]
 - ✓ APCモデル(15年) → [AP]APCモデル(35年)
 - 「#2.7 一番大切なもの(家族)」[女性]
 - ✓ APCモデル(20年) → [AP]APCモデル(30年)
- ▶ データ期間としては、20年程度は必要(日本人の国民性調査規模で)。
 ▶ データ期間が長くなれば、より複雑なモデルが選ばれる。

[AP]は、年齢×時代の交互作用効果を示す

◆ 6. では、10年, 20年, の時点でどうするか？

- モデル選択表(モデルの順番, モデル間のABICの差)をみる。
- リサンプリング・シミュレーションにより、最適モデルが入れ替わる様子、効果推定値の変動の程度をみる。

2時点(5年)では



上位2モデルとAPCモデル(効果推定値が異なる)

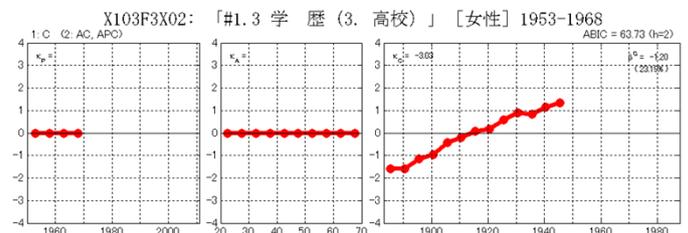
| Model | ABIC | ΔABIC | h | log ₂ σ _A ² | log ₂ σ _P ² | log ₂ σ _C ² |
|-------|----------|----------|---|--|--|--|
| A | 35.7515 | - | 2 | -2.60 | - | - |
| C | 35.8000 | 0.0485 | 2 | - | - | -2.63 |
| AC | 36.8333 | 1.0818 | 3 | -4.18 | - | -3.51 |
| PC | 37.6810 | 1.9295 | 3 | - | -7.38 | -2.63 |
| AP | 37.6939 | 1.9424 | 3 | -2.60 | -8.11 | - |
| APC | 38.8335 | 3.0820 | 4 | -4.18 | -18.77 | -3.51 |
| G | 217.8992 | 182.1477 | 1 | - | - | - |
| P | 219.8993 | 184.1478 | 2 | - | -20.58 | - |

モデル選択表(1位2位でABICの差がほとんどない)

| Model | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|-------|----|----|----|----|----|----|---|----|
| A | 40 | 14 | 9 | 2 | 25 | 10 | - | - |
| AC | 32 | 50 | 18 | - | - | - | - | - |
| C | 26 | 10 | 9 | 10 | 24 | 21 | - | - |
| AP | 2 | 8 | 31 | 18 | 12 | 29 | - | - |
| PC | - | 9 | 16 | 17 | 25 | 33 | - | - |
| APC | - | 9 | 17 | 53 | 14 | 7 | - | - |
| G | - | - | - | - | - | 96 | 4 | - |
| P | - | - | - | - | - | - | 4 | 96 |

リサンプリング・シミュレーションの結果(上位3モデルが拮抗)

4時点(15年)では



上位2モデルとAPCモデル(効果推定値は重なってほとんど同じ)

| Model | ABIC | ΔABIC | h | log ₂ σ _A ² | log ₂ σ _P ² | log ₂ σ _C ² |
|-------|----------|----------|---|--|--|--|
| C | 63.7345 | - | 2 | - | - | -3.03 |
| AC | 65.7345 | 2.0000 | 3 | -23.19 | - | -3.03 |
| PC | 65.7346 | 2.0001 | 3 | - | -22.18 | -3.03 |
| APC | 67.7345 | 4.0000 | 4 | -25.14 | -23.83 | -3.03 |
| AP | 80.1236 | 16.3891 | 3 | -3.62 | -4.11 | - |
| A | 139.7701 | 76.0356 | 2 | -3.71 | - | - |
| P | 440.6044 | 376.8699 | 2 | - | -4.59 | - |
| G | 481.0699 | 417.3354 | 1 | - | - | - |

モデル選択表(上位4モデルはC効果をもつ。1位2位でABICの差が2.0。あてはまりはほとんど変わらないことを示す)

| Model | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|-------|----|----|----|----|----|-----|-----|-----|
| C | 58 | 16 | 19 | 7 | - | - | - | - |
| AC | 23 | 40 | 26 | 8 | 3 | - | - | - |
| PC | 9 | 23 | 35 | 28 | 5 | - | - | - |
| APC | 8 | 18 | 19 | 54 | 1 | - | - | - |
| AP | 2 | 3 | 1 | 3 | 91 | - | - | - |
| A | - | - | - | - | - | 100 | - | - |
| P | - | - | - | - | - | - | 100 | - |
| G | - | - | - | - | - | - | - | 100 |

リサンプリング・シミュレーションの結果(Cモデルが6割1位を占める)