

睡眠見守りセンサーデータの構造方程式モデリングによる因果分析

大石 信弘^{1,*} 山本 直樹¹ 石田 明男² 村上 純¹

A Causal Analysis by Structural Equation Modeling of Sleep Monitoring Sensor Data

Nobuhiro Oishi^{1,*}, Naoki Yamamoto², Akio Ishida³, Jun Murakami²

In this paper, structural equation modeling (SEM) is used to analyze the causal relationship between sleep stages and environmental data. The data used for the analysis is obtained by a care support device used in an elderly care facility. By applying the stepwise method to this data, we were able to find four observation variables that affect sleep stages. The latent variables are then determined by a scree plot. We proposed a causal model in which four observed variables and two latent variables affect sleep levels. Statistical analysis environment R and the lavaan package were used for SEM analysis in this paper. The results show that SEM can be used to build a rational model for the effects of vital signs and environmental conditions on good sleep.

キーワード：睡眠見守りセンサー、因果分析、構造方程式モデリング、R言語

Keywords : sleep monitoring sensor, causal analysis, structural equation modeling, R language

1. はじめに

近年、大量のデータが容易に手に入るビッグデータの時代になった⁽¹⁾。利用可能なデータとしては、ID付POSデータやネット通販の購買データ、ウェブサイトの閲覧統計やSNSでの発信統計、乗用車の自動運転やウェアラブル端末等に搭載された各種センサーによる物理量データや位置情報、生体情報、Nシステムに代表される固定カメラの画像から得られる各種の情報などがある。

観測される情報間の関係性を統計的に分析することで、今まで曖昧だった観測変数間の関係性をはっきりと示すことができるようになった。例えば、重回帰分析、さらには潜在変数を組み込んでモデル化した探索的因子分析(EFA; Exploratory Factor Analysis)や確認的因子分析(CFA; Confirmatory Factor Analysis)などが挙げられる。

観測された情報間には、関係性があるばかりでなく、観測変数と観測変数の間に因果関係が存在するものもある。この因果関係を明らかにする分析法として、グラフィカルモデリング(GM; Graphical Modeling)⁽³⁾や構造方程式モデリング(SEM; Structural Equation Modeling)^(3,6-8)がある。

翻ってこれからの日本が直面する社会的な問題を考えてみると、本格的な高齢化社会と同時に出生率低下が引き起こす労働力人口の減少問題がある。この問題がすでに生じ

ている現場もあり、特に老人介護施設における介護福祉士の人手不足は深刻で、就労者への過度な労働負担による入所者へのサービス低下などの問題が起きている⁽²⁾。その例として、入所者が夜間眠れずにベッドを離床すると、当直勤務の介護士がその対応に追われ、本来他の入所者を世話するための時間が削られるという問題が挙げられる。

そこで、入所者によりよい睡眠をとってもらえれば、介護福祉士の労働負担を少しでも軽くできると考えた。本論文では、老人介護施設で使用されている介護サポート機器から得られる環境データと睡眠の深さとの因果関係を分析することで、よりよい睡眠を得るための条件を見つけることを目的とした。それにより、入所者各人がよりよい睡眠を得られるようになり、健康な生活を送ることができるとともに、入床後の巡回タイミングを適切にでき、一人当たりの介護福祉士にかかる労働負担を軽減することが可能になると期待できる。

本論文におけるデータの分析には、統計解析環境R⁽⁴⁾を用いた。Rはフリーウェアであり、教育的にもこの言語の使用が有益であると考えられる⁽⁵⁾。因果関係の分析手法としては構造方程式モデリング(SEM)を採用し、Rのlavaanパッケージ⁽⁹⁾を用いて実装した。このパッケージは現在も開発が続いているパッケージではあるが、分析結果は厳密で正確であると分析性能については定評がある⁽⁷⁾。また、このパッケージで採用されているモデル記述文法も簡単であるとともに、様々なモデルを表現できる汎用性がある。

2. 因果分析に用いるデータの選択

本論文では、ASD社製の介護施設向け睡眠見守りセンサー「まもる〜の」⁽¹⁰⁾を実際に施設で利用したデータの提供

¹ 電子情報システム工学系
〒861-1102 熊本県合志市須屋 2659-2
Faculty of Electronics and Information Systems Engineering,
2659-2 Suya, Koshi-shi, Kumamoto, Japan 861-1102

² リベラルアーツ系
〒861-1102 熊本県合志市須屋 2659-2
Faculty of Liberal Arts,
2659-2 Suya, Koshi-shi, Kumamoto, Japan 861-1102

* Corresponding author:
E-mail address: oishi@kumamoto-nct.ac.jp (N. Oishi).

を受けて分析を行った。この装置は介護施設のベッドに取り付け、気温、湿度、気圧、照度、生体情報(心拍数、呼吸数、睡眠レベル)、入床データ(起床、入床、離床)を観測し、クラウドへアップロードする。測定間隔は1分間隔モードと5分間隔モードの2つの測定モードがあり、1ヵ月分のデータがクラウド上に自動的に保存・保管される。

今回の分析のために用意したデータは1人分に関する、5分間隔の測定モードで得られた室温 (*tm*)、湿度 (*hm*)、気圧 (*at*)、照度 (*il*)、1分前の呼吸数 (*rr1*)、1分前の心拍数 (*hr1*) および睡眠レベル (*ss*) の7個の観測変数である(括弧内に観測変数名を示す)。*ss* はレベル1からレベル4の4値を取る。その他の観測変数は、正の実数である。分析に用いた時系列データは欠損値のある時点を除く2926時点(約10日分)のものである。

図1に時系列データの一部の観測変数のうち、ある時刻から480分間(8時間)のデータを抜粋して示す。*ss* や *rr1* および *hr1* は値の変化が激しく、一見するとランダムであるかのようにさえ見える。一方、*hm* は変化が少なく、ある時刻を境にして高値から低値へと変わっていることが分かる。

7個の観測データ間関係を見るために、相関図行列を

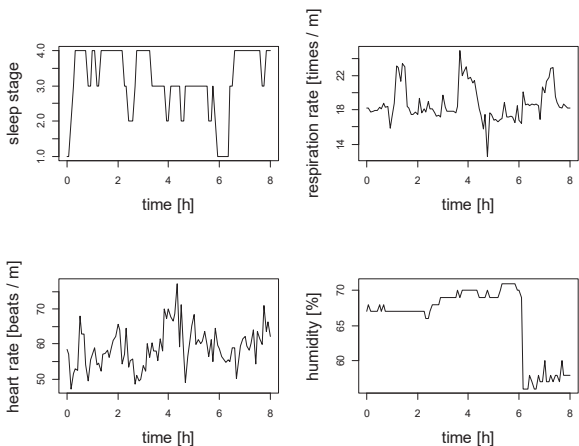


図1 観測データの時系列表示(抜粋)

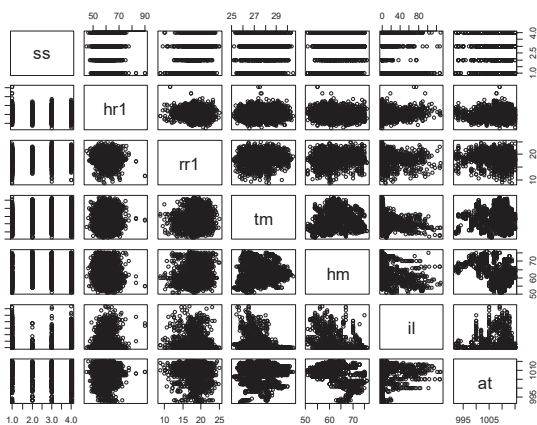


図2 観測データの散布図行列

表1 ステップワイズ法による変数選択の過程

算法	選択した変数	AIC	算法
変数減少法 ↓	<i>hr1, rr1, tm, hm, il, at</i>	775.55	変数増減法 ↑
	<i>hr1, rr1, tm, hm, il</i>	774.22	
	<i>hr1, rr1, hm, il</i>	773.84	
	<i>il, hr1, rr1</i>	780.23	
	<i>il, hr1</i>	817.61	
	<i>il</i>	829.55	
	なし	990.38	

図2に示す。*rr1*には0の値があるが、これは睡眠時無呼吸症などの兆候を示すと考えられるため、外れ値としての処理はしていない。各変数間には強い相関を見ることができず、*ss*と各変数間の相関関係をこの行列から読み取ることが難しい。

このように *ss* との関係性が判然としない7個の観測変数のすべてを使って因果分析をしたところ、モデルの適合度指標が良適合を示さなかった。(例えば後述の指標では、AGFI = 0.936、RMSEA = 0.103など)。そのため、分析に使用する変数を減らすことにした。変数選択⁽¹¹⁾にはステップワイズ法⁽¹²⁾を用い、次式の赤池情報量基準(AIC; Akaike's Information Criterion)をモニターしながら、採用する観測変数を決定した⁽¹²⁾。

$$AIC = m \times \ln \frac{\sum_{i=1}^n \{y^{(i)} - y_{EST}^{(i)}\}^2}{m} + 2 \dots \dots \dots (1)$$

n: サンプル数

m: 説明変数の数

$y^{(i)}$: *i* 番目のサンプルにおける目的変数の値

$y_{EST}^{(i)}$: *i* 番目のサンプルにおける目的変数の推定値

実際には、目的変数を *ss* として、Rのstep関数を用いて実施した。その結果、表1に示すように、*il*、*hm*、*hr1* および *rr1* の4個の変数を採用して *ss* を表した場合にAICが最小となることが分かった。この結果から、目的変数を含めて5個の観測変数を使用して分析を行うことにした。

3. 潜在変数の導入

上述の5個の変数の背後にある構成概念を表すために、潜在変数を導入する。その個数を検討するために、観測変数間の相関係数行列の固有値を求め、図3のスクリーンプロットを得た。この図から、第3固有値以降はほぼなだらかな曲線となるので、潜在変数の個数は2個とした。おおよそ1以上の値を持つ固有値の個数も2個であるので、カイザー基準⁽⁷⁾にも合致している。

そこで、2個の潜在変数が表す構成概念と、それぞれの潜在変数に結び付ける観測変数を以下のようにする。

f_1 (室内環境): 照度 (*il*) および湿度 (*hm*)

f_2 (バイタルサイン): 心拍数 (*hr1*) および呼吸数 (*rr1*)

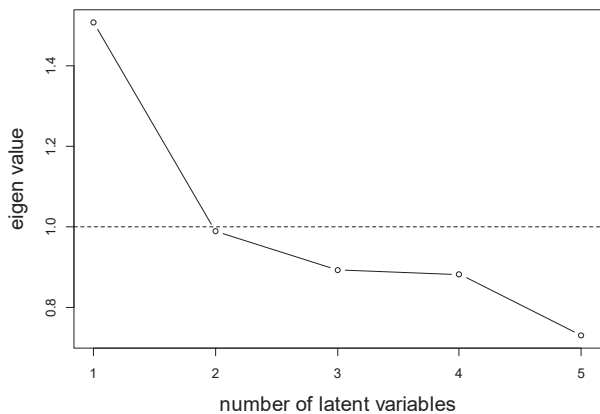


図3 5変数選択時のスクリープロット

4. 提案モデルとパス係数

因果関係を表すには、多重指標モデルや、多重指標多重原因 (MIMIC ; Multiple Indicator Multiple Cause) モデル、部分的最小二乗 (PLS ; Partial Least Squares) モデルなどがある⁷⁾。今回明らかにしたいのは、観測変数である睡眠レベル (*ss*) に、他の潜在変数や観測変数がどう影響を及ぼしているかということであるので、MIMIC モデルや PLS モデルが適しているから、図4に示すモデルを提案する。室内環境 (*f1*) がバイタルサイン (*f2*) に影響を及ぼすことはあまり考えられず、その逆方向の影響も考えにくい。そのため、このモデルにおいて、*f1* と *f2* は互いに直交していると仮定した。このモデルのパス係数および *ss* に結び付いている誤差分散を求めるための測定方程式と構造方程式は次のようになる。

$$\begin{cases} f1 = \alpha_1 \cdot il + \alpha_2 \cdot hm \\ f2 = \alpha_3 \cdot hr1 + \alpha_4 \cdot rr1 \\ ss = \gamma_1 \cdot f1 + \gamma_2 \cdot f2 + e1 \end{cases} \dots\dots\dots (2)$$

ここで、*f1* と *f2* は内生変数ではあるが、誤差変数を伴っていない。これは PLS モデルでは左辺に置かれた内生変数である潜在変数が右辺の外生変数によって定義されると考えるためである。つまり、室内環境 (*f1*) は照度 (*il*) と湿度 (*hm*) により定義されている構成概念と捉え、同様に、

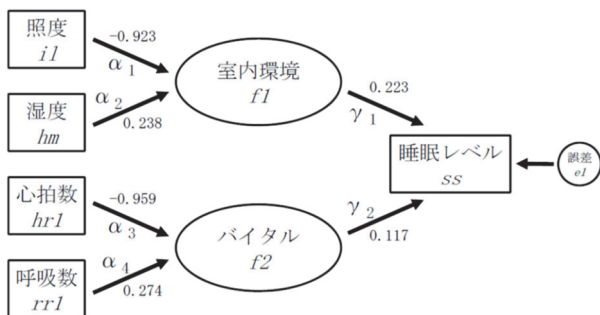


図4 提案モデルのパス図

```
# lavaan パッケージを用いて、提案モデルに SEM を適用する
# 変数選択後、標準化したデータが"data.dat"に格納されている
# 適合度指標 : GFI, AGFI, RMSEA, CFI, SRMR, AIC, BIC

library(lavaan)

# model : 図4の提案モデルを lavaan の文法で記述
model <- '
f1 ~ il + hm
f2 ~ hr1 + rr1
f1 + f2 =~ ss
f1 =~ 0*f1
f2 =~ 0*f2
f1 =~ 0*f2
'

fit <- lavaan::sem(model, data=data.dat, orthogonal=T,
fixed.x=T, std.lv=F)
summary(fit, standardized=T)
fitMeasures(fit, fit.measures =
c("gfi", "agfi", "rmsea", "cfi", "srmr", "aic", "bic"))
```

図5 lavaan のスクリプト

バイタルサイン (*f2*) は心拍数 (*hr1*) と呼吸数 (*rr1*) によって定義される構成概念と捉えている。そのため、*f1* も *f2* も誤差変数を伴わない。

(2)式を、lavaan のモデル記述文法で記述し、lavaan パッケージの *sem* 関数を用いて母数を推定した。推定には最尤 (ML ; Maximum Likelihood) 推定法を用いた。なお、解析に用いた観測データは標準化したものを用いた。このスクリプトを図5に示す。標準化した母数の推定値は表2のようになった。表中で、パス係数 α_1 と α_3 が負の値となっている。これは、照度 (*il*) が強く (明るく) なれば室内環境の値 (*f1*) が小さくなり、逆に、暗くなれば室内環境の値が大きくなることを意味する。同様に、心拍数 (*hr1*) が大きく (早く) なればバイタルサインの値 (*f2*) が小さくなり、心拍数が小さく (遅く) なればバイタルサインの値が大きくなることを意味している。パス係数の大きさは、その値が大きいくほど、影響が大きいことを表している。これらのパス係数は図4にも記入した。

提案モデルの適合度指標⁷⁾を表3に示す。GFI (Goodness of Fit Index)、AGFI (Adjusted GFI) および CFI (Comparative Fit Index) は観測変数の分散に対するモデルの説明率という観点から分析の精度を評価する指数で、0 から 1 までの値を取り、0.95 より大であると良適合と判断される。この値が 1.000 となっているのは、提案モデルの自由度が 0 であるためである。RMSEA (Root Mean Square Error of Approximation) と SRMR (Standardized Root Mean Square

表2 標準化した母数の推定値

α_1	α_2	α_3	α_4	γ_1	γ_2	<i>e1</i> の分散
-0.923	0.238	-0.959	0.274	0.223	0.117	0.928

表3 提案モデルの適合度指標

GFI	AGFI	CFI	RMSEA	SRMR
1.000	1.000	1.000	0.000	0.000

Residual) は実際のデータとモデルによる予測値とのずれの小ささに着目した評価指数で、0.05 より小であると良適合と判断される。この値が 0.000 となっているのも、自由度が 0 であるためである。

5. 分析結果の解釈

分析の結果得られたパス係数をもとに、図4の提案モデルのパス図を解釈する。室内環境 ($f1$) は照度 (il) および湿度 (hm) で定義されているが、その大きさは-0.923 と 0.238 であり、室内環境は照度に大きく左右されることが分かる。バイタルサイン ($f2$) は心拍数 ($hr1$) および呼吸数 ($rr1$) で定義されているが、その大きさは-0.595 と 0.274 であり、バイタルサインの値は心拍数に左右される。睡眠レベル (ss) は室内環境から 0.223 の影響を受けながら、バイタルサインの値からも 0.117 の影響を受けていることから、睡眠レベルはバイタルサインよりも室内環境により大きく依存していると言える。

各観測変数 il 、 hm 、 $hr1$ および $rr1$ が ss に及ぼす影響の総合効果を以下に見積もる。

$$il \rightarrow ss \text{ の総合効果} : -0.923 \times 0.223 = -0.206$$

$$hm \rightarrow ss \text{ の総合効果} : 0.238 \times 0.223 = 0.053$$

$$hr1 \rightarrow ss \text{ の総合効果} : -0.959 \times 0.117 = -0.112$$

$$rr1 \rightarrow ss \text{ の総合効果} : 0.274 \times 0.117 = 0.032$$

睡眠レベルに最も影響を与えるのは照度で、その次に大きな影響を与えるのは心拍数であり、どちらも負の値である。つまり、部屋を暗くすることが睡眠レベルを上げるためには最も効果があり、次に心拍数を小さくすることが効果的であると言える。

6. まとめ

睡眠レベルに影響を及ぼすたくさんの観測変数の中から、ステップワイズ法により観測変数を選択することで、構造方程式モデリングによる因果関係の分析を行うことができた。分析結果からは、バイタルサインよりも室内環境がより大きく睡眠に影響するとともに、深い睡眠を得るためには、部屋を暗くした上で、落ち着いてゆったりとした雰囲気にするなどにより心拍数を小さくするとよいことが分かった。

今回はパッケージ `lavaan` を用いて分析を行ったが、構造方程式の記述が容易で、R により因果分析が割合簡単に実行可能であることも確認できた。R によりデータ分析を行う際には、`lavaan` による因果分析も広く普及するとよいと考えている。

今後の課題として、今回提案したモデルでは睡眠レベル

が深くなっていく過程については明らかにしていないため、成長曲線モデルにより睡眠レベルが深くなっていく過程を明らかにしたい。

(令和元年9月19日受付)

(令和元年12月5日受理)

参考文献

- (1) 総務省:「平成29年版情報通信白書」, pp.52-62 (2018)
- (2) 厚生労働省:「平成30年版厚生労働白書—障害や病気などと向き合い、全ての人が活躍できる社会に—」, pp.369-398 (2019)
- (3) 小島隆矢, 山本将史:「Excelで学ぶ 共分散構造分析とグラフィカルモデリング」, pp.89-179, オーム社 (2013).
- (4) “The R Project for Statistical Computing”, <https://www.r-project.org/>, Retrieved Sep. 10, 2019.
- (5) 石田明男, 山本直樹, 大石信弘, 村上純:「多次元データ分解の手法を用いた立体パズルの解法(その4)」, 初等数学, 第86号, pp.20-24(2019). (継続中)
- (6) 山田剛史, 杉澤武俊, 村井純一郎:「Rによるやさしい統計学」, pp.309-319, オーム社 (2008).
- (7) 豊田秀樹:「共分散構造分析 [R 偏] —構造方程式モデリング—」, pp.18-195, 東京図書 (2014).
- (8) 豊田秀樹, 前田忠彦, 柳井晴夫:「原因をさぐる統計学 —共分散構造分析入門」, pp.99-132, 講談社 (1992).
- (9) Yves Rosseel: “The lavaan tutorial”, pp.8-15, Ghent University(2019).
- (10) 「まもるへの」, <http://mamoruno.miel.care/>, Retrieved Sep. 10, 2019.
- (11) 水野欽司:「多変量データ解析講義」, pp.61-69, 朝倉書店 (1996)
- (12) 「AICを使った変数選択」, <http://www.hnami.or.tv/d/index.php?radvance>, Retrieved Sep. 10, 2019.