

論文・解説

07 社員意識調査データの因果構造モデル化による要因分析**Data-Driven Sensitivity Analysis of Structural Equation Modeling:
A Case Study of Bayesian Network Analysis**小川 史恵^{*1} 古川 善也^{*2} 中島 健一郎^{*3}
Fumie Ogawa Yoshiya Furukawa Ken'ichiro Nakashima**要 約**

近年、マツダではそれぞれの部門で働く人々が仕事を通じて幸福を感じるといった構造の因果と活性化をうながすため Well-Being を確保できるようにしたいという目的が共有されている。従業員も風通しの良い風土づくり、職場の改善につなげるため、人事から一年に一度「グローバル社員意識調査（アンケート）」が実施されている。しかし、職場環境を改善する施策にこのデータを十分に活かせていない状態であった。そこで環境の抜本的な改善を後押しする道筋をつかもうと本研究を試みた。Well-Being をゴール指標として想定しどのような変数名や因子を構造化すれば構築できるか、について構造方程式モデリング（Structural Equation Modeling: SEM）を用いた分析を行った。この SEM の構造ではゴール指標と関係する因子として、各アンケートの項目名でもある仕事のパフォーマンス（生産性やクリエイティビティ）等がいくつかの心理的要因を介して従業員幸福感に影響することが明確化された。また本 OUTPUT は、各部門の本部長や労使懇談会などで有効なツールとして活用された。当初はデータの分析手法・結果の可視化方法を改良して個人回答の背景にある職場の現状を理解することに焦点を当てていたが、本分析の反響から①個人ではなく職場の環境に落とし込むことで改善に取り組む方の心理的な負担を軽減する、②全社共通フォーマットでお互いの現状の開示や比較がしやすくなり部門を超えたコミュニケーションが増えたことなどの現場への副次的な価値を生み出してきていることが分かり、全社の職場改善活動をより前向きな方向へ変化させることができた。

Abstract

In recent years, securing well-being in the workplace is a shared purpose at Mazda to inspire employees working in all departments. With a view to improving the workplace environment, the company conducts an annual Employee Attitude Survey. However, the survey results were not fully utilized in measures to improve the mental health environment of employees. To solve it, this study attempted to find a way to facilitate fundamental improvement in the environment, using the Structural Equation Model (SEM), with employee well-being as the target indicator. The SEM made it clear that job performance (productivity and creativity) is influenced by several factors related to well-being. This outcome was used as an effective tool by department general managers and in labor-management roundtables. In addition, although the initial focus was on understanding the current situation in the workplace behind the individual responses by improving the data analysis methods and visualization of the results, the feedback of the study outcome showed that: (1) Focusing on the workplace environment, rather than individuals, can reduce the psychological burden on those working on the improvement; (2) The company-common format has made it easier to disclose and compare each other's current status and has created secondary value for the field, such as increased communication across departments. This study has resulted in a more positive change in workplace improvement activities throughout the company.

Key words : Well-Being, SEM, Bayesian approach

*1 統合制御開発本部
Integrated Control System Development Div.

*2 福岡大学
Fukuoka University

*3 広島大学
Hiroshima University

1. はじめに

近年、人々の仕事のみならず生活を通じて幸福を感じるといった構造の因果と活性化を促すため Well-Being を確保できるようにしたいという風潮がある。そこで毎年全従業員向けに行われているアンケートで得られた結果を可視化するための手法の開発を目的とした研究を行った。毎年アンケートでデータを取得しているものの職場の細分化により問題がどこに潜んでいるかはつかみにくく、データより得られた知見を十分に活かせていないという課題が存在していた。そこで、筆者らはこうした課題を解決するための一つの方法として、構造方程式モデリング (Structural Equation Modeling: SEM) を用いたデータ分析による検討を行った⁽¹⁾⁽²⁾。本手法では、回答者の感性や意識に関するアンケート項目への回答データを観測変数として用いた。それに加えて、各分析での目的変数に応じて観測変数の背後に想定される潜在変数も分析に用いた。SEM では実際に測定を行った複数の観測変数から潜在変数を推定して分析モデルの中で用いることができる。これにより、先行研究を加味しながら収集したデータと統計モデルにより推測を利用したデータドリブンなアプローチが実現できる。更に単なるコーチングやファシリテーションに終わらず専門集団が経験を活かして働けるようにするために、このアプローチをどのように活かすことができるか、その産業応用を社内でも議論した。Fig. 1 に本手法の概要図を示す。

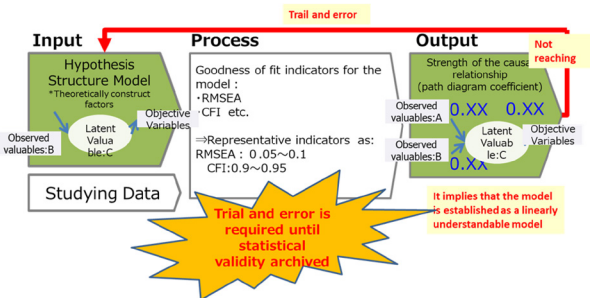


Fig. 1 The Process of Statistical Validity Archive

2. 方法

これまで行ってきたアンケート分析では、担当者の提供するアドバイスは限られており、そうした取り組みの中でデータが十分に活用されていたとはいえない。本研究のアプローチでは、この点を改善するために SEM による初期段階の分析で得られた適合度指標に基づいてモデル修正を行うとともに、そのモデルに対して社内の専門集団からのフィードバックをもらい修正の是非を議論した。この議論を効率的に進めながらアンケート項目間の因果関係を探るために結果の出力の自動化及び SEM モデルを効率的に構築する方法を検討した。

2.1 アンケートデータの取得とモデル化

まず、調査対象者に対して調査会社経由で日々の生活や仕事のなかでどんなふうを感じるかについてさまざまな角度からのアンケート項目 (40 項目前後) を、5 段階評価で回答を求めた。最終的な分析対象のアンケート項目の選定は 2 種類の因果推定法を基に行っている。1 つ目の方法は Linear Non-Gaussian Acyclic Model: LiNGAM であり、2 つ目はマツダで開発したベイジアンネットワークを基軸にする Graph Structured Analysis: GSA による手法で試行した。その後 SEM によるモデリングを行っている。

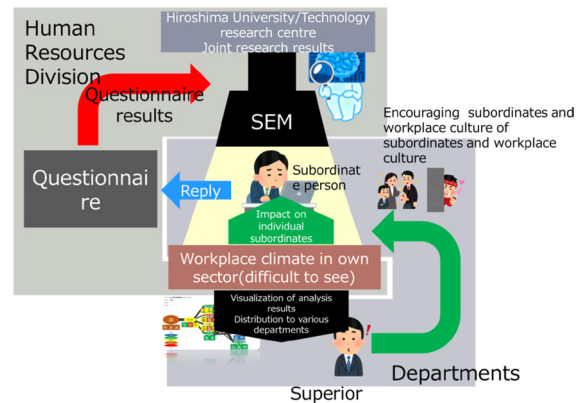


Fig. 2 One Scene of Research Data

2.2 SEM による因果的プロセスの探索

次に SEM の構造について述べる。SEM は直接観測できない潜在変数を導入し、潜在変数と観測変数との間の変数関係を推定する統計的なアプローチである。本検討では、「幸福度」を導くという仮説を立てており、仮説モデル: SEM (パス図) 及び観測変数: アンケートからの取得データ, 潜在変数: アンケート各項目から「幸福度」に至る変数名を仮説担当者が任意に決めることができる。それぞれ変数間の関連の程度を示した分散共分散行列の計算を内部で行っている。ここで測定モデルとしては各指標の因子に対する負荷量という表現を用いることが一般的である。構造モデルのパス係数と区別するために λ を使用する。各変数のスケールを標準化した場合 (各変数の平均を 0, 標準偏差を 1 に変換する操作を標準化という), 負荷量が高いほどその指標が対応する因子をよく反映していることを意味する。Fig. 2 のモデルは構造方程式として表せるが汎用性を高めるため GUI ソフトウェア (IBM AMOS Ver28, 29) を使用して検討している。R 言語などでも同様のアプローチが可能である。また内部のアルゴリズムは基本的な形は SEM と同じでパス係数の負荷量と原因因子の因子の積に誤差項 $e_{1,2,3,4}$ が加わった形になっている (下記の例の場合)。測定モデルでは、通常は各因子につき複数の指標が仮定されるため方程式の数もその数だけ設定される⁽³⁾。

式(1)~(4)においては例えば変数 F_1 : 所属する誇り,

が変化したときそれぞれの観測変数 ($X_1 \sim X_4$) の値も $\lambda_{11} \sim \lambda_{41}$ の分だけ変化することが表される。つまり、各観測変数にそのような形で影響を与える潜在変数 F_1 を仮定するという意味になっている。構造モデルと同様にパス係数（負荷量）に加えて外生変数（因子としている）の分散・共分散と内生変数（観測変数としている）の誤差分散・誤差共分散もパラメータとして推定される。ここでは共分散が仮定されていないので、分散のみがパラメータとなっている。構造モデルと同様に外生変数である因子については因子そのものの分散や内生変数である各指標について誤差変動の分散がパラメータとなる。これらのパラメータを用いて観測変数の SEM 分析が内部ではなされている。

$$X_1 = \lambda_{11}F_1 + e_1 \quad (1)$$

$$X_2 = \lambda_{21}F_1 + e_2 \quad (2)$$

$$X_3 = \lambda_{31}F_1 + e_3 \quad (3)$$

$$X_4 = \lambda_{41}F_1 + e_4 \quad (4)$$

平行して本研究では、SEM で得られた結果の構造モデル（パス図）を読み手に対してどのように呈示するかをデザインすることも重要視した。収集したアンケートの質問項目は段階評価であり、各項目への回答分布にばらつきが生じる。回答分布には濃淡があるため、尺度に基づく意識調査のアンケートごとの濃淡について平均値や分散分析を用いて、データの散らばりから閾値を決めた。ここでは SEM 図から得られる係数とかけ合わせて 5 段階（High~Low まで）を求めた。最終的に各部署に配布される時には個人の回答を基にした可視化支援システムとして活用できるうえ、本システムは SEM に情報を入れると個人情報データは一切なくなる点が強みとなっている。更に SEM モデルを作成する際の自動化を目指して前処理としていくつかの方法をトライした。そのような前処理の方法として採用しているのが LiNGAM と GSA であり、それぞれの特徴を Table 1 に整理した。両者の詳細な計算理論については参考文献リストを紹介するととどめ、本報告では割愛する。

Table 1 Comparison for GSA and LiNGAM

	GSA	LiNGAM
Theory	Bayesian network	Linear non Gaussian acyclic model; LiNGAM
Assumption	No need to assume probability distributions and no making hypothesis about structures	Need to assume probability and no making hypothesis about structures(maybe)
Linear or non-linear	Non-Linear	Linear
Shape	Working with discrete variables	Working with continuous variables
Types of algorithms	Based by BDeu-maximization (Taboo-search)	Based by Lasso

3. 結果

前章で述べた方法に基づいた結果について下記に述べる。前処理の工程では、アンケートデータからからベイジアンネットワークによるグループ化が完了したあとに SEM モデル化を実施した。SEM のモデルの評価には CFI (Comparative Fit Index) や RMSEA (Root Mean Square Error of Approximation) などの適合度指標を用いた。SEM では母数の推定後にモデルの評価を行う。モデルの評価は全体的評価と部分的評価に大きく分かれるが、まず全体的評価としてモデルがデータに適しているかを確認し、部分的評価では個々の母数の推定値が適切かを検討する⁽³⁾。全体的評価にはモデルとデータとの当てはまりの度合いを示す適合度（指標）を利用した。CFI: 0~1 までの値をとり、1 に近いほどいいモデルと解釈できる。一般的に 0.8~0.9 以上が良適合の判断基準として用いられる。RMSEA: 一般的に 0.05 以下が良適合の判断基準として用いられる。部分的評価としては標準誤差・母数の検定などは分析結果を見ながらモデルの検討に活かしていった。

3.1 Test Case1 Modelling

標準誤差・母数の検定・信頼区間の検定を行い、仮説としてどんな分散状態のデータでも SEM 形状に依存しないと仮定した。テストケース 1 では、都道府県を問わず、サンプルの N 数である 4295 の利用者アンケートの結果から 4 グループに分けた質問を求めている。ここでの前処理は因果推定として LiNGAM⁽⁴⁾⁽⁵⁾ を使用した。

質問: 27 項目, 5 件法による回答とした。

【質問内容の例】

- How do you think neighborhood feeling of you? (e.g., Satisfaction of the life)
- Consideration to others? (e.g., Circumstances)

Fig. 3 に示したように SEM モデルのデザイン（下記のようなパス図を描くこと）及び解析を行った。配置する四角形の変数は観測変数、楕円形の変数は潜在変数を意味している（因子とも呼ぶ）は直接測定されず、複数の観測変数を介して測定されている (Fig. 3, 4, 6)。本モデルの適合度は CFI: 0.590, RMSEA: 0.109 となり十分な水準には達していなかった⁽⁶⁾。

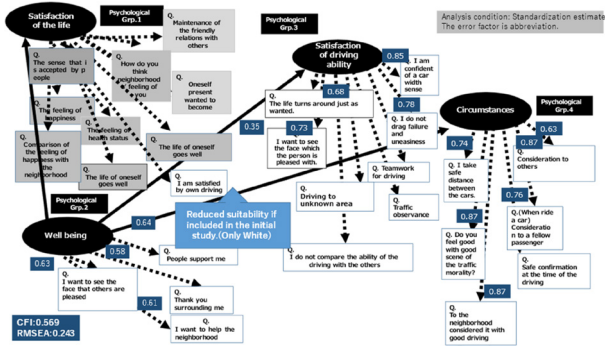


Fig. 3 Result of Test Case1, Modeling of LiNGAM to SEM

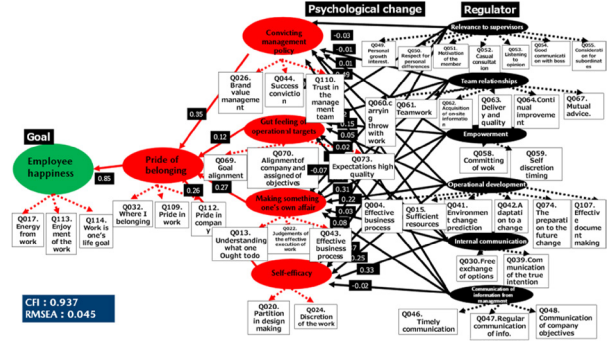


Fig. 4 Result of Test Case2, the Approach of GSA and SEM

3.2 Test Case2 Modelling

テストケース2では、データのプライバシー保護の理由からダミーデータを用いている⁽⁷⁾。社内活用の際には社員調査（任意）により収集したデータについて、個人が特定できない状態で行っている（調査会社で個人のラベルが付かない状態で加工）。サンプルのN数は10736であり、項目ごとのグループに分けた質問を実施している。前処理はGSA⁽⁸⁾⁽⁹⁾を使用し、SEMモデルのデザイン解析を行った。

質問：43項目、5件法による回答とした。

【質問内容の例】

- Trust in the management team (e.g., Convicting management policy)
- Partition in design making (e.g., Self-efficacy)

3.1 同様に Test Case2 の因果推定を GSA の検定方法で確認したところ Fig. 4 のように前段項目は4つ、後段は6つの潜在変数が測定した項目からは推定されるといった結果を因子分析から得た。GSAを前処理としSEMモデルのデザイン解析を行った。本モデルの適合度はCFI: 0.937, RMSEA: 0.045となり、初期のモデル適合度よりも最終的な活用版は工夫を重ねて高い値を確保することができた。

本章ではGSAに基づく変数間の依存関係の推論と活用のアプローチを検討し、モデルの構築後に各母集団におけるモデルの適合度を得るSEMモデルを作成することができた。CFIやRMSEAの適合度が最良のモデルを探すことにしたため、自動化でうまくいかない場合は何種類ものパス図を作成する過程も大事にした。

4. 社内活用への応用

本手法の社内活用として、Test Case2の因果推定からSEMを用いた分析を行った。SEMモデルの解析によりパス係数が自動的に求まる（詳細は2章に記述）。Fig. 5の工程の後、SEMで求まるパス係数を用いて5段階（High~Low）の評価を求め、最終的に各部署に配布される際には自動的に色別の強弱を示す分析シートを提供した。その結果、各部門やチームの強み（人の感情）を色分けし、この分析を基に各部門の本部長から「数値データだけでなく役立つ資料になった」という声が多く寄せられた。Fig. 6に活用シートの例を示す。本構造における潜在変数に向かうベクトルの強さや直接観測が可能なものはパス係数で強弱を求めることができるようになっている。

現代において製造業には多くの世代、職種の方が関わりキャリアも考え方も多様である⁽¹⁰⁾。より一人一人が仕事に生きがいを感じられるように、多くの会社の方から職場の風土改善が重要だという意見が多い。この取り組みの成果が上がり、社内では対話を重ねる活動に軸足を移し始めた。今後は対話を通じて風土を変えようという流れを感じられる。

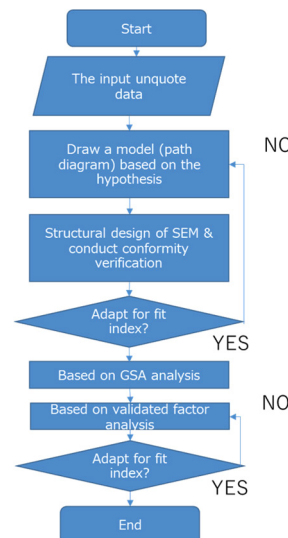


Fig. 5 The Flowchart of Approach for Causal SEM System

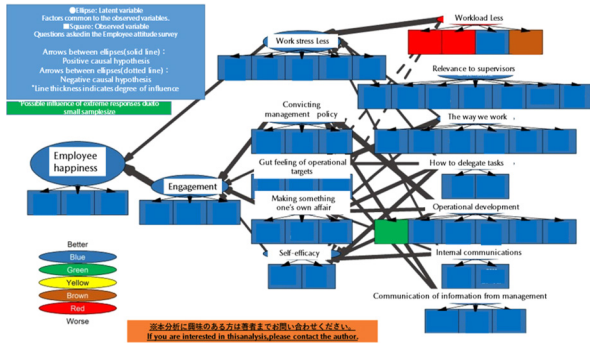


Fig. 6 Well-Being Sheet of Mazda

5. 考察

今回デザインしたシステムでは「従業員幸福感」を統計でいう目的変数に置いている。この観測変数の名称は仮説担当者が任意に決めることができる。SEMでは、変数間のパスや共分散の組み合わせが何とおりも発生するため、先行研究の知見等を踏まえ、その組み合わせの候補を絞る必要がある。今回採用したベイジアンネットワークによるアプローチでもこの組み合わせの候補は複数提案されることが分かり、モデル構築がデータ依存であることが改めて確認された。Test Case1のデータは中心部の水準に偏る分布が多く、ガウス分布に近い変数が多くなっていた。このような関係性を見落としてしまうとデータ集団の計上手段を間違えてしまう可能性もあり、散布図の確認も必要だと考えた。一方でTest Case2のデータ群は分散が大きく、サンプル数が少ない場合でも良好な適合度のCFIやRMSEAの値が得られた。貢献できた点や残る技術的課題を下記にまとめる。

- SEMの仮説構造を効率的に更新して適合できる設計の糸口がつかめた。
- 潜在変数を間接的に推定する（測定された観測変数から計算する）というSEMによるモデル構築のために、ベイジアンネットワークや因子分析を前処理として利用する強みが示唆された。
- 社員意識調査のデータに対してSEMを行うことで、変数間の関係性を示すモデルを可視化し社内でのデータ活用につなげた。
- 他のケースで適合度を確保できるかは分からないため、別のサンプルで同じモデルが適用できるかどうか確認することが今後の課題である。

6. おわりに

本共創の取り組みを4年間社内でも継続してきた。この分析を用いることで、仕事のパフォーマンス（生産性や創造性）に幸福感のいくつかの因子が影響することを明確にできた。また、この成果は各部門の本部長や労使懇談会で有効なツールとして活用された。当初はデータの

分析手法と結果の可視化方法を改良し、個人回答の背景にある職場の現状を理解することに焦点を当てていたが、社内での反響から以下の副次的な価値が生まれている。全社共通フォーマットにより、現状の開示・比較が容易になり、部門間のコミュニケーションが増加した。これにより、全社の職場改善活動がより前向きに変化している。個人ではなく職場環境に焦点を当てることで、改善に取り組む人の心理的負担が軽減された。社員意識調査の分析を従業員に共有して本活動を活用することで、相互理解が進むことを願っている。この取り組みは社内だけでなく学会でも大きな反響を得た。今後も会社の枠を超えた共創活動や部門を超えた意見交換を継続できるような取り組みを進めていきたい。

これまで広島大学・福岡大学とマツダの産学連携研究に関わってくださった関係のみならず全員にお礼申し上げます。特にプロジェクトリーダーであった吉永冬彦氏にはマツダ在職時に本研究を牽引していただいた。更に2024年のSICE ANNUAL CONFERENCE 2024や日本グループ・ダイナミクス学会第70回大会においても皆様に現場で活用したいという意見を多く頂戴し、会場でも多くの興味深い意見をいただいた。感謝申し上げます。

参考文献

- K. A. Hallgren, C. J. McCabe, K. M. King, D. C. Atkins, Beyond Path Diagrams: Enhancing Applied Structural Equation Modeling Research through Data Visualization, *Addictive behaviors*, Vol.94, pp.74-82, July, (2019)
- D. A. Fife, S. M. Brunwasser, E. C. Merkle, Seeing the impossible: Visualizing latent variable models with flexplavaan. *Psychological Approachs*, 28(6), pp.1456-1477, (2023)
- H. Toyota, *Structural Equation Modeling AMOS Edition*, Tokyo-Tosho, (2014)
- S. Shimizu, P. O Hoyer, et al., A Linear Non Gaussian Acyclic Model for causal Discovery, *Journal of Machine Learning research* 7, pp.2003-2030, (2006)
- J. Xiang, et al., A Lasso for Learning a Sparse Bayesian Network Structure for Continuous Variables, *Advances in Neural Information Processing Systems* 26, NIPS (2013)
- 小川史恵・古川善成・吉永冬彦・中島健一郎, “データドリブンに基づく感性評価についての探索的研究” 日本グループ・ダイナミクス学会第70回大会発表論文集, pp.124-125, (2024)
- F. Ogawa, Y. Furukawa, F. Yoshinaga, K. Tomii, K. Nakashima, Study of sensitivity evaluation based on the data driven, 63rd Annual Conference of the

Society of Instrument and Control Engineers of Japan (SICE), pp.81-85, (2024)

- (8) M. Ueno, Bayesian Network, Corona Publishing Co., Ltd., (2013)
- (9) M. Mera, Y. Nakamura, et al., Study on Approach of Utilization of Bayesian Networks for Analyzing Time Series Data, Journal of The Japan Society for Industrial and Applied Mathematics, Vol.31, No.2, pp.76-104, (2021)
- (10) 宇田川元一：他者と働くー「わかりあえなさ」から始める組織論，ニュースピックス，(2019, 2024)

■著者■



小川 史恵



古川 善也



中島 健一郎