

2019年10月26日
研究集会「Demographic Change」

テンソル解析を用いた 死因別将来死亡率の同時推定

明治大学 先端数理科学研究科 博士前期課程2年 鈴木孝太郎

明治大学 総合数理学部 松山直樹

目次

- 研究動機
- Lee-Carterモデルの概説
- 先行研究
 - 死因別モデル
 - テンソル型モデル
- モデルの提案
- 実際の解析
 - 使用データ
 - 推定結果
- まとめ
- 参考文献

研究動機

- 研究動機
- Lee-Carterモデルの概説
- 先行研究
- モデルの提案
- 実際の解析
- まとめ
- 参考文献

- 将来死亡率の推定は人口推計や死亡リスク管理に有用
- 死亡率の推移を表すモデルとしてLee-Carterモデルが知られている
- Lee-Carterを含む多くのモデルが年齢・暦年のみに依存するモデル

死因を考慮することで
モデルの予測精度を上げられないか？

研究動機

- 研究動機
- Lee-Carterモデルの概説
- 先行研究
- モデルの提案
- 実際の解析
- まとめ
- 参考文献

この発想は過去に日本の人口推計で用いられていた以下の方式に繋がることがいずれもGompertzモデルのような関数フィッティングであった

- 年齢別死因別死亡率補外方式
- 標準化死因別死亡率補外方式

しかし、この方法は

- 推定パラメータの多さ
- 死因分類の改定における影響
- 死因別データのデータ数の少なさ

等の問題とLee-Carterモデルに代表されるリレーショナルモデルの発展から用いられなくなった

研究動機

- 研究動機
- Lee-Carterモデルの概説
- 先行研究
- モデルの提案
- 実際の解析
- まとめ
- 参考文献

現在死因別死亡率推定を行う研究では
全体死亡率で用いられる方法を死因別データに転用するものが多い

例(死因別Lee-Carterモデル) : $\log m_{x,t,c} = \alpha_{x,c} + \beta_{x,c}\kappa_{t,c}$

この方法は死因間の相互関係を無視しているため
将来推定には使えないという欠点がある

モデルが大規模になってしまうため
推定の安定性に欠けるという欠点もある

目的 : 死因別死亡率の同時推定モデルの提案

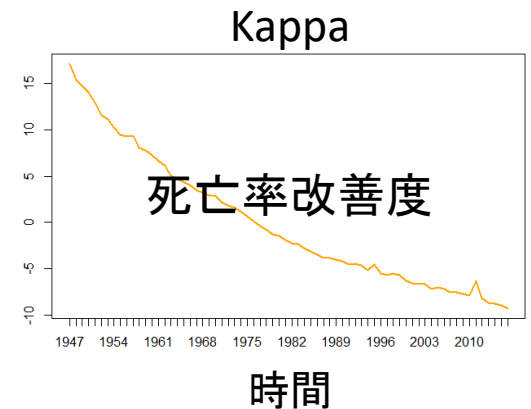
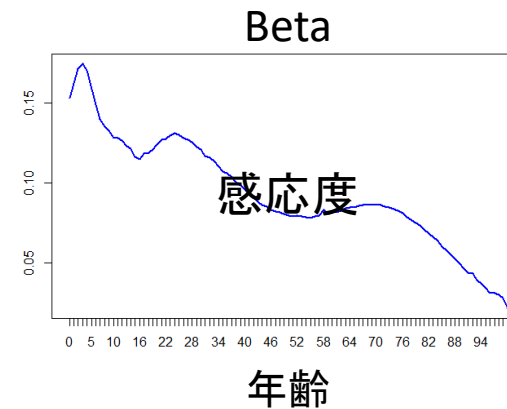
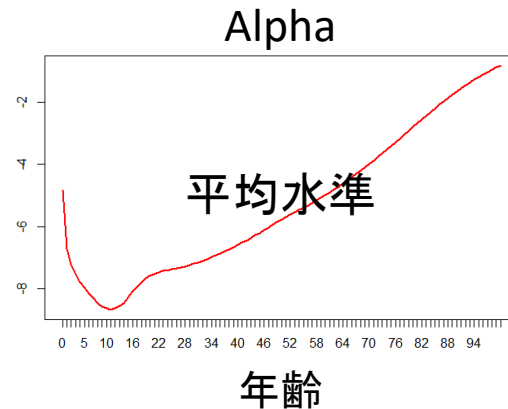
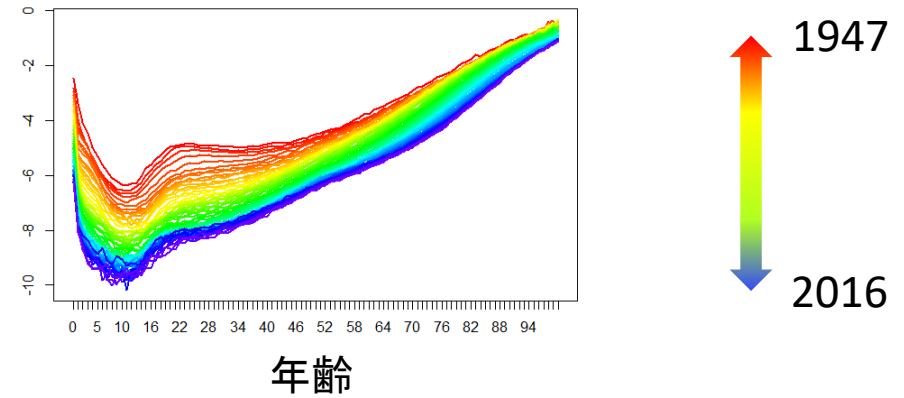
Lee-Carter モデルとは

- 研究動機
- Lee-Carterモデルの概説
- 先行研究
- モデルの提案
- 実際の解析
- まとめ
- 参考文献

Lee and Carter (1992)により考案された
対数死亡率を3要素に分解するモデル

高い解釈可能性と実装のしやすさから実務でも使われている

$$\log m_{x,t,c} = \alpha_{x,c} + \beta_{x,c} \kappa_{t,c}$$

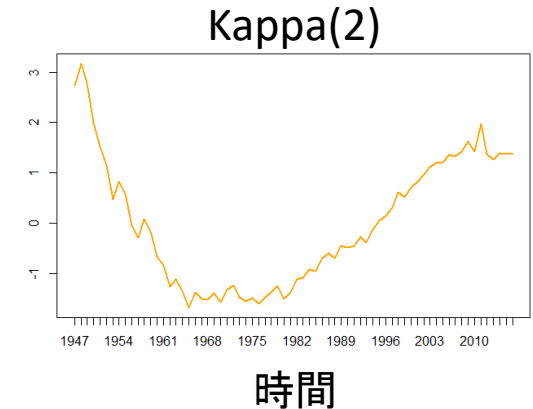
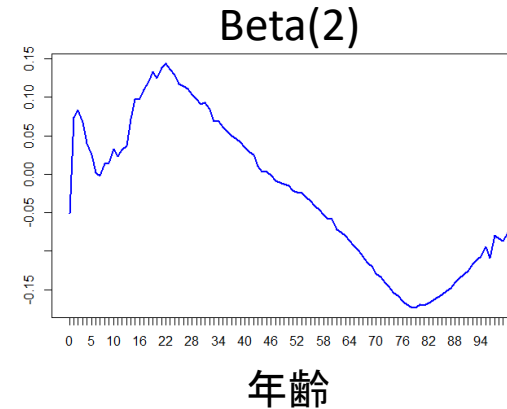
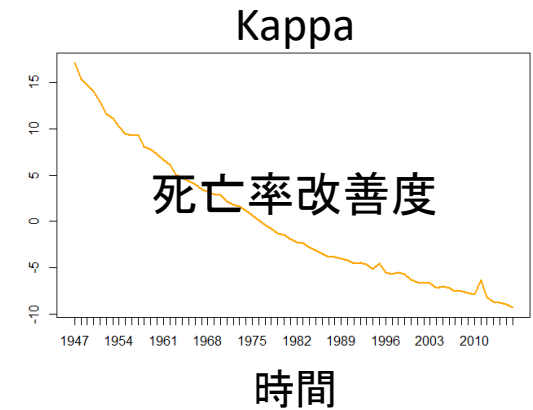
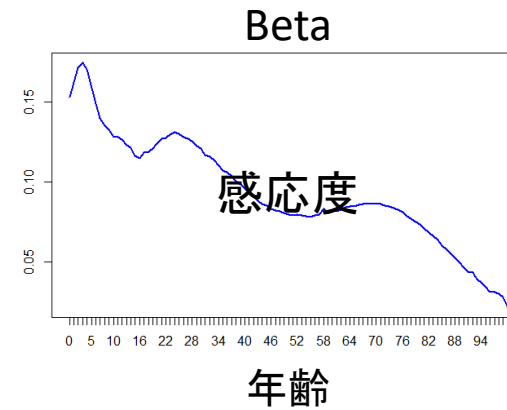
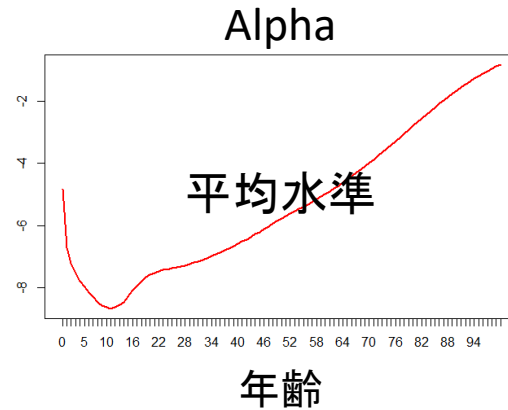


拡張Lee-Carterモデル

- 研究動機
- Lee-Carterモデルの概説
- 先行研究
- モデルの提案
- 実際の解析
- まとめ
- 参考文献

● 低ランク近似に基づく項を加えるモデル(Renshaw and Haberman(2003))

$$\bullet \log m_{x,t} = \alpha_x + \beta_x^{(1)} \kappa_t^{(1)} + \beta_x^{(2)} \kappa_t^{(2)} + \varepsilon_{x,t}$$



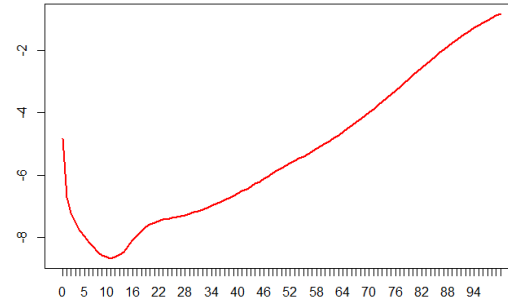
パラメータの解釈が難しい

男女間の比較

- 研究動機
- Lee-Carterモデルの概説
- 先行研究
- モデルの提案
- 実際の解析
- まとめ
- 参考文献

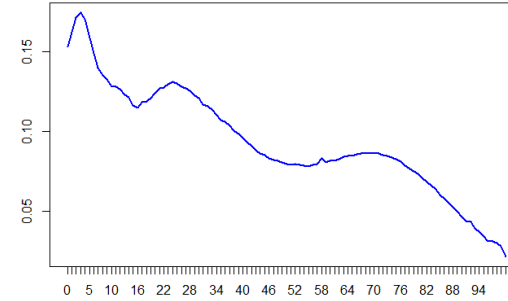
女性

Alpha



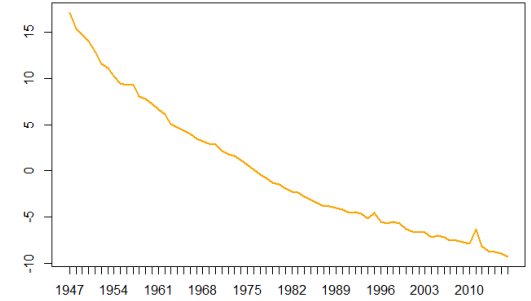
年齢

Beta



年齢

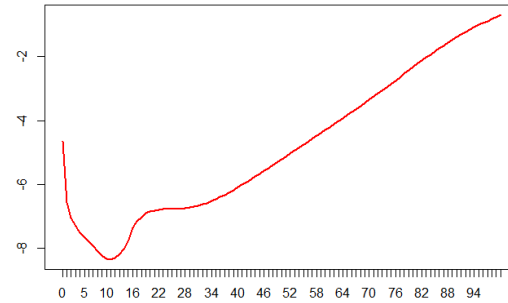
Kappa



時間

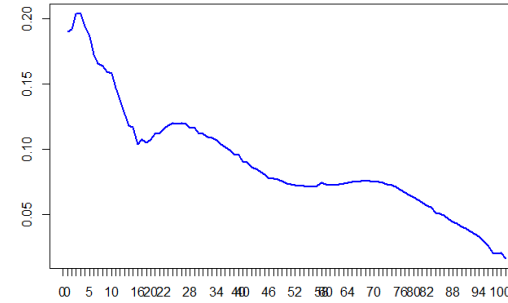
男性

Alpha



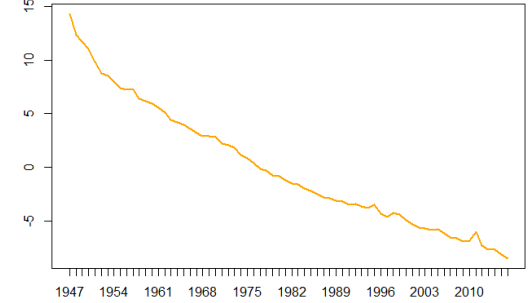
年齢

Beta



年齢

Kappa

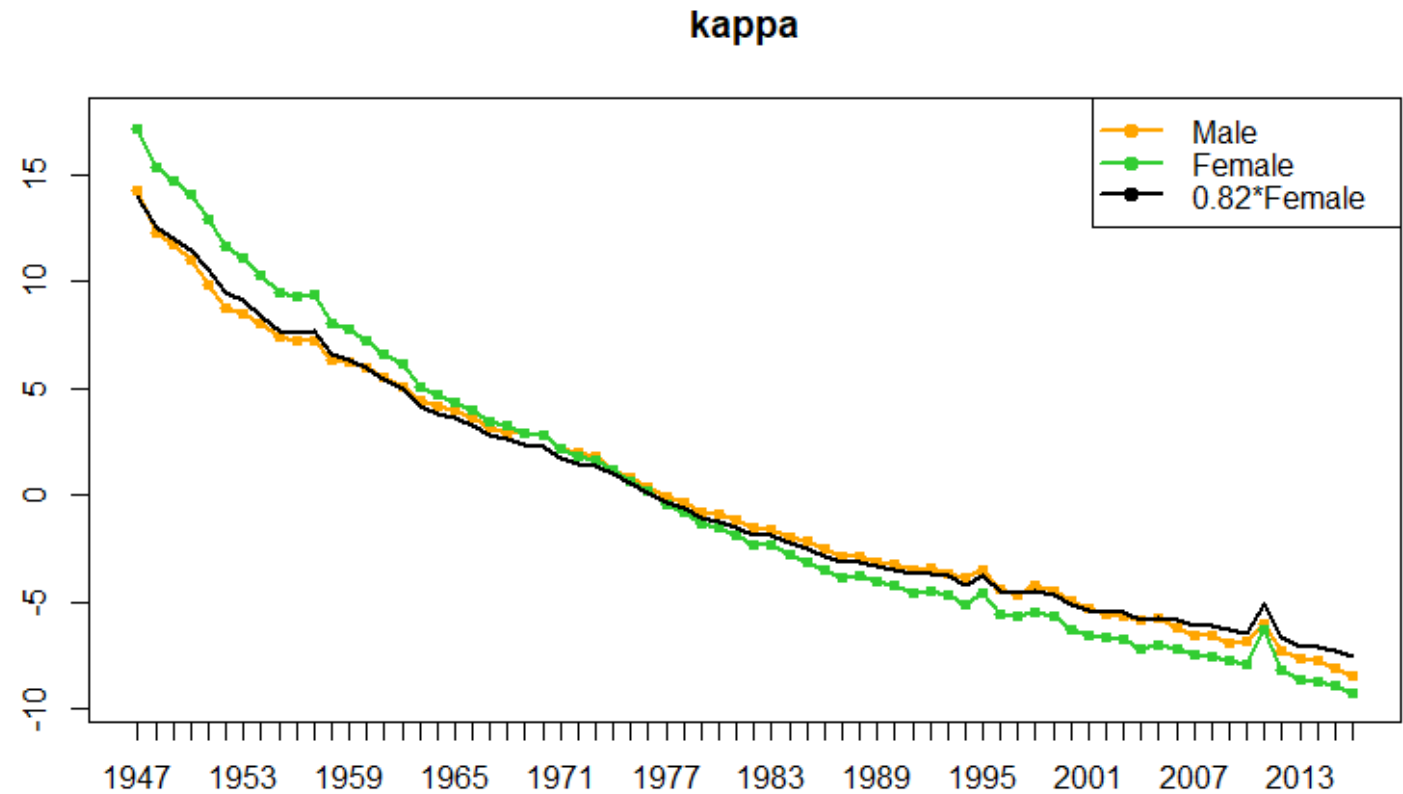


時間

Kappaは似た形状をしている

男女間の比較

- 研究動機
- Lee-Carterモデルの概説
- 先行研究
- モデルの提案
- 実際の解析
- まとめ
- 参考文献



スケールの違いを除いて同等であることが示唆



同一のパラメータでドライブされていると
考えられるのではないか

先行研究① 死因別モデル

□ 研究動機

□ Lee-Carterモデルの概説

□ 先行研究

- 死因別モデル
- テンソル型モデル

□ モデルの提案

□ 実際の解析

□ まとめ

□ 参考文献

- 死因別の死亡者数データに独立にLee-Carterモデルを当てはめたもの
- 死因別解析は以下の点で有効と考える
 - 死因ごとのリスク検知
 - 個々のトレンドを反映した商品開発
 - 個々のトレンドを反映した全体死亡率の予測
- オリジナルな死因別Lee-Carterモデルの先行研究は見つからなかった
 - Girosi and King(2007)に詳細な分析結果がまとめられている

Girosi and Kingによる実証によれば

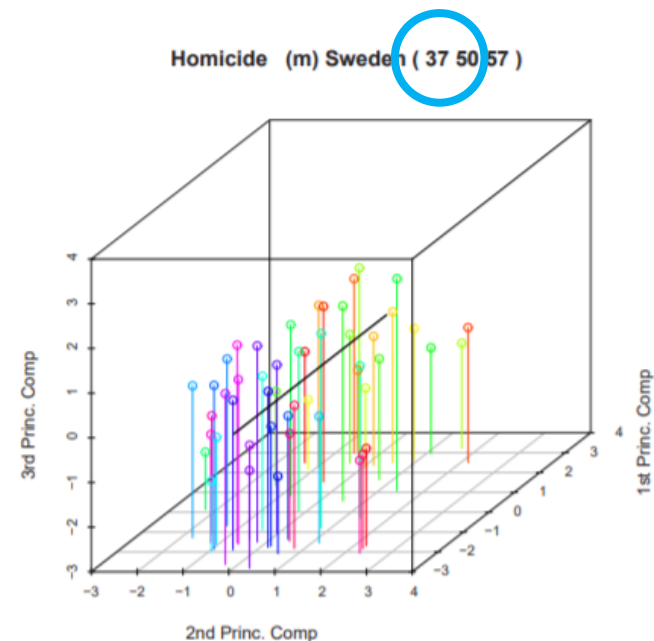
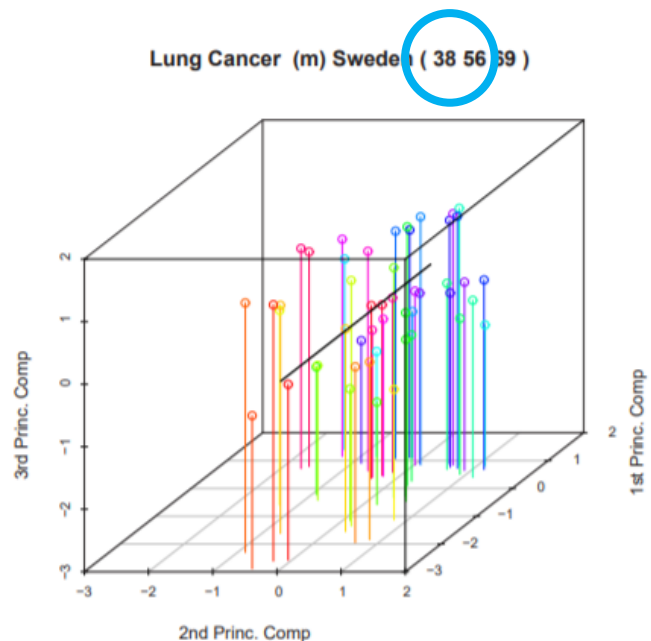
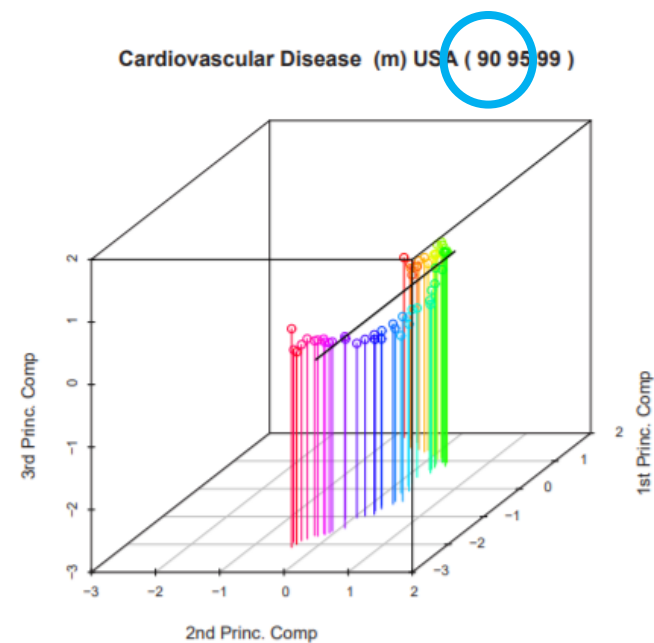
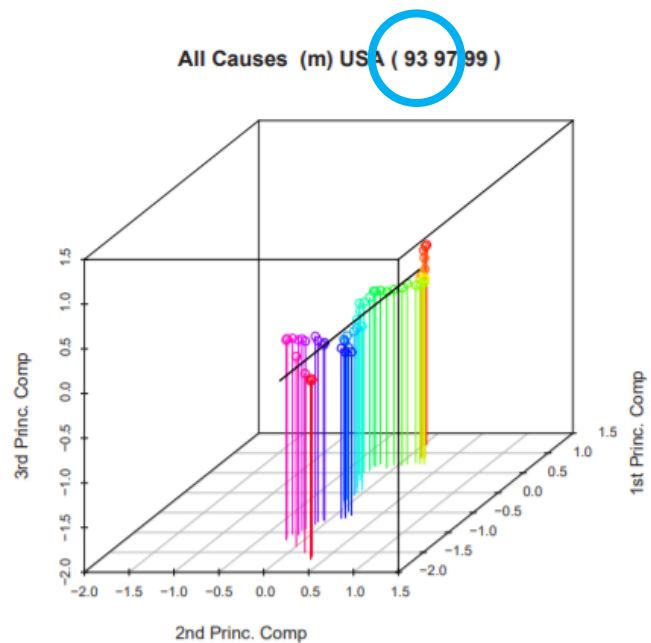
「Lee-Carterモデルは全死因のデータにはよく当てはまるが、死因別のデータには当てはまるものと当てはまらないものがある」と結論付け、

「死因別データに日常的に用いるべきではない」とした

他にも死因間の相互関係が無視されている、個々の死因のデータ数の少なさによる推定精度の不安定性などが問題視されている

先行研究① 死因別モデル

- 研究動機
- Lee-Carterモデルの概説
- 先行研究
 - 死因別モデル
 - テンソル型モデル
- モデルの提案
- 実際の解析
- まとめ
- 参考文献



先行研究② テンソル型 モデル

- 研究動機
- Lee-Carterモデルの概説
- 先行研究
 - 死因別モデル
 - テンソル型モデル
- モデルの提案
- 実際の解析
- まとめ
- 参考文献

- 野村(2018)によりテンソル解析を応用したモデルが提案された
この論文では年齢, 時間に都道府県の要素を追加し

$$\log m_{x,t,p} = \alpha_x + \sum_i \beta_x^{(i)} \kappa_t^{(i)} \gamma_p^{(i)}$$

という3階テンソルに対するモデルとなっている。

Lee-Carterモデルが行列の特異値分解を基礎としているのに対し
このモデルはテンソルの高次特異値分解という手法を基礎としている

- 1次近似で説明しきれない(論文では近似精度の面で3次としている)
 - パラメータの解釈が難しい
- といった問題を抱えている。

提案モデル

- 研究動機
- Lee-Carterモデルの概説
- 先行研究
- モデルの提案
- 実際の解析
- まとめ
- 参考文献

- 以上の先行研究と背景から、次のようなモデルを提案する

$$\log(1 - m_{x,t,c,s}) = \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K z_{i,j,k} \beta_x^{(i)} \kappa_t \lambda_c^{(j)} \xi_s^{(k)} + \varepsilon_{x,t,c,s}$$

- 観測死亡率に0が多く含まれるため $1 - m_{x,t,c,s}$ に対するモデルとした
- 誤差項にはテンソル正規分布を仮定した。
- 残差行列に特異値分解 → Lee and Carter(1992)
- 残差テンソルに高次特異値分解(CP分解) → 野村(2018)
- テンソルに高次特異値分解(Tucker分解) → 今回の提案モデル

テンソル正規分布

- 研究動機
- Lee-Carterモデルの概説
- 先行研究
- モデルの提案
- 実際の解析
- まとめ
- 参考文献

- テンソル正規分布は次の密度関数をもつ分布である

$$E_t = \{\varepsilon_{x,t,c,s}\}, \quad U_i^T U_i = \Sigma_i^{-1}$$
$$f(E_t) = C \exp \left[-\frac{1}{2} \| E_t \times_1 U_1 \times_2 U_2 \times_3 U_3 \|_F^2 \right]$$

- 共分散構造を軸ごとに推定することができるという特徴
 - 年齢間の共分散
 - 死因間の共分散
 - 性別間の共分散を独立に表現できる
- 本研究は年齢・性別間を対角行列とし死因間を一般の共分散行列とした

テンソル 正規分布

- 研究動機
- Lee-Carterモデルの概説
- 先行研究
- モデルの提案
- 実際の解析
- まとめ
- 参考文献

- クロネッカー積の性質から次の式が成り立つ

$$\Sigma_3 \otimes \Sigma_2 \otimes \Sigma_1 = \frac{1}{s_1 s_2} \Sigma_3 \otimes s_2 \Sigma_2 \otimes s_1 \Sigma_1$$

- スケールの一意性が無いため分散の値をそのまま評価できない
- ただし相関行列はそのまま比較することが可能

使用データ

- 研究動機
- Lee-Carterモデルの概説
- 先行研究
- 提案モデル
- 実際の解析
 - 使用データ
 - 解析結果
- まとめ
- 参考文献

本研究ではWHOが公開しているWHO mortality databaseより日本人の年齢, 暦年, 死因基本分類(ICD), 性別ごとの死亡者数・生存者数データを用いた.

死因基本分類は約10年おきに改定されており、後方互換性がないため今回の研究では現在の死因基本分類(ICD10)のデータ(1995年から2015年までのデータ)を用いた

またICD10は詳細(最大7万コード)であり個々の死因データ数が極端に少ないため厚生労働省が提示しているICD10準拠基本分類表の22項目へデータを圧縮して分析を行った

使用データ

- 研究動機
- Lee-Carterモデルの概説
- 先行研究
- 提案モデル
- 実際の解析
 - 使用データ
 - 解析結果
- まとめ
- 参考文献

ICD10対応標準病名マスター

1 ICD10分類

1 ICD10分類

	分類ID	分類表記
1	A00-B99	感染症及び寄生虫症 (1713)
2	C00-D48	新生物<腫瘍> (3351)
3	D50-D89	血液及び造血器の疾患並びに免疫機構の障害 (518)
4	E00-E90	内分泌, 栄養及び代謝疾患 (1297)
5	F00-F99	精神及び行動の障害 (795)
6	G00-G99	神経系の疾患 (1170)
7	H00-H59	眼及び付属器の疾患 (1270)
8	H60-H95	耳及び乳様突起の疾患 (308)
9	I00-I99	循環器系の疾患 (1209)
10	J00-J99	呼吸器系の疾患 (764)
11	K00-K93	消化器系の疾患 (1880)
12	L00-L99	皮膚及び皮下組織の疾患 (884)
13	M00-M99	筋骨格系及び結合組織の疾患 (1811)
14	N00-N99	腎尿路生殖器系の疾患 (988)
15	O00-O99	妊娠, 分娩及び産じょく<褥> (602)
16	P00-P96	周産期に発生した病態 (392)
17	Q00-Q99	先天奇形, 変形及び染色体異常 (1635)
18	R00-R99	症状, 徴候及び異常臨床所見・異常検査所見で他に分類されないもの (1310)
19	S00-T99	損傷, 中毒及びその他の外因の影響 (4902)
20	V01-Y98	傷病及び死亡の外因 (99)
21	Z00-Z99	健康状態に影響を及ぼす要因及び保健サービスの利用 (227)
22	U00-U99	特殊目的用コード (31)

使用データ

- 研究動機
- Lee-Carterモデルの概説
- 先行研究
- 提案モデル
- 実際の解析
 - 使用データ
 - 解析結果
- まとめ
- 参考文献

WHOが公表しているデータは年齢が各歳のデータではないが
本研究では各歳に展開するなどせずそのままデータを扱った

年齢区分の構成は下記の通り

番号	年齢区分
1	0歳
2	1歳から4歳
3	5歳から9歳
4	10歳から14歳
● ● ●	(以降5歳区分)
18	80歳から84歳
19	85歳以上

この結果今回用いるデータのテンソルは
次元 $19 \times 21 \times 19 \times 2$ の4階テンソルとなった

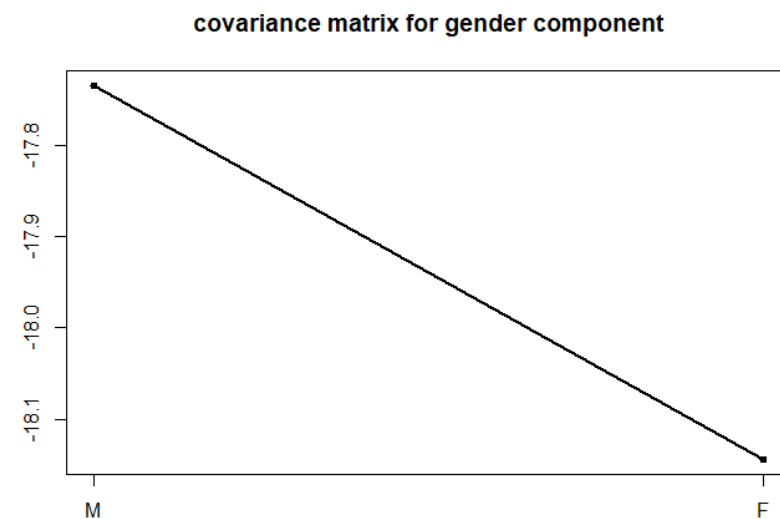
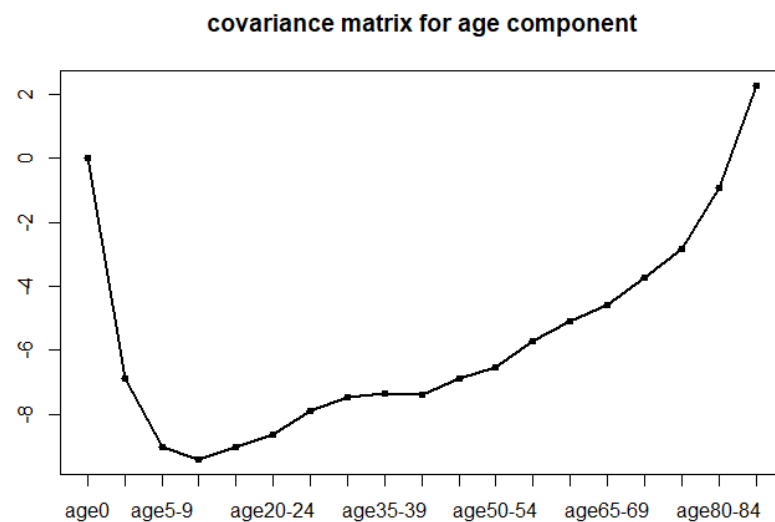
推定結果

- 研究動機
- Lee-Carterモデルの概説
- 先行研究
- 提案モデル
- 実際の解析
 - 使用データ
 - 解析結果
- まとめ
- 参考文献

推定結果は下記の通り

モデル	対数尤度	パラメータ数	AIC
提案モデル	163577	1296	-324562
死因別Lee-Carter	4844	2204	-5360

また年齢・性別間の分散(対数スケール)は下記の通り



κ_t の制約条件

- 研究動機
- Lee-Carterモデルの概説
- 先行研究
- 提案モデル
- 実際の解析
 - 使用データ
 - 解析結果
- まとめ
- 参考文献

提案モデル(誤差項を除く)の κ_t 以外の項を

$$\hat{\beta}_{x,c,s} = \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K z_{i,j,k} \beta_x^{(i)} \lambda_c^{(j)} \xi_s^{(k)}$$

のようにおくと

$$\begin{aligned} \log(1 - m_{x,t,c,s}) &= \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K z_{i,j,k} \beta_x^{(i)} \kappa_t \lambda_c^{(j)} \xi_s^{(k)} \\ &= \hat{\beta}_{x,c,s} \kappa_t \end{aligned}$$

と表現できる

ここで $0 \leq m_{x,t,c,s} < 1$ であるため $\hat{\beta}_{x,c,s} \kappa_t \leq 0$ を満たす必要がある
 本研究では $\kappa_t \geq 0$ を仮定し、 $\log \kappa_t$ に対してARIMAモデルを当てはめた
 AICによりモデル選択をしたところ、ARIMA(0,1,1)が選択された

ARIMAモデル の推定結果

- 研究動機
- Lee-Carterモデルの概説
- 先行研究
- 提案モデル
- 実際の解析
 - 使用データ
 - 解析結果
- まとめ
- 参考文献

得られたモデルの式と推定されたパラメータは下記の通りである

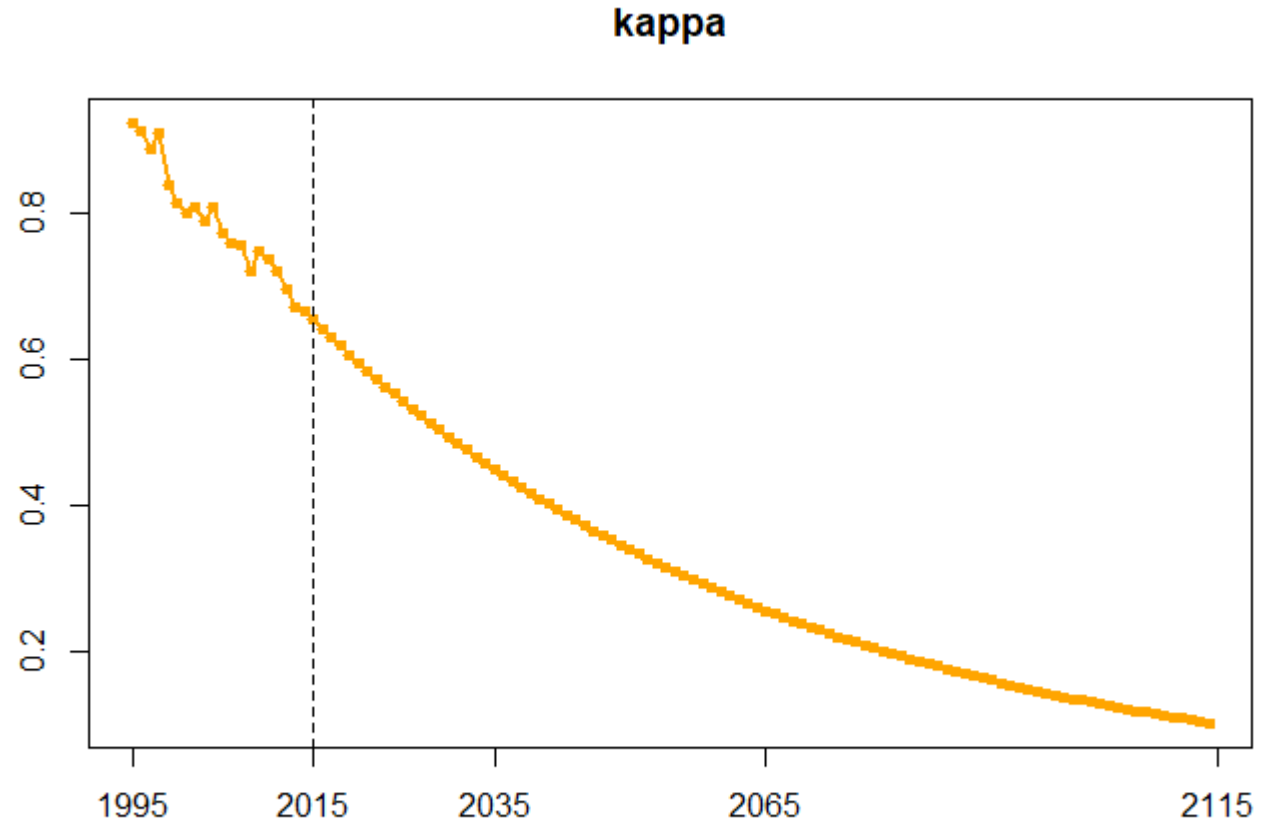
$$\log \kappa_t = \log \kappa_{t-1} + d + \epsilon_t + \theta \epsilon_{t-1}, \quad \epsilon_t \sim N(0, \sigma_\epsilon^2)$$

パラメータ	推定値	標準誤差
d	-0.0188	0.0033
θ	-0.5220	0.2628
σ_ϵ^2	0.0008265	-

κ_t の将来推定

- 研究動機
- Lee-Carterモデルの概説
- 先行研究
- 提案モデル
- 実際の解析
 - 使用データ
 - 解析結果
- まとめ
- 参考文献

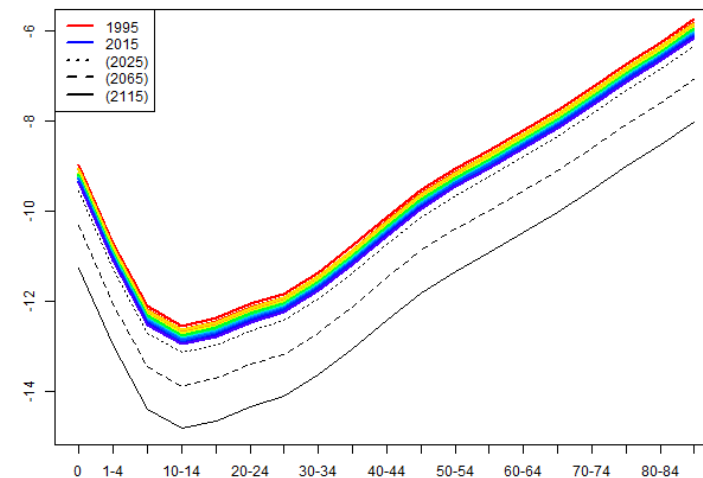
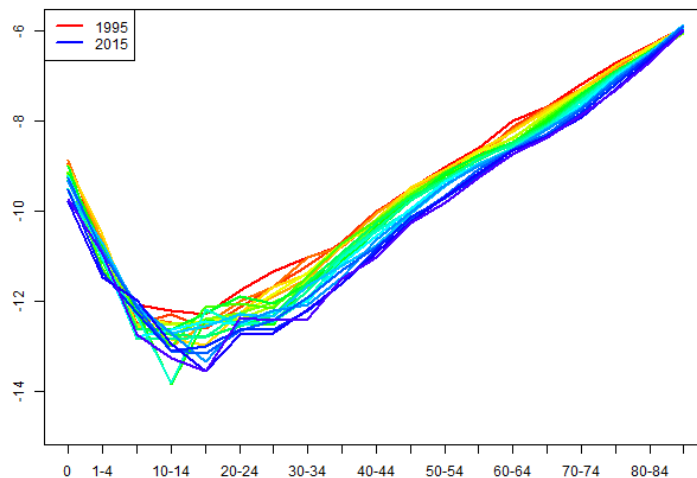
Kappaの平均構造は下記の通り



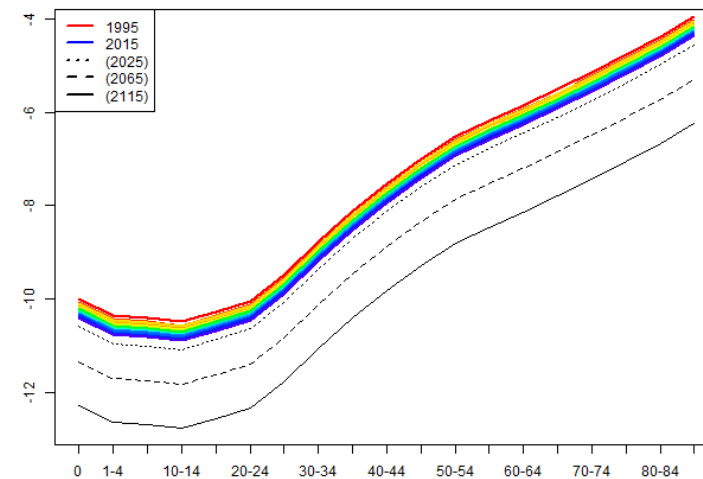
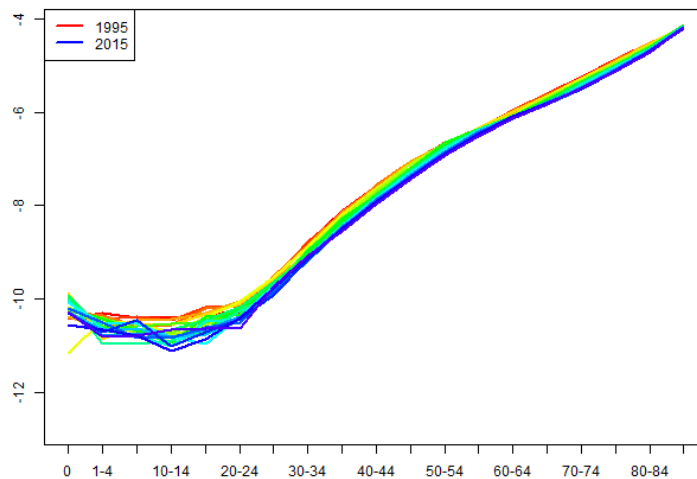
将来推定結果

- 研究動機
- Lee-Carterモデルの概説
- 先行研究
- 提案モデル
- 実際の解析
 - 使用データ
 - 解析結果
- まとめ
- 参考文献

感染症及び寄生虫症：男性



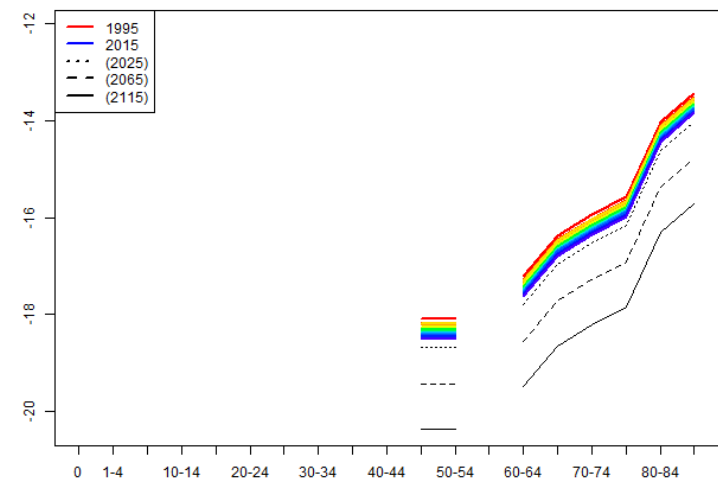
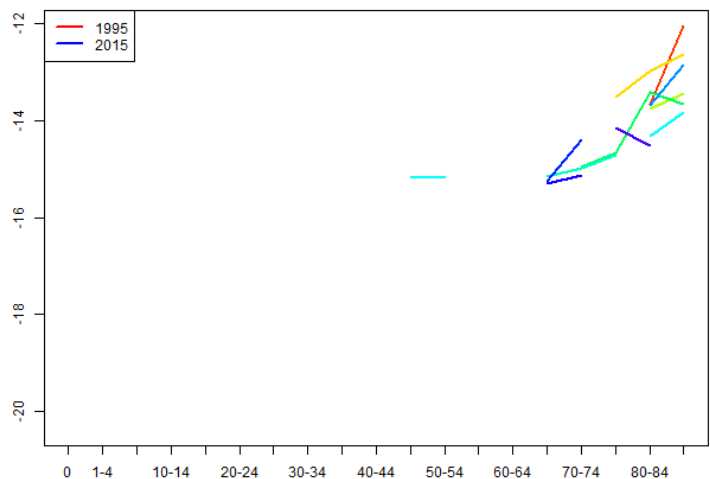
新生物<腫瘍>：女性



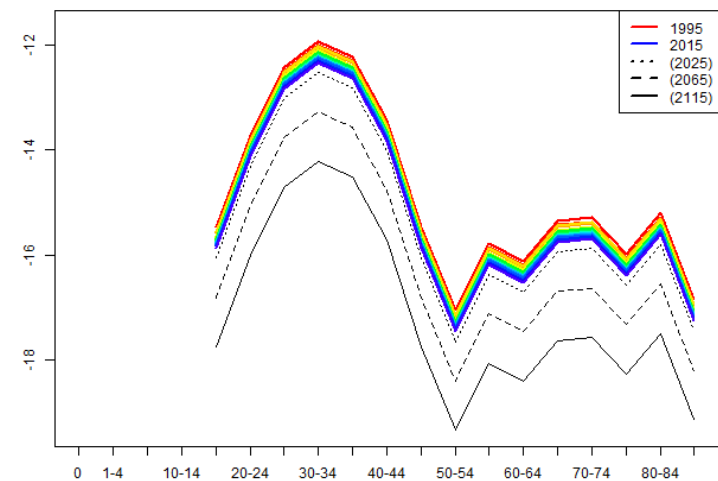
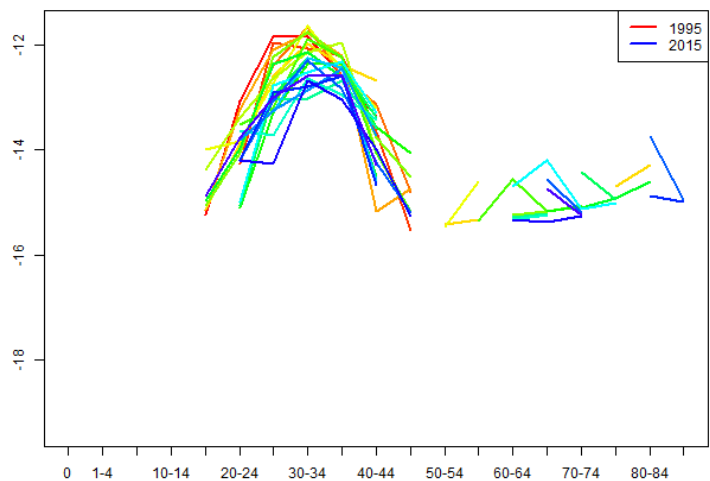
将来推定結果

- 研究動機
- Lee-Carterモデルの概説
- 先行研究
- 提案モデル
- 実際の解析
 - 使用データ
 - 解析結果
- まとめ
- 参考文献

眼及び付属器の疾患：男性



妊娠, 分娩および産じょく：女性



時間平均項のあるモデルとの比較

- 研究動機
- Lee-Carterモデルの概説
- 先行研究
- 提案モデル
- 実際の解析
 - 使用データ
 - 解析結果
- まとめ
- 参考文献

提案モデル

$$\log(1 - m_{x,t,c,s}) = \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K z_{i,j,k} \beta_x^{(i)} \kappa_t \lambda_c^{(j)} \xi_s^{(k)} + \varepsilon_{x,t,c,s}$$

比較モデル

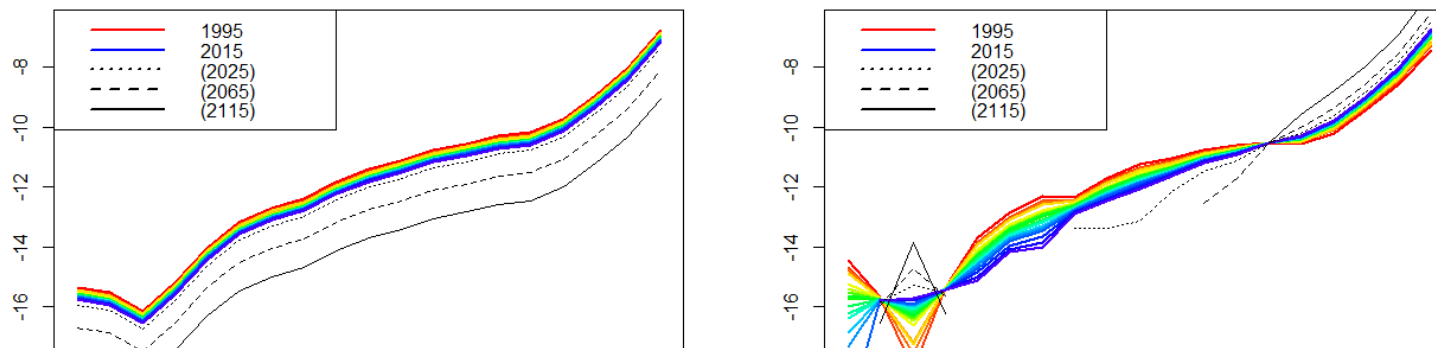
$$\log(1 - m_{x,t,c,s}) = \alpha_{x,c,s} + \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K z_{i,j,k} \beta_x^{(i)} \kappa_x \lambda_c^{(j)} \xi_s^{(k)} + \varepsilon_{x,t,c,s}$$

モデル	対数尤度	パラメータ数	AIC
提案モデル	163577	1296	-324562
比較モデル	168500	2018	-332964

時間平均項のあるモデルとの比較

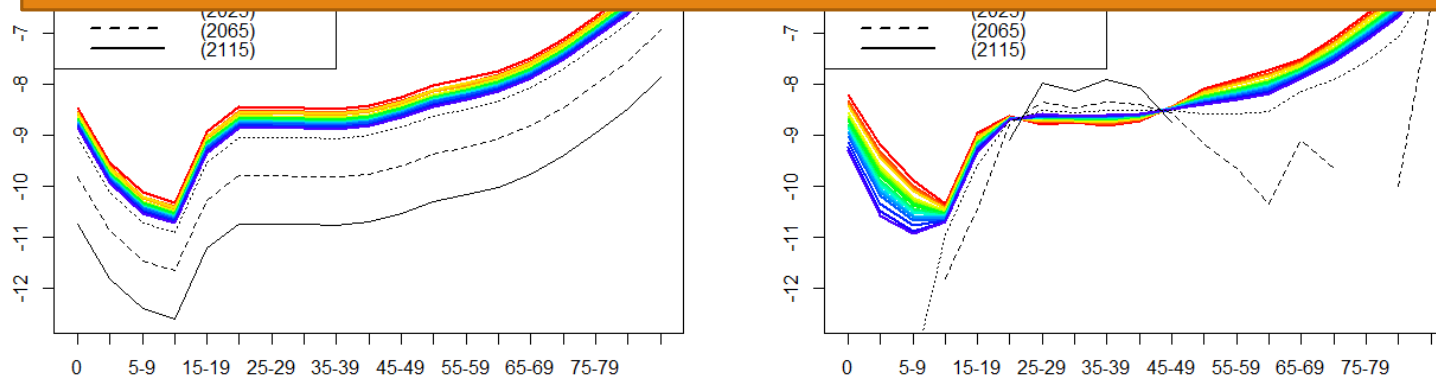
- 研究動機
- Lee-Carterモデルの概説
- 先行研究
- 提案モデル
- 実際の解析
 - 使用データ
 - 解析結果
- まとめ
- 参考文献

精神及び行動の障害



時間平均項のあるモデルによる長期予測はグラフの概形を保てず死亡率改善を過大評価していることが示唆された

このことから長期の死亡率予測には適さないと判断した

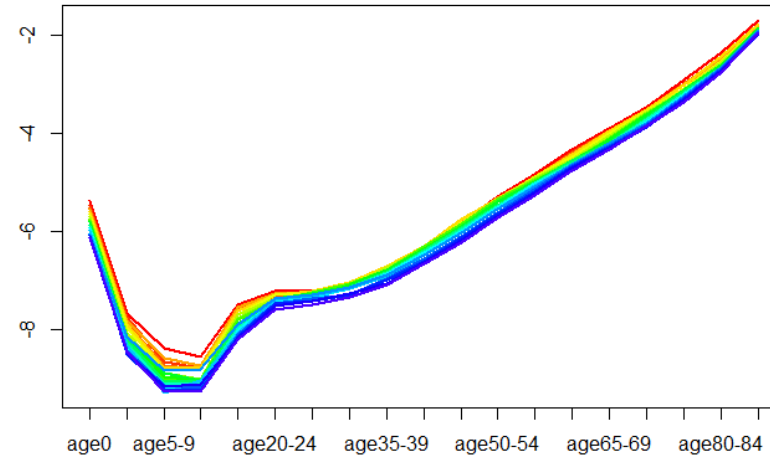


Lee-Carter モデルとの 比較

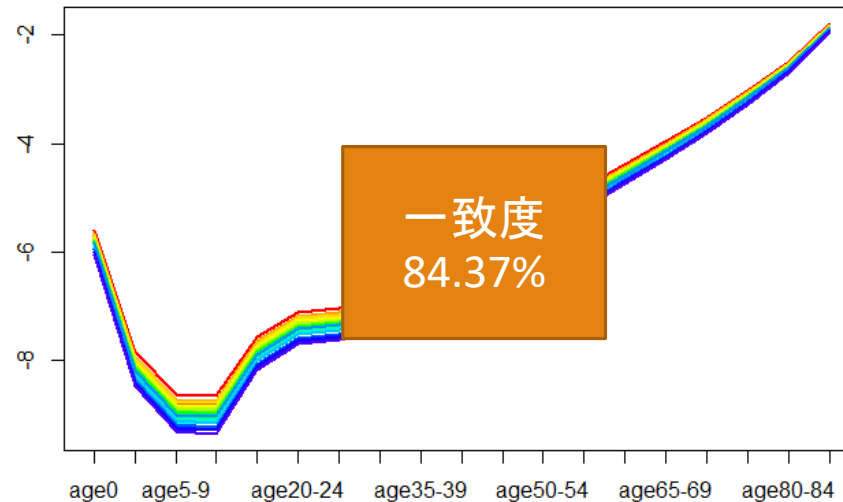
- 研究動機
- Lee-Carterモデルの概説
- 先行研究
- 提案モデル
- 実際の解析
 - 使用データ
 - 解析結果
- まとめ
- 参考文献

死因別死亡率を合算した全体死亡率をLee-Carterモデルと比較する

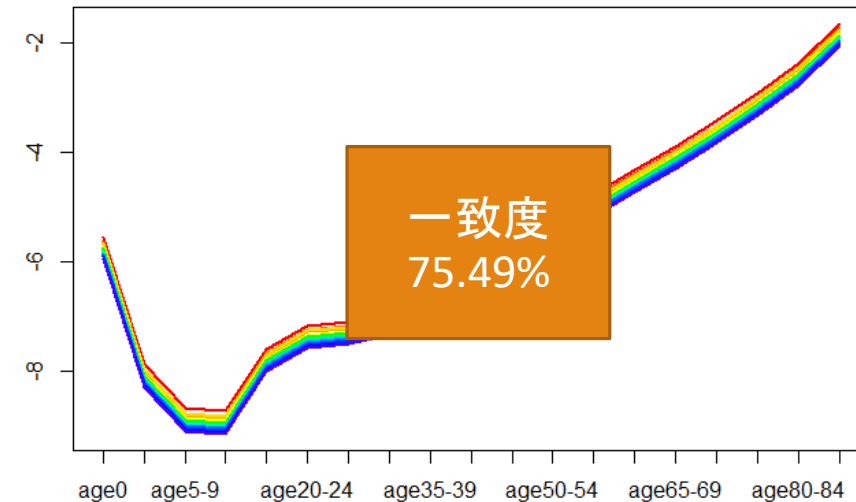
観測データ



Lee-Carterモデルによる推定値



提案モデルによる推定値

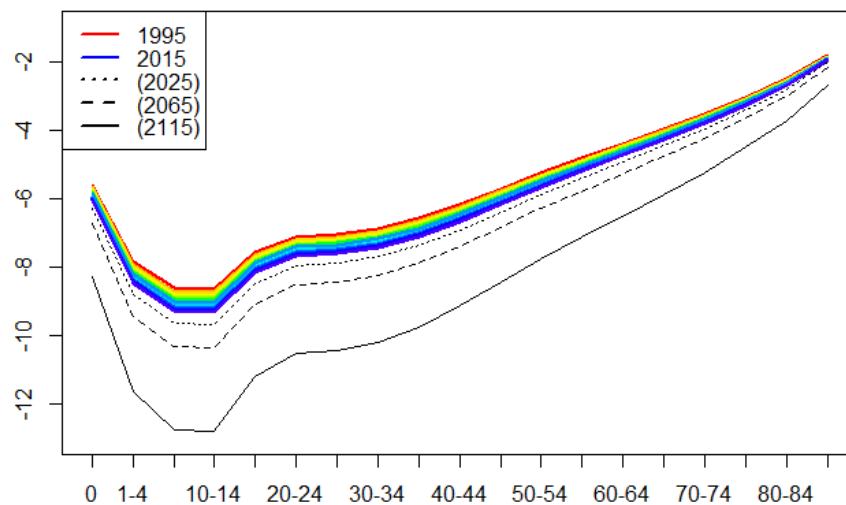


Lee-Carter モデルとの 比較

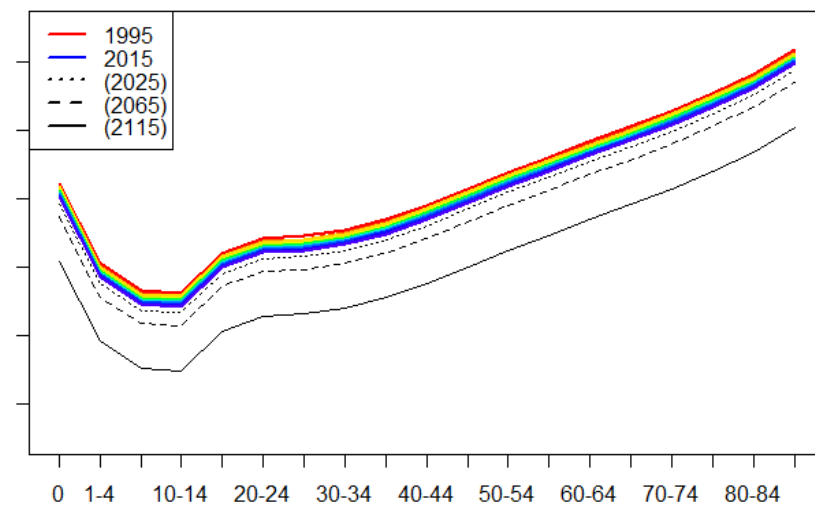
- 研究動機
- Lee-Carterモデルの概説
- 先行研究
- 提案モデル
- 実際の解析
 - 使用データ
 - 解析結果
- まとめ
- 参考文献

死因別死亡率を合算した全体死亡率をLee-Carterモデルと比較する
予測平均による比較

Lee-Carterモデルによる推定値



提案モデルによる推定値

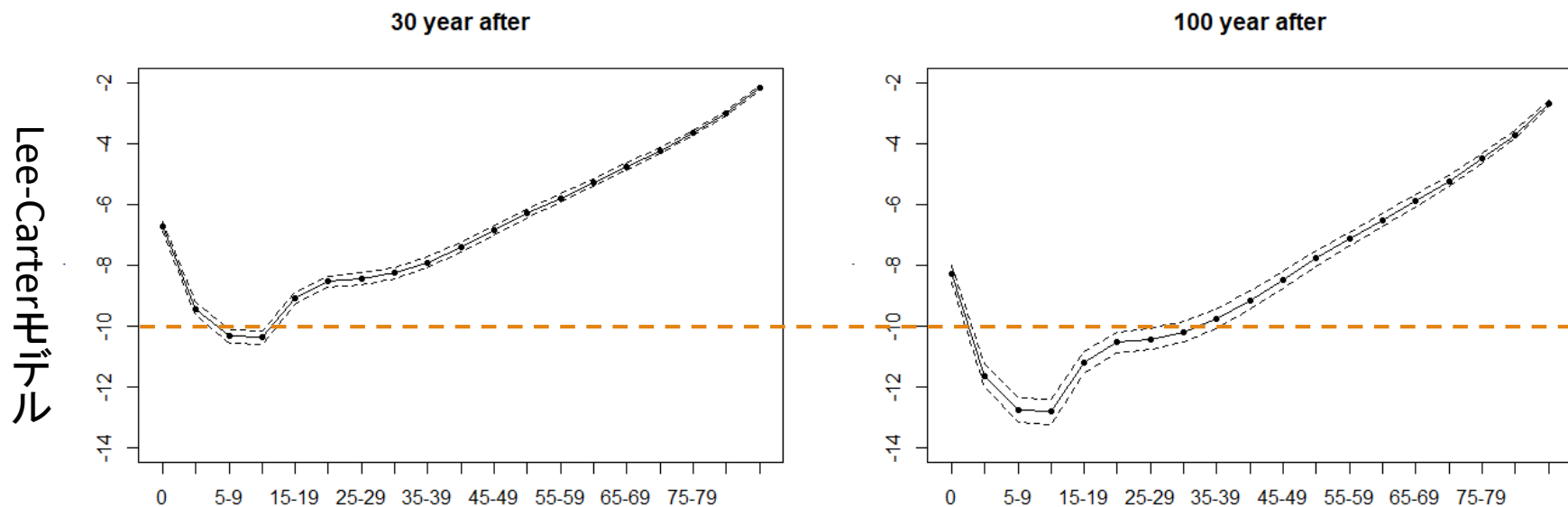
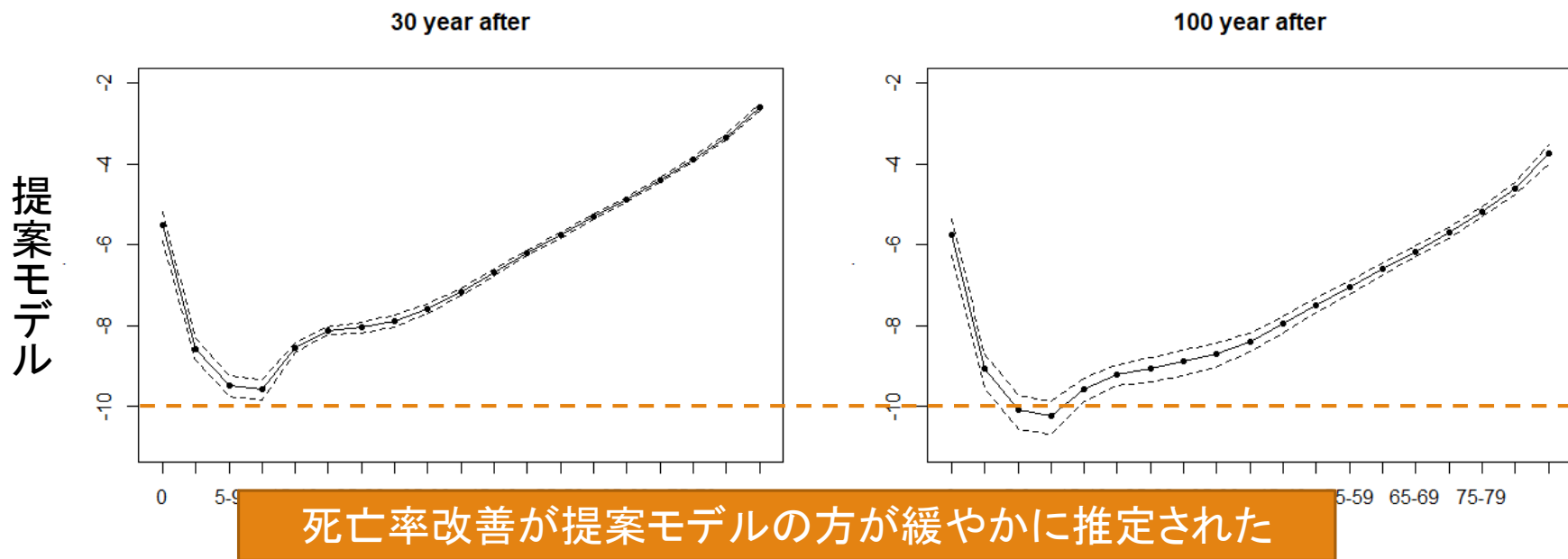


低年齢層と高年齢層で違いが見られた

Lee-Carter モデルとの 比較

- 研究動機
- Lee-Carterモデルの概説
- 先行研究
- 提案モデル
- 実際の解析
 - 使用データ
 - 解析結果
- まとめ
- 参考文献

シミュレーションによる比較



まとめ

- 研究動機
- Lee-Carterモデルの概説
- 先行研究
- 提案モデル
- 実際の解析
- まとめ
- 参考文献

- Lee-Carterモデルとテンソル分解のアイデアから新たな死亡率モデルの提案を行った
- 死因別死亡率を同時推定し、かつテンソル正規分布を用いて死因間の相互関係をモデルに取り入れることで全体死亡率と統合的な推定が可能
- 死因別Lee-Carterモデルと比較してAICを非常に低くすることが出来た
- 全体死亡率をLee-Carterモデルと比較して長期予測で形を崩さない推定が出来た
- 死因ごとの死亡率トレンドを把握することで国の医療政策や保険会社の商品政策に利用できる
- ✓ 死因分類の大幅な改定があった場合の比較可能性の問題
- ✓ 誤差項に正規分布を仮定する妥当性

参考文献

□ 研究動機

□ Lee-Carterモデルの概説

□ 先行研究

□ 提案モデル

□ 実際の解析

□ まとめ

□ 参考文献

1. Booth, H., Maindonald, J., and Smith, L. (2002). Applying Lee-Carter under conditions of variable mortality decline. *Population studies*, 56(3), 325-336.
2. Brouhns, N., Denuit, M., and Vermunt, J. K. (2002). A Poisson log-bilinear regression approach to the construction of projected lifetables. *Insurance: Mathematics and economics*, 31(3), 373-393.
3. Cairns, A. J. (2000). A discussion of parameter and model uncertainty in insurance. *Insurance: Mathematics and Economics*, 27(3), 313-330.
4. Girosi, F., and King, G. (2007). Understanding the Lee-Carter mortality forecasting method. Gking. Harvard. Edu.
5. Hunter, G., and Luis, C. (2018). An expectation-maximization algorithm for the matrix normal distribution with an application in remote sensing. *Journal of Multivariate Analysis*, 167, 31-48.
6. Kolda, T. G., and Bader, B. W. (2009). Tensor decompositions and applications. *SIAM review*, 51(3), 455-500.
7. Lee, R. D., and Carter, L. R. (1992). Modeling and forecasting US mortality. *Journal of the American statistical association*, 87(419), 659-671.
8. Lee, R. (2000). The Lee-Carter method for forecasting mortality, with various extensions and applications. *North American actuarial journal*, 4(1), 80-91.
9. Li, J. S. H., Hardy, M. R., and Tan, K. S. (2009). Uncertainty in mortality forecasting: an extension to the classical Lee-Carter approach. *ASTIN Bulletin: The Journal of the IAA*, 39(1), 137-164.
10. Manceur, A. M., and Dutilleul, P. (2013). Maximum likelihood estimation for the tensor normal distribution: Algorithm, minimum sample size, and empirical bias and dispersion. *Journal of Computational and Applied Mathematics*, 239, 37-49.
11. Renshaw, A. E., and Haberman, S. (2003). Lee-Carter mortality forecasting with age-specific enhancement. *Insurance: Mathematics and Economics*, 33(2), 255-272.
12. Renshaw, A. E., and Haberman, S. (2006). A cohort-based extension to the Lee-Carter model for mortality reduction factors. *Insurance: Mathematics and economics*, 38(3), 556-570.
13. 国立社会保障・人口問題研究所(2017). 「日本の将来推計人口」, 『人口問題研究資料』
14. 野村俊一(2018). 「テンソル分解を利用したLee-Carterモデルの拡張と都道府県別生命表解析」, 『JARIP研究発表大会2018予稿集』

ご清聴ありがとうございました。
