

統計的因果推論に生じた奇妙な現象ともやもや感

—潜在変数測定値の連鎖処理と構造モデル適応に伴う変容との乖離問題—

深 瀬 澄

要旨

近年、統計的因果推論の進歩により、過去の実験結果に重大な誤りが判明した事例も散見する。文化系大学でも統計的因果推論の重要性は認識され、数理・データサイエンス・AI教育による考察力、問題解決力、価値創造力の育成強化が期待される。観測不能な潜在変数間の統計的因果推論では、かつては因子分析により尺度得点化した変数を重回帰分析に用いる分析手法が一般的であったが、2000年前後から構造方程式モデルが普及し、一部の研究者からは従来手法の問題点が指摘された。しかし、今なお従来手法が実証分析のテキストや論文で見られ、重要な警鐘が忘れられつつある。

本研究では、学生の実験指導において、構造方程式モデリングの分析中に奇妙な現象が観察され、その検証実験においてまたしても奇妙な現象が生じたことから、「モデルの構造中で潜在変数の測定値（因子得点）は変容する」という新しい仮説を着想し、従来手法と構造方程式モデリングでは計算に使用する潜在変数のデータの値が異なる可能性があり、そのために両者の分析結果が整合しない可能性があることを検証した。この結果は、数理・データサイエンス・AI教育のみならず、データのフロー処理の根底に影響する可能性もあり、さらなる詳細な検討が必要と思われる。

キーワード：

潜在変数、統計的因果推論、確証的因子分析、尺度得点化、構造方程式モデル

I はじめに

1. 本研究の学術的背景

現代社会において「データサイエンス」という言葉が膾炙し、統計教育への期待もかつてないほど高まった。ただし、特に文科系大学における数理・データサイエンス・AI教育では、統計理論が軽視され、パッケージソフトの使用が偏重される傾向がある。例えば、オープンソースのR言語やPythonでは高度な分析ツールのライブラリ

が充実しており、複数の分析ツールを連動させたフロー処理が一般的に教えられている。だが、統計的因果推論の分析においては、かつては因子分析により尺度得点化した変数を重回帰分析に用いる分析手法が一般的であったが、2000年前後から構造方程式モデルが普及し、狩野（2002）、小杉（2018）など一部の研究者からは従来手法の問題点が指摘された。しかしながら、今なお従来の手法が実証分析のテキストや論文に散見し、先人による重要な警鐘が教育現場に届かず忘れられつつある。

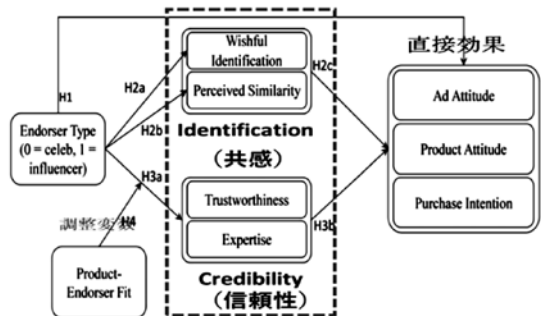
担当教員の専門分野によっては、このような分析の是非について理論的な検討がされず用いられてきた可能性がある。その結果、誤用が疑われる教材や事例も散見され、致命的な誤りが生じる可能性がある。このような状況が続けば、データサイエンスの実践において誤りが拡散し、政策決定やビジネス戦略に重大な影響を及ぼし、社会全体のデータ活用能力に悪影響を及ぼす恐れがある。

2. 研究目的

本研究は、文系学生のデータサイエンス教育において、複数の統計的手法を連続的に組み合わせて処理することにより生じる問題に着目し、徹底的に検討することによって、特に統計的因果推論の誤用を防止することを目的としている。現代の大学教育において強く求められるのは、知識の修得よりも問題解決力や価値創造力の養成であり、統計的因果推論の手法はそれを支援する有力な分析ツールとなり得る。しかし、

図表 I-1 *International Journal of Advertising* 掲載論文における分析モデル

一部の研究者は、確証的因子分析を用いて潜在変数を尺度得点化し、因果関係の分析する分析手法を問題視したが、図のように、査読付きのジャーナル誌でさえも、下位尺度得点を測定値としてパス解析に利用する従来の分析が今なお行われている。



出所：Alexander P. Schouten, Loes Janssen & Maegan Verspaget (2020), “Celebrity vs. Influencer endorsements in advertising: the role of identification, credibility, and Product-Endorser fit”

この分析ツールは諸刃の剣であり、不適切に使用され誤った分析結果が導かれれば、逆に悪影響を与える危険性もある。誤用について詳細に検討することは極めて重要である。特に、統計理論に詳しくない文科系学生に対して、統計的因果推論の誤用を防ぐための「安全点検」的アプローチを導入することで、より正確で信頼性の高い教育を提供する。本研究は、交通や医療など他の分野での安全性確保と同様に重要であり、データサイエンス教育の基盤を強化するものである。

3. 着想に至った経緯

着想の契機となったのは、当時修士2年であった庄豊源氏の構造方程式モデリングを用いた実証研究の指導中に観察された奇妙な現象である。彼女の研究テーマは「製品普及におけるライブコマースを利用したキャスターの起用」であり、得られた知見は学術的および実務的な貢献が非常に高い。実際、彼女の修士論文は2023年度修了生の学長賞に選ばれるなど、秀作と評価されている。しかし、分析モデルには直接効果が想定されておらず、媒介効果の分析結果から結論が導かれていた。海外のジャーナル誌にもこのような論文は多数見られ問題はない。筆者の興味から直接効果についても分析を試みるよう指示し、自らもモデルを修正して試行したところ、奇妙な現象が発生した。この分析結果に確信が持てず、検証を試みたところ、さらに奇妙で、伝統的な因果推論のみならずプログラミングにおけるデータ送り処理等の根底を揺るがしかねない現象が生じた。庄豊源氏の修士論文の提出期限が迫り、分析結果の誤りも懸念されるため、指示した分析を論文審査の対象外とし、補論として扱うこととした。

本研究はこのような背景から着想されたものである。次章に本研究着想の契機となった庄豊源論文の概要および経緯を示す。

II. 庄豊源論文の概要

1. 庄豊源 (2024) の研究目的と含意

庄豊源 (2024) は、「製品普及におけるライブコマースを利用したキャスターの起用」をテーマとするマーケティング分野の実証研究であり、経済学部出身の彼女により、経営学専攻の学位請求論文として提出された修士論文である。

ライブコマースとは、商品推奨者（キャスター）がSNSを駆使して、映像、音声により消費者に商品情報を双方向で伝えながら、同時に消費者間でも弹幕と呼ばれる文字情報を飛び交わして商品知識を交換できる多次元の情報を駆使した販売形式である。

日本市場では未だ普及していないが、商品に対する信頼感が弱い中国市場では、情報の非対称性を解消しやすいライブコマースによる流通取引が急速に拡大している。2025年の取引総額規模は42兆円に達すると予測されている¹。1回のライブ配信で数億～数十億円を売り上げる達人キャスター（商品推奨者）も存在するが、ライブコマースの売上額は、起用するキャスターのタイプによって、大きく左右される。

庄豊源（2024）は、ライブコマースによる販売形式においてキャスターの重要性に着目し、販売対象商品、ターゲットとする購買層との相性から、起用すべき商品推奨者の最適な選択を研究目的とした。

ライブコマースを、商品ブランド、消費者、キャスターの三者が集う「プラットフォーム」として捉え、販売対象とする商品カテゴリと、ターゲットとする消費者タイプの組合せに対して、効果的なキャスターを選択することを研究目的とした。

図表Ⅱ-1 中国で著名な達人キャスター、ライブ配信での売上実績



出所：左：NHK クローズアップ現代、<https://www.nhk.or.jp/gendai/articles/4529/>

右：新周刊 -APP「李佳琦106億、薇娅83億、有你份吗？/ 日本肯德基，薯条没了」、<https://www.newweekly.com.cn/article/shp0728301308>（最終閲覧日2024年1月）

¹ 「爆発的成長を遂げる中国ライブコマース市場」、<https://reinforz.co.jp/bizmedia/16952/>

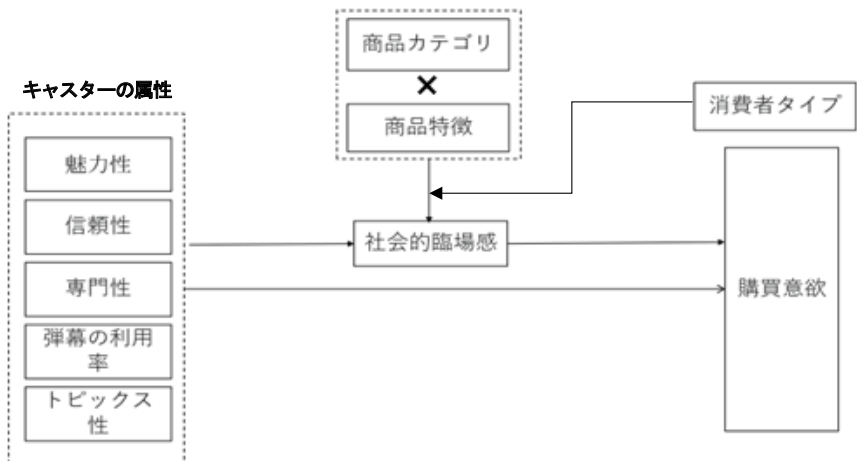
庄豊源 (2024) は、図表Ⅱ－2 のような概念フレームワークを想定し、以下に掲げる学術的問いに答えることにより、キャスターが消費者に与える影響が明らかになるだけでなく、様々なタイプのキャスターが独自の能力を開発する一助となり、ブランドに合致するキャスター選択の実務にも資するとしている。

- i) キャスターは、自己の個性をどのように生かして能力を開発すべきか。
- ii) 商品ブランドに、どのような属性をもつ達人キャスターが合致するのか。
- iii) 消費者は、社会臨場感を高めるためにどのような属性のキャスターのライブに参加すべきか。

2. 先行研究および仮説設定

ライブコマースと従来の e コマースとの大きな違いは、キャスターが商品を推奨することと、他の視聴者の参加と反応が感じられることである。庄豊源 (2024) では、SNS を駆使してキャスターと参加した消費者が一体となることによって「社会臨場感」が醸成されると仮定し、これをライブコマースにおける重要な媒介変数として用いた。

図表Ⅱ-2 庄豊源 (2024) の概念フレームワーク



出所：庄豊源 (2024)、第Ⅴ章掲載図に筆者が一部修正を加えた。

「社会臨場感 (social presence)」とは、Short, Williams & Christie (1976) の著書『情報通信の社会心理学』に登場した概念であり、「オンラインコミュニケーションにおいて個人が他者の実在を知覚する度合い」を指す。その後、インターネット技術の普及に伴い、研究対象が、遠隔教育、マーケティングをはじめとする広範な分野に拡大している。

Short ら (2007) は、社会臨場感に影響を与える重要な 2 つの次元として、即時性と親密性の概念を関連付けた。即時性とは、個人と交流対象者の心理的距離と直接感を指し、反応時間が短縮すると他者との接触感を高めるとともに仲介感を低下させる。また、親密性は、他者への気遣い、信頼、自己の表現、他者との関係構築の程度として捉えられ、伝統的なコミュニケーション環境の下では、トピックス、アイコンタクト、微笑みなどを通じて実現される。また、季丹 (2016) は、社会臨場感が高いほど、オンラインでの人的なコミュニケーションにおいて相互親密性が強く知覚され、相手に対する信頼感が高まると述べている。

庄豊源 (2024) は、事前調査として実施された庄豊源 (2023) ² を基に、以下の仮説を設定した。まず、「H1 : ライブコマースにおいて、キャスターから発信される情報の特性 (信頼性、専門性、魅力性) は、個人と交流対象者の間の親密さを深める」。次に、社会臨場感に関する知見に基づき、「H2 : キャスターから発信される情報の特性により、視聴者との親密感が強まると、視聴者が知覚する仲介感が弱まり、強い社会臨場感が醸成される」、さらに、これにより「H3 : キャスターおよびキャスターが推奨する商品に対する信頼が高まり、消費者の購買意欲が促進される」という仮説が付加され、研究室の先輩学生の研究を継承し、倪テッケツ (2020) が修士論文において言及したライブコマースにおける弾幕の効果も考慮されている。

3. 研究方法

(1) 分析モデルおよび分析方法

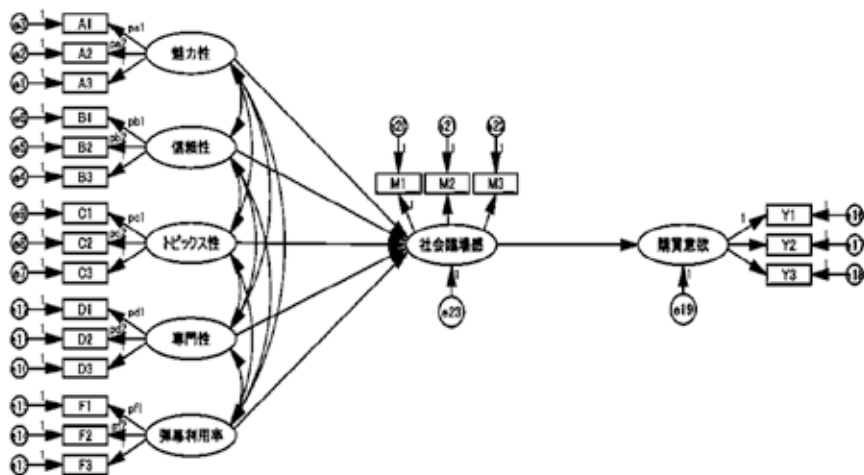
庄豊源 (2024) では、ライブコマースにおけるキャスター、商品、消費者の 3 者間における因果関係を、図表Ⅱ-3 に示す「社会臨場感」を媒介変数とする PLS 型構造

² 日本ホスピタリティ学会第30回全国大会、口頭報告予稿集 (2022年8月、熊本学園大学)

方程式モデル (Partial Least Squares Model) を用いて媒介効果を分析した。分析モデルにおいて、丸枠で示したキャストの属性に関する「魅力性」、「信頼性」、「トピックス」、「専門性」、「弹幕利用」と、「社会臨場感」、「購買意欲」はいずれも潜在変数であり、A群～Y群は観測変数である。A群～Y群は潜在変数を測定するための5件法によるリッカート尺度となっていて、それぞれの質問項目は図表Ⅱ-4に示される。

前5つの潜在変数に関して、消費者はキャストが発信するこれらの情報の特性から、キャストの特徴を判断していると仮定している。しかし、これらの知覚は購買意欲に直接影響しない。「魅力性」と「信頼性」は応答性、「専門性」と「弹幕利用」は即時性の軸方向に知覚を促し、「トピックス」による話題性も加わり、媒介変数である「社会臨場感」が醸成される。これにより躊躇が緩み、購買意欲が促進されるのである。

図表Ⅱ-3 分析モデル1 (直接効果がない構造方程式モデル)



出所：庄豊源 (2024)。Amos により作成。注：各潜在変数から3つの観測変数へのパスのうち3番目を固定。

³ Rogers (1962) における消費者分類：①イノベーター (Innovators：革新者)、②アーリーアダプター (Early Adopters：初期採用層)、③アーリーマジョリティ (Early Majority：前期追従層)、④レイトマジョリティ (Late Majority：後期追従層)、⑤ラガード (Laggards：遅滞層)

統計的因果推論に生じた奇妙な現象ともややや感（深瀬）

図表Ⅱ-4 リッカート尺度（アンケート調査票の質問項目）

潜在変数	変数	質問項目
魅力性	A1	キャスターはとても美しい・かっこいいか
	A2	キャスターの口コミよりも、価格よりも、イメージを優先するか
	A3	キャスターの性格的な魅力が好きになったり、夢中になったりすることができるか
信頼性	B1	キャスターは推奨している製品に対して、豊富な使用経験があると思うか
	B2	キャスターのおすすめ商品を信じて、あなたに役立つと信じているか
	B3	ライブの製品価格は合理的だと思うか
トビックス性	C1	あなたがよく見るキャスターは知名度が高いと思うか
	C2	キャスターは主導的で影響力のある地位を持っていると思うか
	C3	キャスターのトビックス性（ファン数）が高いので商品を買うのか
専門性	D1	キャスターは商品について詳しく、専門的に解説できるか
	D2	キャスターはあなたの要求に応じてサービスを提供してくれるか。 （例えばキャスターが質問に答えるときは、視聴者の要求に応じて製品の詳細を説明するなど）
	D3	商品選択に関して、キャスターは具体的なアドバイスができるか
弹幕の利用率	F1	ライブコマースを見るとき、キャスターが弹幕を出すように誘導するか
	F2	ライブコマースを見る際に、弹幕を利用するか
	F3	ライブコマースを見る際に、弹幕を通じてコメントした商品を購入したいか
社会臨場感	M1	キャスターと対面して交流しているような感覚になるか
	M2	あなたは楽しく満足して、全身全霊を傾けられるか
	M3	キャスターの行為によって、幾分かあなたの気持ちが影響されるか
購買意欲	Y1	好きなキャスターを他の人に推薦したいか
	Y2	ライブコマースで新製品を見たとき、キャスターが推薦する製品ならば購入したいか
	Y3	購入するつもりはなかった製品でも、キャスターのライブコマースを見て購入したくなるか
商品タイプ		ライブコマースにおいて、あなたは一般的にどの商品タイプに注目しているか
消費者タイプ		新しい製品が出たら、どのタイミングで購入して使用しているか （前期消費者・中期消費者・後期消費者・末期消費者）
推薦度		最近購入した製品を他の人に推薦したいか
満足度		最近購入した商品に満足しているか

出所：庄豊源（2024）。

注1：上記の質問項目以外に回答者のプロフィールについても尋ねている。

注2：実際には中国で質問したが、邦訳したものである。

回答結果に消費者のタイプが影響する可能性を軽減するために、回答者をタイプ別に階層化して分析することで無作為化を図った。米国流の研究では Rogers (1962) が提唱したイノベーター理論に基づき、商品購入態度が早い順に5分類³される。しかし、中国市場を対象とする庄豊源 (2024) では、周永沢 (2020) の製品イメージ形成研究における製品消費への介入時期に基づき、消費者タイプを前期消費者・中期消費者・後期消費者・末端保守消費者の4タイプに分類した。キャスターの特性による訴求力について、ターゲットとするこれらの消費者タイプによる違いを多母集団同時分析を用いて比較した。

次に、販売対象とする商品ブランドについては、商品特徴 (高い推薦度／高い満足度) と商品タイプ (日用消費財／耐久消費財) を組み合わせて、図表Ⅱ-5 に示す4ブランド (Ⅰ・Ⅱ・Ⅲ・Ⅳ) に分類した。販売対象とするそれぞれの商品ブランドについて、消費者タイプ別の多母集団同時分析を4回繰り返して分析を行った。

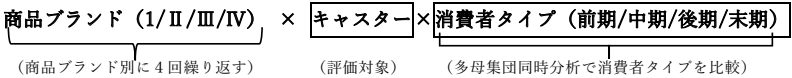
上述したキャスターの特性、商品ブランド (カテゴリ×特徴、Ⅰ・Ⅱ・Ⅲ・Ⅳ)、消費者タイプ (前期消費者・中期消費者・後期消費者・末端保守消費者) の3要因を組み合わせた因果関係の分析方法を要約すると、図表Ⅱ-6 のように図式化できる。

図表Ⅱ-5 商品特徴と商品カテゴリによるブランドの組み合わせ

商品特徴	商品カテゴリ	
	日常消費財	耐久消費財
高い推薦度 (お値打)	商品ブランドⅠ	商品ブランドⅡ
高い満足度 (高品質)	商品ブランドⅢ	商品ブランドⅣ

出所：庄豊源 (2024)。

図表Ⅱ-6 商品ブランド、キャスター、消費者タイプを組み合わせた分析の枠組み



(3) アンケート調査概要と分析に用いたデータ

図表Ⅱ-3 に示した分析モデルに使用するデータを収集するため、2023年9月から11月まで中国で Web によるアンケート調査を実施し、図表Ⅱ-4 (前掲) に示した項目について5件法で尋ねた。3188件のアンケート回答が集まり、回答者の属性は図表

図表Ⅱ-7 アンケート回答者の属性

変数	項目	回答者数	構成比
性別	男	1,574	48.5
	女	1,641	51.5
年齢	18歳未満	191	6.0
	18-25歳	753	23.6
	26-35歳	1,004	31.5
	36歳-50歳	800	25.1
	51歳及び以上	440	13.8
	学生	322	10.1
職業	国営企業職員 / 公務員	346	10.9
	民間企業職員	987	31.0
	自由職業者	539	16.9
	専業主婦	620	19.4
	その他	374	11.7
	20,000円以下（日本円換算した）	300	9.4
月収	20,001-60,000円（日本円換算した）	597	18.7
	60,001-100,000円（日本円換算した）	1,015	31.8
	100,001-200,000円（日本円換算した）	644	20.2
	200,001円以上（日本円換算した）	632	19.8
	中学卒業以下	359	11.3
	高校卒業	652	20.5
学歴	専門学校卒業	941	29.5
	大学卒業	751	23.6
	大学院生修了以上	485	15.2
商品のタイプ	日用消費財	1,599	50.2
	耐久消費財	1,589	49.8
商品の特徴	高い推薦度	1,601	50.2
	高い満足度	1,587	49.8
消費者のタイプ	前期消費者	796	25.0
	中期消費者	781	24.5
	末期消費者	808	25.3
	末端保守消費者	803	25.2

出所：庄豊源（2024）

筆者注：多母集団同時分析により、各消費者タイプの態度を比較するにあたり、より厳密には、ランダム化の成立を確認し、性別、年齢、年収等による交絡要因の影響がないことを検討しておくべきだった。

Ⅱ-7のとおりである。得られたデータについて、構造方程式モデル分析の前提となる正規性（歪度と尖度）、変数間の相関性（KMO 値）、分散均一性（パートレットの球面性検定）、尺度の信頼性（クロンバック α 係数）尺度の妥当性（項目テスト相関）などを検討し、結果は全て良好であった。

4. 庄豊源（2023）の分析結果

（1）全体モデルの分析結果

まず、全体モデルについて最尤法による推定を行った。分析結果は図表Ⅱ-8に示されるように、適合度の指標は全て1に近い値を示し、誤差も3%以下で良好な結

果となった。また、因果関係の強さを示すパスについても、Wald 検定の結果、p 値が全て0.1% 未満で有意となり、削除すべきパスはなかった。標準化パス係数の値より、「社会臨場感」の醸成に影響するキャスターの情報発信の特性を比較すると、「トビックス性」(0.312) の影響が他の特性よりも幾分強いことが判明した。さらに、醸成された「社会臨場感」(0.788) が消費者の購買意欲に強く影響することも確認された。

図表 II-8 モデル適合度 (左表) と主効果のパラメータ推定結果 (右表)

参考指標	統計値	P 値	パス係数		Estimate	S.E.	C.R.	P
Chi-square/df	3.653 ***		魅力性	→	社会臨場感	0.213	0.015	13.916 ***
GFI	0.983 ***		信頼性	→	社会臨場感	0.223	0.016	13.736 ***
AGFI	0.975 ***		専門性	→	社会臨場感	0.252	0.018	13.669 ***
CFI	0.989 ***		弹幕の利用率	→	社会臨場感	0.259	0.02	13.223 ***
NFI	0.985 ***		トビックス性	→	社会臨場感	0.312	0.019	16.696 ***
RMSEA	0.029 ***		社会臨場感	→	購買意欲	0.788	0.02	39.33 ***

出所：庄豊源 (2024)

注 1：有意水準のラベル表示について、***： $p<0.001$ 、**: $p<0.05$ 、+： $p<0.1$

注 2：各潜在変数から観測変数への因子負荷量を示す8本のパスを省略しているが、全て1% 水準で有意。

さらに、キャスターの情報発信の特性が「社会臨場感」を媒介して消費者の購買意欲を促進するという間接効果が、収集した標本に影響されず一般的に存在することを確認するため、ブートストラップ検定を行った。3188件のアンケート回答からランダムにサンプルを抽出し、上記の分析を2000回反復して間接効果が生じることを検証した。ブートストラップ検定の結果は図表 II-9 に示され、各潜在変数からの間接効果の P 値は全て 5% 未満となった。以上の結果より、分析モデルにおいて間接効果は偶然ではなく一般的に存在することが示され、仮説 H 1 ～ H 3 を支持する結果が得られた。

図表 II-9 2000回の無作為抽出によるブートストラップ検定結果

結果となる内生変数		原因となる説明変数				
		トビックス性	弹幕利用	専門性	信頼性	魅力性
媒介変数	社会臨場感	↓	↓	↓	↓	↓
目的変数	購買意欲	0.001	0.001	0.001	0.001	0.001

出所：庄豊源 (2024)、Amos 出力結果からの数値抜き出し。

（２）多母集団同時分析

次に、キャスターの情報発信の訴求が、消費者タイプ（前期・中期・後期・末端）の違いによりどのように異なるのか、多母集団同時分析によって比較した。このような分析を、販売対象とする商品ブランドⅠ～Ⅳのそれぞれについて行った。

構造方程式モデルは「測定モデル」間の因果関係について、「構造モデル」を構築してパス解析する分析手法である。「測定モデル」（因子分析の部分）のパス係数は因子負荷量に対応し、「構造モデル」（重回帰分析の部分）のパス係数は偏回帰係数に対応する。多母集団同時分析において、グループ間のパラメータ（パス係数、共分散、分散）に等値制約を課して共通化すると、推定する未知数の個数が減り自由度が高まるメリットと、誤差が大きくなるデメリットがある。

そこで、販売対象商品を「商品ブランドⅠ」とし、図表Ⅱ-10に示す6通りの等値制約を比較した。キャスターの訴求力を消費者タイプで比較するにあたり、「測定モデル」の因子負荷量（パス係数）は等値制約を付けて測定不変とする必要がある。モデル適合度は「構造モデル」のウェイト制約が最良だが、等値制約を付けてしまうと消費者タイプ別の訴求力の比較ができなくなる。モデルの自由度も考慮して、本分析においてはBCC値（2403.7）が最も低い「測定モデルのウェイト」に等値制約を設けたモデルが最も妥当である。

図表Ⅱ-10 モデル比較（配置不変性、等値制約の検討）

モデル	モデル適合					モデル比較	
	RMR	GFI	AGFI	PGFI	CFI	AIC	BCC
制約なし	0.064	0.894	0.858	0.67	0.948	2439.9	2500.87
測定モデルのウェイト	0.066	0.893	0.87	0.712	0.949	2354.5	2403.7
構造モデルのウェイト*	0.062	0.902	0.877	0.719	0.95	2354.5	2404
構造モデルの共分散	0.104	0.88	0.859	0.748	0.942	2426.6	2463.19
構造モデルの残差	0.105	0.879	0.858	0.752	0.941	2415.9	2450.82
測定モデルの残差	0.105	0.87	0.86	0.808	0.939	2420.6	2437.81

配置不変性、測定不変性等を検討し、測定モデルについてはウェイトに等値制約を課し、構造モデルについては制約を設けずに多母集団同時分析を行った。分析結果はいずれの消費者タイプについても、構造モデルのウェイトすなわちパス係数の推定値

が0.1%水準で有意であることが示された。他の商品ブランドについても同様に分析を行い、推定された標準化パス係数は図表Ⅱ-11に集約されている。

分析結果を視覚化して解釈するため、数量化理論Ⅲ類を用いて、消費者タイプと訴求しやすいキャスターの情報発信の特性とを図表Ⅱ-12に示す座標平面上に対応づけた。なお、配列を縦横の軸方向に近づけるために、一部バリマックス回転をかけた。

庄豊源 (2024) は、ライブコマースによる販売形式においてキャスターの重要性に着目し、販売対象商品、ターゲットとする購買層との相性から、起用すべき商品推奨者の最適な選択を研究目的とした。考察については省略するが、この結果に研究目的に対する成果が集約されており、研究の貢献として、消費者タイプとキャスターの情報発信特性の関係を明らかにし、実務および学術的に有益な知見を提供している。

図表Ⅱ-11 多母集団同時分析による主効果の推定結果（商品ブランド別）

商品ブランドⅠ：日常消費財×高推薦度（お値打）

	前期消費者	中期消費者	後期消費者	末期消費者
魅力性	0.136	0.161	0.275	0.141
信頼性	0.18	0.232	0.254	0.292
専門性	0.225	0.176	0.212	0.258
弾幕の利用率	0.345	0.264	0.134	0.279
トピックス性	0.385	0.318	0.283	0.19
社会臨場感	0.743	0.769	0.852	0.838

商品ブランドⅢ：日常消費財×高満足度（高品質）

	前期消費者	中期消費者	後期消費者	末期消費者
魅力性	0.361	0.166	0.146	0.125
信頼性	0.158	0.187	0.252	0.263
専門性	0.172	0.202	0.336	0.276
弾幕の利用率	0.261	0.27	0.26	0.295
トピックス性	0.301	0.345	0.17	0.194
社会臨場感	0.688	0.781	0.843	0.834

商品ブランドⅡ：耐久消費財×高推薦度（お値打）

	前期消費者	中期消費者	後期消費者	末期消費者
魅力性	0.273	0.185	0.254	0.245
信頼性	0.213	0.171	0.286	0.304
専門性	0.245	0.26	0.125	0.293
弾幕の利用率	0.164	0.182	0.245	0.144
トピックス性	0.262	0.269	0.247	0.18
社会臨場感	0.785	0.878	0.841	0.846

商品ブランドⅣ：耐久消費財×高満足度（高品質）

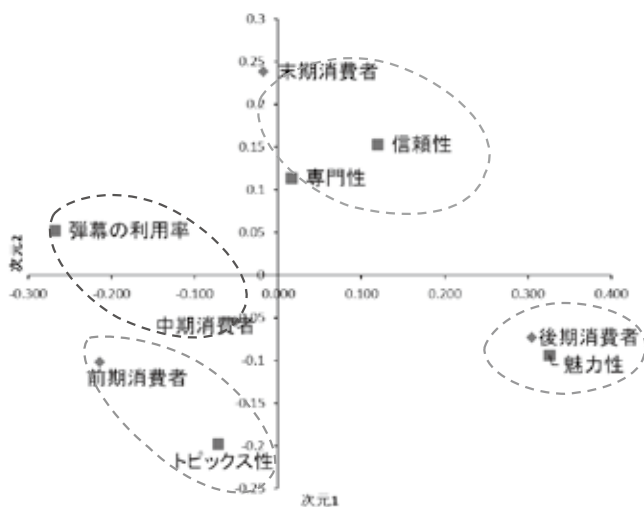
	前期消費者	中期消費者	後期消費者	末期消費者
魅力性	0.22	0.175	0.236	0.227
信頼性	0.2	0.193	0.25	0.271
専門性	0.278	0.344	0.279	0.329
弾幕の利用率	0.228	0.195	0.228	0.216
トピックス性	0.386	0.283	0.167	0.117
社会臨場感	0.718	0.844	0.822	0.82

出所：庄豊源 (2024)。ただし、一部に誤りがあり筆者が再計算を行なった。

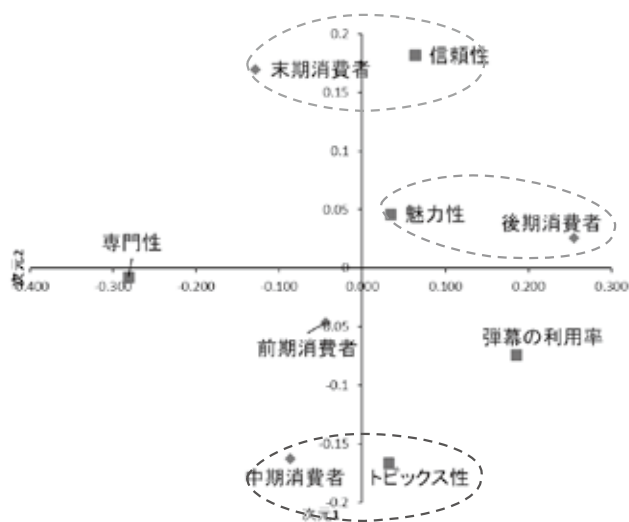
統計的因果推論に生じた奇妙な現象ともややや感（深瀬）

図表Ⅱ-12 数量化Ⅲ類によるカスターの情報発信特性と消費者タイプの対応付け

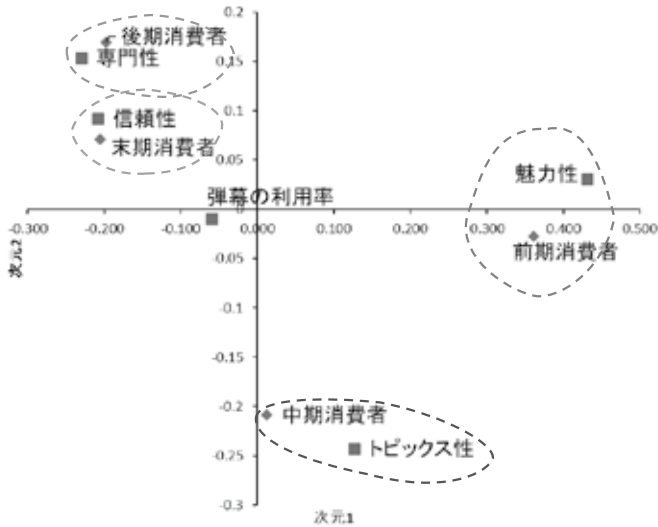
①商品ブランドⅠ：日常消費財×高満足度（お値打）



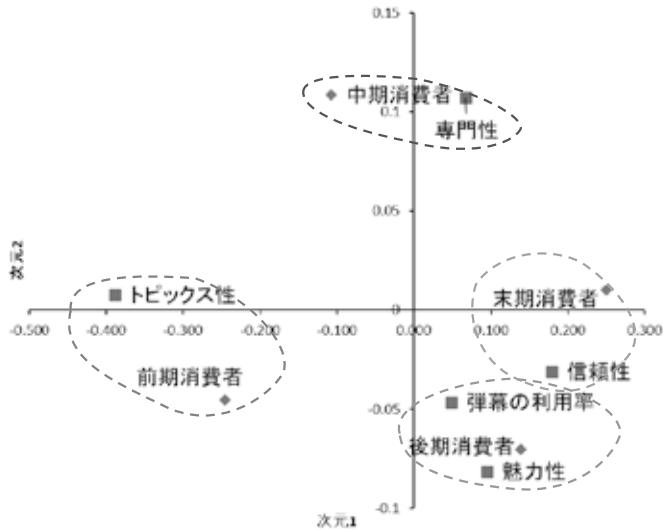
②商品ブランドⅡ：耐久消費財×高推薦度（お値打）



③商品ブランドⅢ：日常消費財×高推薦度（高品質）



④商品ブランドⅣ：耐久消費財×高推薦度（お値打）



注：見やすくするために、一部バリマックス回転をかけた。

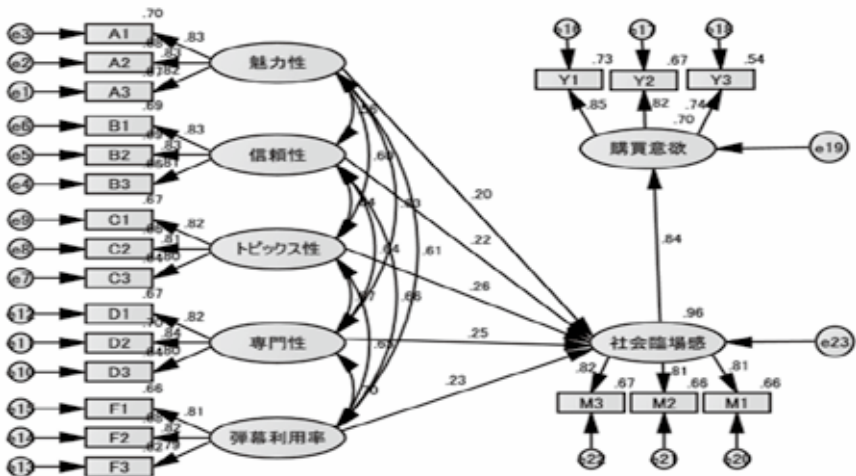
Ⅲ. 分析モデルに観察された奇妙な現象

1. 直接効果の分析において生じた奇妙な現象

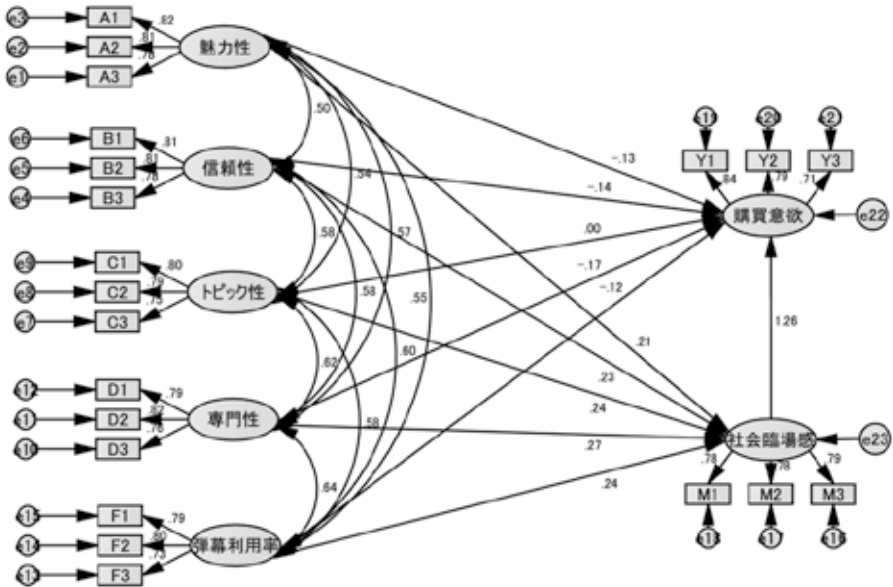
庄豊源（2024）は、図表Ⅲ-1に示す基本モデルを用いて、キャスターの情報発信の特性が「社会臨場感」を媒介し、間接効果として「購買意欲」を促進することを検証し、研究目的を達成して有益な知見を導き出した。しかし、媒介変数を経由しない直接効果の分析がされていなかったため、筆者はこの点に興味を持ち、追加調査を指導した。この追加分析に関連して、3つの奇妙な現象が観察された。

彼女の分析結果は、図表Ⅲ-2に示されるモデル2bの標準化パス係数に示されている。「トピック性」（ $p=0.058$ ）が5%水準で有意とならなかったことを除き、「購買意欲」への直接効果の標準化パス係数が全てマイナス値で有意となり、キャスターの魅力的な特性が「購買意欲」を抑制するという予想外の結果となったのである。

図表Ⅲ-1 モデル1 庄（2024）基本モデル



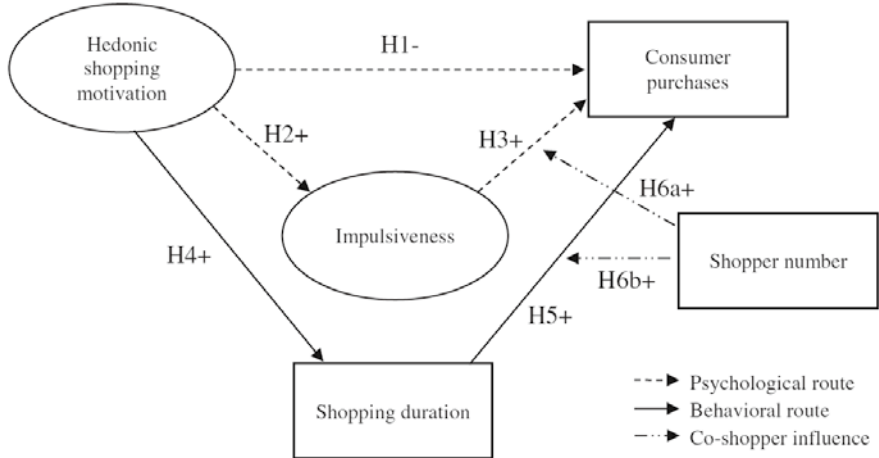
図表Ⅲ-2 モデル2b (購買意欲への直接効果と間接効果あり)



このような現象に関連する先行研究として、笠井・科野 (2017)⁴による非日常的店舗での購買行動において発生するブラウジング効果に関する口頭報告がある。笠井・科野 (2017) におけるブラウジング効果とは、情報や気晴らしを求めて消費者が店舗内を回遊し探索する行動により、長時間滞在するにもかかわらず、店内での商品購入につながらない現象をいう。Williams et al. (1985) などによれば、快楽的ショッピングの目標は、①自己満足を求めて得ること、②新しいトレンドを学び社会的経験を追求すること、③交渉や社会的交流、④日常生活からの逃避、⑤自由を楽しむことなどであり、購買を伴わないショッピングでも快楽的な充足感と楽しみが得られるとしている。さらに、Yim, et. al. (2014) は、ブラウジング効果は同伴者には影響されないとしている。

⁴ 2018年度本学卒業生。当時は3年生で学内の「2017年度学生研究発表大会」において口頭報告し、準優勝を受賞している。

図表Ⅲ-3 Yim, et.al. (2014) による快楽消費におけるブラウジング効果



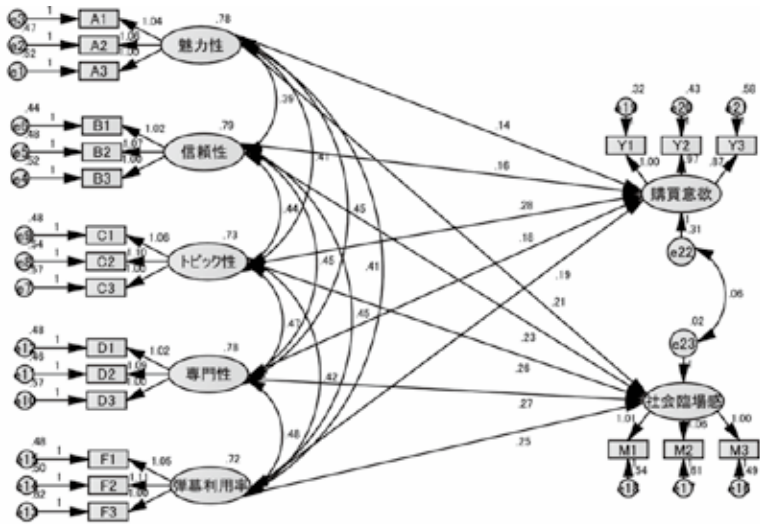
出所：Mark Yi-Cheon Yim & Seung-Chul Yoo & Paul L. Sauer & Joo Hwan Seo (2014) p.511, Fig. 1 Hedonic shopping motivation and purchases in a supermarket

図表Ⅲ-2の分析結果が正しいとすれば、ライブコマースにおいても Yim, et. al. (2014) におけるブラウジング効果 (H1-) に類似する現象が生じ、他の参加者の存在を知覚して「社会臨場感」が高まっても、購買意欲が抑制されるとも考えられる。

しかし、未だ確信が持てないので、「社会臨場感」から「購買意欲」へのパスを1本外し、図表Ⅲ-4に示す間接効果のないモデル2aに単純化し、キャスターの情報発信の特性が、媒介変数と目的変数にどのように影響しているのかを確認した。その結果、最初の奇妙な現象が観察された。

図表Ⅲ-5に示す分析結果によれば、「社会臨場感」への直接効果にはほとんど変化はみられなかったが、「購買意欲」への直接効果がマイナス値からプラス値へと反転した。この分析結果からは直接効果が抑制効果なのか促進効果なのか結論を下せない。しかし、図表Ⅲ-5に示すように、標本数が3200近くもあり十分に多いこともあり、標準誤差は0.05以内に収まりそれほど大きくはない。パスのp値は全て0.1%未満で強く有意であり、特に不安定なわけでもない。謎はさらに深まった。

図表Ⅲ-4 モデル2a (直接効果のみ、間接効果なし)



出所：筆者作成。

注：e22とe23の誤差項間の共分散（相関）を除去しても分析結果には影響しないが、これを付加することで、AGFI値、BIC値が若干改善される。

図表Ⅲ-5 間接効果のパスを外す前後での推定結果の変化

		モデル2b：直接効果＋間接効果				モデル2a：直接効果のみ			
原因変数	結果変数	推定値	標準誤差	検定統計量	確率	推定値	標準誤差	検定統計量	確率
魅力性	社会	0.213	0.016	13.368	***	0.206	0.015	13.574	***
信頼性	会	0.234	0.017	13.754	***	0.226	0.016	13.956	***
トピック性	臨	0.247	0.018	13.519	***	0.268	0.018	14.868	***
専門性	場	0.273	0.019	14.185	***	0.267	0.018	14.535	***
弾幕利用率	感	0.251	0.02	12.794	***	0.25	0.019	13.332	***
魅力性	購買意欲	-0.129	0.037	-3.531	***	0.137	0.021	6.502	***
信頼性		-0.136	0.041	-3.361	***	0.158	0.022	7.045	***
トピック性		-0.17	0.047	-3.615	***	0.287	0.025	11.424	***
専門性		0	—	—	—	0.179	0.025	7.093	***
弾幕利用率		-0.124	0.044	-2.827	0.01	0.197	0.026	7.594	***
社会臨場感		1.245	0.108	11.53	***				

出所：筆者作成。

注1：有意水準のラベル表示について、***：p<0.001、**：p<0.05、+：p<0.1

2. 潜在変数の尺度得点化により生じた奇妙な現象

前節で予想外の分析結果となったモデル 2b（図表Ⅲ-2、再掲）について、本節では潜在変数間の構造モデルは変更せず、測定モデル（因子分析ユニット）を尺度得点化して因子得点に置き換え、図表Ⅲ-7に示されるモデル 2b-f に単純化して考察した。潜在変数の因子得点の測定については、確証的因子分析（CFA）を用いて最尤法で推定した。図表Ⅲ-6に示されるように、因子負荷量行列は単一構造となり、マクドナルドの ω 係数値も全て0.8 前後で、測定尺度の信頼性も高いことが確認された。

図表Ⅲ-6 確証的因子分析による測定モデルの尺度得点化（因子負荷量）

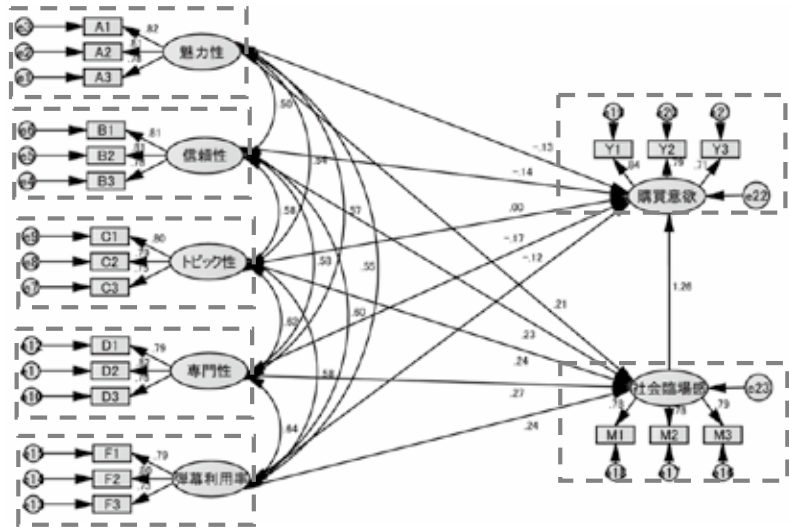
潜在変数 (因子)		魅力性	信頼性	トビツク性	専門性	弾幕利 用率	社会臨 場感	購買意 欲	共通性
観測変数		Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7	
キ ャ ス タ ー の 情 報 発 信 の 特 性	A1	.817	.000	.000	.000	.000	.000	.000	.667
	A2	.811	.000	.000	.000	.000	.000	.000	.661
	A3	.761	.000	.000	.000	.000	.000	.000	.593
	B1	.000	.805	.000	.000	.000	.000	.000	.651
	B2	.000	.819	.000	.000	.000	.000	.000	.666
	B3	.000	.767	.000	.000	.000	.000	.000	.596
	C1	.000	.000	.715	.000	.000	.000	.000	.637
	C2	.000	.000	.745	.000	.000	.000	.000	.645
	C3	.000	.000	.639	.000	.000	.000	.000	.543
	D1	.000	.000	.000	.794	.000	.000	.000	.630
	D2	.000	.000	.000	.823	.000	.000	.000	.678
	D3	.000	.000	.000	.754	.000	.000	.000	.569
	F1	.000	.000	.000	.000	.831	.000	.000	.691
	F2	.000	.000	.000	.000	.818	.000	.000	.670
	F3	.000	.000	.000	.000	.638	.000	.000	.407
媒介 変数	M1	.000	.000	.000	.000	.000	.743	.000	.552
	M2	.000	.000	.000	.000	.000	.840	.000	.706
	M3	.000	.000	.000	.000	.000	.742	.000	.550
目的 変数	Y1	.000	.000	.000	.000	.000	.000	.769	.592
	Y2	.000	.000	.000	.000	.000	.000	.844	.712
	Y3	.000	.000	.000	.000	.000	.000	.739	.546

尺度の信頼度

	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
α 係数	.842	.840	.819	.832	.816	.818	.827
ω 係数	.799	.784	.789	.834	.809	.821	.828

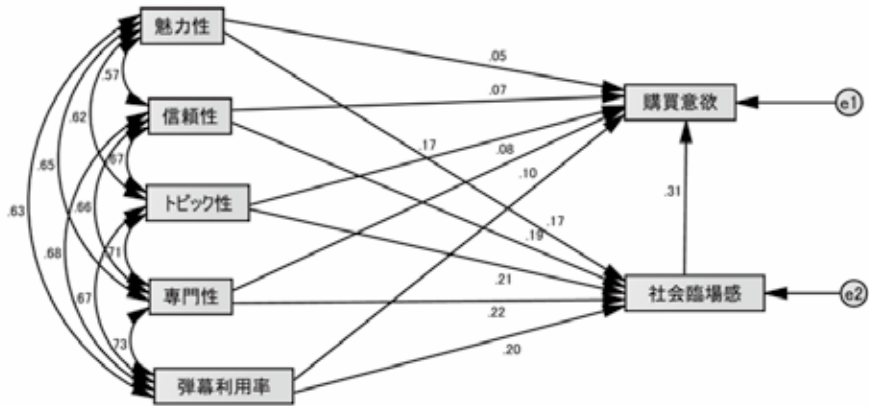
出所：筆者作成。

図表Ⅲ-2 (再掲) モデル2b (購買意欲への直接効果と間接効果あり)



出所：庄豊源 (2024)

図表Ⅲ-7 モデル2b-f (モデル2b における測定モデルを因子得点に置換)



出所：筆者作成。

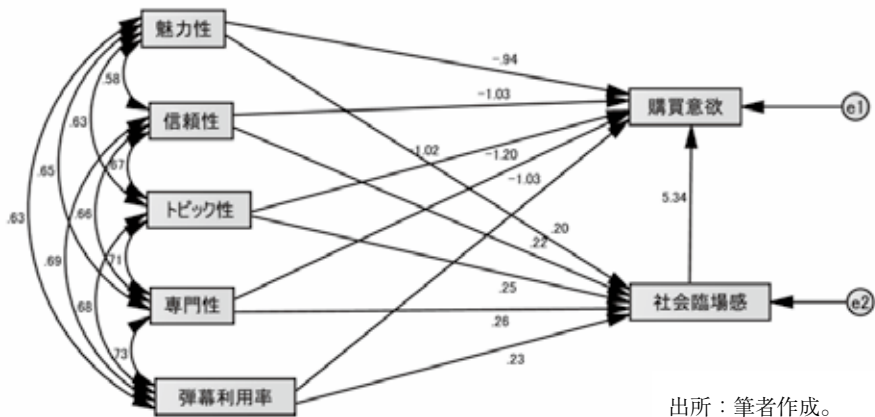
図表Ⅲ-7に示されるように、モデル2bにおいてはマイナス値となっていた「購買意欲」への直接効果のパス係数が、モデル2b-fでは有意ではなかった「トピック性」も含めて、全てプラス値に反転した。すなわち、モデル2bとモデル2b-fは因子間の構造が全く同じモデルであり、同じ観測データを使用しても、モデル2b内で因子得点を測定して分析した場合には「購買意欲」へのパス係数がマイナス値となるが、モデル2b-fの外で予め確証的因子分析(CFA)により因子得点を測定して分析した場合には全てプラス値になるという、2つ目の奇妙な現象が観察された。

そこで、モデル2b-fの使用データをモデル2b内の因子得点に換えて分析を行った。

ただし、筆者が分析に使用したAmos19では、構造方程式モデル中における潜在変数の因子得点は出力されない。そこで、豊田(2007)に従い、平均値で中心化した観測変数行列(X)に因子得点ウェイト行列(W)を乗じて、因子スコア行列(F)を推定した⁵。

分析結果は図表Ⅲ-8に示され、「購買意欲」へのパス係数が再びプラス値に反転した。すなわち、同一のモデルでありながら、因子得点の測定方法の違いで導かれる分析結果が真逆になるという、3つ目の奇妙な現象が観察された。

図表Ⅲ-8 モデル2b-s (モデル2bにおける測定モデルを因子スコアに置換)



⁵ 7個の潜在変数のそれぞれをリッカート尺度に採用した3項目の観測変数で測定し、オペザベーションは3188件なので、行列 X は21列×3188行、行列 W は21行×7列、行列 F は7列×3188行となる。

Ⅳ. 構造モデルに関する理論および考察

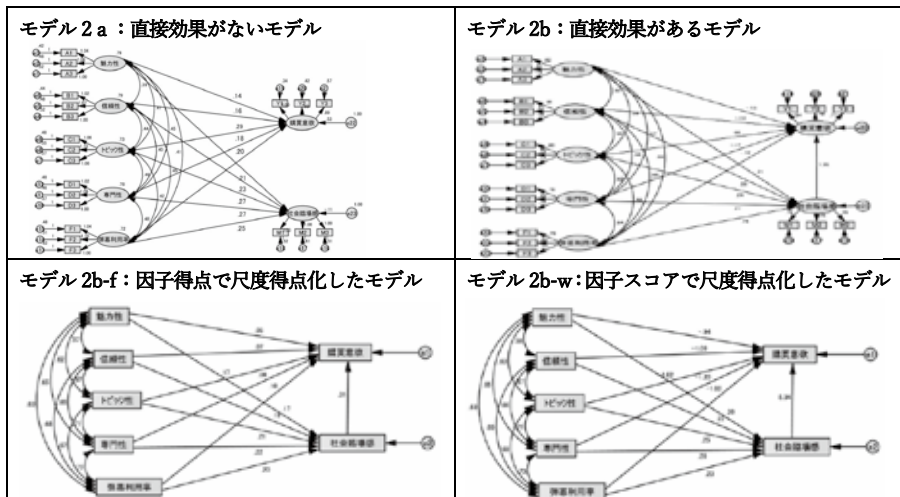
前章で述べた、構造方程式モデリングにおいて観察された一見奇妙に思える現象は以下のように要約される。

【現象1】 モデル2aにおいて、「カスタマーの諸特性」から「購買意欲」への直接効果はプラス値である。しかし、モデル2bに修正し、「社会臨場感」を経由する間接効果のパスを1本付加すると、同じ直接効果がマイナス値に反転する。

【現象2】 「カスタマーの諸特性」から「購買意欲」への直接効果と間接効果を想定したモデル2bにおいて、直接効果はマイナス値であった。しかし、モデル2b-fのように尺度得点化して、測定モデルの部分の素子を因子得点（確証的因子分析により測定）に置換したところ、同じ「購買意欲」への直接効果がプラス値に反転した。

【現象3】 尺度得点化の測定値に、モデル2b-fでは確証的因子分析による因子得点を用いたが、モデル2b-wでは構造方程式モデル中の因子スコアを用いたところ、同じ「購買意欲」への直接効果がマイナス値に反転した。

図表Ⅳ-1 分析モデル



これらのうち、現象1は「構造モデル」、現象2と現象3は「測定モデル」に関する現象であるが、いずれもキャスターの諸特性（原因変数）と購買意欲（結果変数）の因果関係が、モデルの状況によって変化してしまい、不明確であることが問題となっている。本章では、これらのうち、「構造モデル」において観察された現象1について、統計的因果推論に関する既往研究を踏まえて理論的な考察を試みる。

1. 統計的因果推論の研究系譜




原因と結果の因果関係を推論することを目的とした統計的分析の枠組みを総称して、統計的因果推論という。具体的には、実験データや観察データから得られた情報を基に、事象の因果効果を統計的に推定する手法を指す。統計的因果推論の理論的枠組みは20世紀後半から大きく発展したが、特に重要な提唱者である Judea Pearl 流と Donald Rubin & Paul Rosenbaum 流による少なくとも二つの研究系譜が存在し（図表IV）、他に Donald Campbell の系譜もある。

Pearl 流の系譜は、構造的因果モデル（Structural Causal Model：SCM）の枠組みによる因果推論手法であり、有向非巡回グラフ（Directed Acyclic Graph：DAG）を用いた数学的な分析がされている。Pearl（1995）の "Causal Diagrams for Empirical Research" における因果ダイアグラムを用いたバック・ドア・パス基準などがその典型である。本研究において、構造モデルの問題である現象1を考察するには、このような Pearl 流の構造的因果モデルによるアプローチが合致していると考えられる。

一方、Rubin & Rosenbaum 流の系譜は、ポテンシャルアウトカムを用いた欠損値補完の枠組みによる因果推論手法であり、反事実モデル（counterfactual model）を用いて、原因が結果に与える影響を評価する。例えば、ある処置を受けた場合と受けなかった場合の結果の差を比較することで因果効果の有無を推定する。Rubin & Rosenbaum 流の統計的因果推論の具体的な手法には、ランダム化比較実験（Randomized Controlled Trial：RCT）、自然実験（Natural Experiment）、傾向スコアマッチング（Propensity Score Matching）、逆確率重み付け（Inverse Probability Weighting：IPW）、操作変数法（Instrumental Variable Methods：IV）、回帰不連続デザイン（Regression Discontinuity Design）などがある。

統計的因果推論は政策評価や医療研究など多岐にわたる分野で応用され、2000年代

図表Ⅳ-2 確率的因果推論の提唱者

	Judea Pearl 1936～		Donald Bruce Rubin 1943～		Paul R. Rosenbaum
	カリフォルニア(UCLA) 確率的因果的推論 バックドア基準 人工知能に基礎的貢献		ハーバード、統計学名譽教 傾向スコア・マッチングなど、 ランダム化実験について経済学 者インベンズとの共著で教科書		ペンシルベニア(ウォートン) 統計学教授 ハーバード大Rubinの門下 傾向スコア・マッチング

図表Ⅳ-3 「統計的因果推論」関連の実証研究のノーベル経済学賞受賞者

2000年		James Joseph Heckman (プリンストン)		Daniel Little McFadden (ミネソタ)
		ミクロ計量、2段階推定法による内生性問題解消		ミクロ計量、離散選択分析理論とその計算手法の開発
2003年		Robert Fry Engle III (カリフォルニア)		Clive William John Granger (カリフォルニア)
		金融市場における時系列分析手法		時系列分析における因果性の検定
2011年		Thomas John Sargent (ニューヨーク)		Christopher Albert Sims (プリンストン)
		マクロ経済の原因と結果をめぐる実証研究		ベクトル自己回帰モデルによる動学的推計、検定
2019年		Abhijit Vinayak Banerjee (マサチューセッツ工科大学)		Esther Duflo (夫人) (マサチューセッツ工科大学) 開発途上国のミクロ開発経済
		David Edward Card (カリフォルニア)		Joshua Angrist (マサチューセッツ工科大学) 反実仮想モデル
2021年				Michael Kremer (ハーバード) ランダム化比較実験の利用
				Guido Wilhelmus Imbens (スタンフォード) 反実仮想モデル

に入り、経済学を中心とする社会科学分野の実証研究に統計的因果推論を用いた研究者のノーベル経済学賞の受賞が相次いだ（図表Ⅳ-3）。

2. 構造モデルにおける現象 1 に関連する理論

2.1 パス解析による考察

(1) パス解析の基本モデルにおける相関関係と因果関係

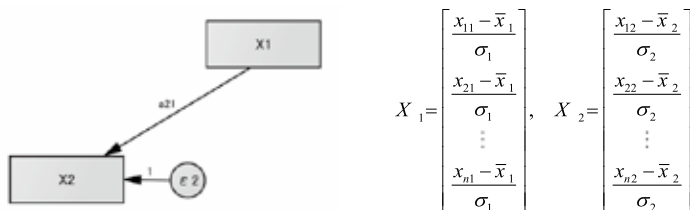
パス解析の基本モデルとして、図表Ⅳ-4 のパス図に示す単回帰モデルを想定する。

パス図中の X_1 , X_2 は、説明変数（原因変数） x_1 と目的変数（結果変数） x_2 について、平均と標準偏差を求めて標準化した変数ベクトルであり、2 つの変数の関係は (4.1) 式で表わされる。

$$X_2 = \alpha_{21} X_1 + \varepsilon_2 \cdots (4.1)$$

ただし、 α_{21} は標準化パス係数（標準化回帰係数）、 ε_2 は誤差項

図表Ⅳ-4 パス解析の基本モデル



(4.1) 式の両辺に X_1 を乗じて期待値をとると、回帰式の傾きを表すパラメータ α_{21} 、すなわちパスダイアグラムにおけるパス係数 α_{21} となる。

$$E[X_1 X_2] = \alpha_{21} \underbrace{E[X_1 X_1]}_{=1} + \underbrace{E[X_1 \varepsilon_2]}_{=0} = \alpha_{21} \cdots (4.2)$$

一方、 X_1 , X_2 の内積は単相関係数 r_{12} となり、これは確率を含まないスカラーであるため、期待値をとってもその値は変わらない。また、説明変数 X_1 と誤差項 ε_2 は無相関を仮定し、右辺第 2 項の内積は 0 となる。

$$\begin{aligned}
 E[X_1 X_2] &= \frac{1}{n-1} \sum_{i=1}^n (x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2) \\
 &= \frac{\frac{1}{n-1} \sum_{i=1}^n (x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2)}{\sigma_1 \sigma_2} \\
 &= \frac{\frac{1}{n-1} \sum_{i=1}^n (x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2)}{\sqrt{\frac{\sum_{i=1}^n (x_{1i} - \bar{x}_1)^2}{n-1}} \sqrt{\frac{\sum_{i=1}^n (x_{2i} - \bar{x}_2)^2}{n-1}}} = \frac{\sum_{i=1}^n (x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2)}{\sqrt{\sum_{i=1}^n (x_{1i} - \bar{x}_1)^2} \sqrt{\sum_{i=1}^n (x_{2i} - \bar{x}_2)^2}} = r_{12} \dots (4.3)
 \end{aligned}$$

よって、(4.2) 式、(4.3) 式より、

$$\alpha_{21} = r_{12} \dots (4.4)$$

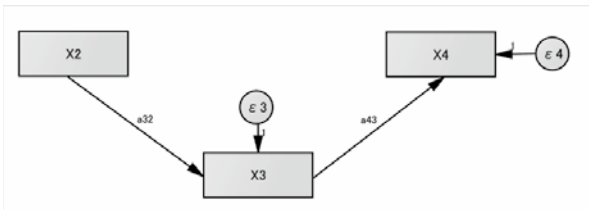
(4.4) 式より、回帰式の傾きを表すパラメータ α_{21} は、説明変数 X_1 と目的変数 X_2 との単回帰係数 r_{12} に一致する。すなわち、交絡因子の影響を受けない状況下では、単回帰モデルにおいて、説明変数 X_1 と、目的変数 X_2 は因果関係にある。

(2) 媒介変数を含む逐次連鎖型モデル

図表Ⅳ-5 のパス図に示す、標準化した3つの観測変数 X_2 , X_3 , X_4 に関する逐次的な連鎖関係において、 X_3 を媒介変数とし、その前後で、前段階を $X_2 \rightarrow X_3$ 方向の単回帰モデル、後段階を $X_3 \rightarrow X_4$ 方向の単回帰モデルとして考察する。

それぞれのモデルにおける変数間の因果関係は、(4.5) 式および (4.6) 式で表される。

図表Ⅳ-5 変数 X_2 , X_3 , X_4 が逐次的な連鎖関係にあるケース



$$\begin{cases} X_3 = \alpha_{32} X_2 + \epsilon_3 \dots (4.5) \end{cases}$$

$$\begin{cases} X_4 = \alpha_{43} X_3 + \epsilon_4 \dots (4.6) \end{cases}$$

(4.6) 式の両辺に X_2 を乗じて期待値をとると、

統計的因果推論に生じた奇妙な現象ともやもや感（深瀬）

$$E[X_2 X_4] = \alpha_{43} \underbrace{E[X_2 X_3]}_{=r_{23}} + \underbrace{E[X_2 \varepsilon_4]}_{=0}$$

$$\therefore r_{24} = \alpha_{43} r_{23} = \alpha_{43} \alpha_{32} = r_{23} r_{34} \dots (4.7)$$

(4.7) 式より、交絡因子の影響を受けない状況下では、媒介変数 X_3 を経由する $X_2 \rightarrow X_4$ への間接効果の影響の強さは、両変数間の単相関係数 r_{24} に等しく、これを X_2 と X_3 との単相関係数 r_{23} と、 X_3 と X_4 との相関係数 r_{34} の積に分解することができる。

一般的に、標準化された観測変数 X_1, X_2, \dots, X_n による n 個の変数間における逐次的な多段階の連鎖関係について、交絡因子の影響を受けない状況下では、漸化計算より、 $X_1 \rightarrow X_n$ の間接効果の影響力を、多段階の単相関係数の積として求めることができる。

$$r_{1n} = \alpha_{nn-1} \alpha_{n-1n-2} \dots \alpha_{21} = \prod_{i=1}^{n-1} \alpha_{(n+1-i)(n-i)} \dots (4.8)$$

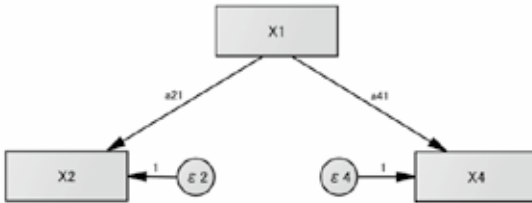
$$= r_{12} r_{23} \dots r_{n-1n} = \prod_{i=1}^{n-1} r_{i(i+1)}$$

(3) 交絡因子の影響を受ける分岐型モデル

1) 擬似相関

図表IV-6において、標準化した観測変数 X_2, X_4 は無相関であるが、共に交絡因子 X_1 からの影響を受けるケースについて考察する。

図表IV-6 変数 X_2, X_4 が共に交絡因子 X_1 の影響を受けるケース



$X_1 \rightarrow X_2$ の単回帰モデルを (4.1) 式、 $X_1 \rightarrow X_4$ の単回帰モデルを (4.9) 式で表す。

$$\begin{cases} X_2 = \alpha_{21}X_1 + \varepsilon_2 \dots (4.1) \text{再} \\ X_4 = \alpha_{41}X_1 + \varepsilon_4 \dots (4.9) \end{cases}$$

(4.9) 式の両辺に X_2 を乗じて期待値をとると、

$$\begin{aligned} E[X_2X_4] &= \alpha_{41} \underbrace{E[X_2X_1]}_{=r_{12}} + \underbrace{E[X_2\varepsilon_4]}_{=0} \\ \therefore r_{24} &= \alpha_{41}r_{12} = \alpha_{41}\alpha_{21} = r_{12}r_{14} \dots (4.10) \end{aligned}$$

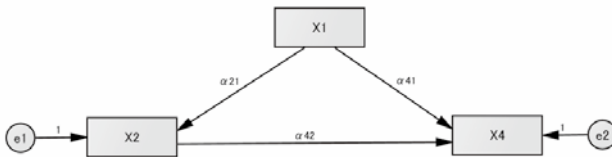
(4.10) 式より、 X_2 , X_4 の間に因果関係がなくても、共に共通の交絡因子 X_1 の影響を受けることで疑似相関 r_{24} が発生し、その強さは、 $X_1 \rightarrow X_2$ の標準化パス係数 α_{21} (すなわち相関係数 r_{12}) と、 $X_1 \rightarrow X_4$ の標準化パス係数 α_{41} (すなわち相関係数 r_{14}) の積に分解できる。この結果は、連鎖型モデルのケース (図表Ⅳ-5, 前掲) と同じである。

ii) 交絡バイアス

図表Ⅳ-7 において、因果関係のある 2 つの標準化した観測変数 X_2 , X_4 について、共に交絡因子 X_1 からの影響を受けるケースについて考察する。

$X_1 \rightarrow X_2$ の単回帰モデルを (4.9) 式、 X_4 への重回帰モデルを (4.11) 式で表す。

図表Ⅳ-7 因果関係に交絡因子が影響するケース



$$\begin{cases} X_2 = \alpha_{21}X_1 + \varepsilon_2 \dots (4.1) \text{再} \\ X_4 = \alpha_{41}X_1 + \alpha_{42}X_2 + \varepsilon_4 \dots (4.11) \end{cases}$$

(4.11) 式に (4.1) 式を代入すると、

$$\begin{aligned} X_4 &= \alpha_{41}X_1 + \alpha_{42}(\alpha_{21}X_1 + \varepsilon_2) + \varepsilon_4 \\ &= \underbrace{\alpha_{41}X_1 + \alpha_{42}\alpha_{21}X_1}_{\text{多重共線性問題}} + \alpha_{42}\varepsilon_2 + \varepsilon_4 \dots (4.12) \end{aligned}$$

(4.12) 式は、 r_{21} が 1 に近い値の場合、多重共線性問題が発生する可能性があることを示している。

また、(4.11) 式の両辺に X_1 , X_2 を乗じて、それぞれについて期待値をとると、

$$E[X_1 X_4] = \alpha_{41} \underbrace{E[X_1 X_1]}_{=1} + \alpha_{42} \underbrace{E[X_1 X_2]}_{=r_{12}} + \underbrace{E[X_1 \varepsilon_4]}_0$$

$$\therefore \alpha_{41} = \alpha_{41} + \alpha_{42} r_{12} = \alpha_{41} + \underbrace{r_{21} r_{24}}_{\text{交絡バイアス}} \dots (4.13)$$

$$E[X_2 X_4] = \alpha_{41} \underbrace{E[X_2 X_1]}_{=r_{12}} + \alpha_{42} \underbrace{E[X_2 X_2]}_{=1} + \underbrace{E[X_2 \varepsilon_4]}_0$$

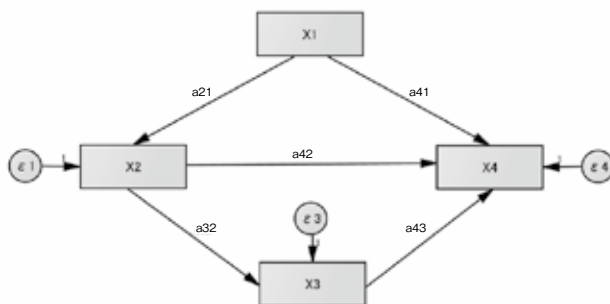
$$\therefore \alpha_{42} = \alpha_{41} r_{12} + \alpha_{42} = \alpha_{42} + \underbrace{r_{41} r_{12}}_{\text{交絡バイアス}} \dots (4.14)$$

(4.14) 式より、 $X_2 \rightarrow X_4$ のパラメータ α_{42} の推定値に、また、(4.13) 式より $X_1 \rightarrow X_4$ のパラメータ α_{41} 推定値に、それぞれ単相関の交絡バイアスが生じる危険性がある。

(4) 交絡因子による疑似相関の影響を受ける場合の単相関の分解

図表IV-8 のパス図は、原因変数 X_2 、結果変数 X_4 は共に交絡因子 X_1 からの影響を受け、 $X_2 \rightarrow X_4$ の因果関係については、直接効果と、媒介変数 X_3 を経由する間接効果による関係をもつケースを示している。

図表IV-8 交絡因子と媒介変数を含むケース



変数間の関係は (4.1) 式、(4.5) 式、(4.15) 式で表される。

$$\begin{cases} X_2 = \alpha_{21}X_1 + \varepsilon_2 \dots (4.1) \text{再} \\ X_3 = \alpha_{32}X_2 + \varepsilon_3 \dots (4.5) \text{再} \\ X_4 = \alpha_{41}X_1 + \alpha_{42}X_2 + \alpha_{43}X_3 + \varepsilon_4 \dots (4.15) \end{cases}$$

(4.15) 式の X_2, X_3 に、(4.1) 式、(4.5) 式を代入すると、

$$\begin{aligned} X_4 &= \alpha_{41}X_1 + \alpha_{42}X_2 + \alpha_{43}X_3 + \varepsilon_4 \\ &= \alpha_{41}X_1 + \alpha_{42}(\underbrace{\alpha_{21}X_1 + \varepsilon_2}_{X_2}) + \alpha_{43}(\underbrace{\alpha_{32}X_2 + \varepsilon_3}_{X_3}) + \varepsilon_4 \\ &= \alpha_{41}X_1 + \alpha_{42}(\alpha_{21}X_1 + \varepsilon_2) + \alpha_{43}\left(\alpha_{32}(\underbrace{\alpha_{21}X_1 + \varepsilon_2}_{X_2}) + \varepsilon_3\right) + \varepsilon_4 \\ &= \left(\underbrace{\alpha_{41}}_{X_1 \rightarrow X_4} + \underbrace{\alpha_{21}\alpha_{42}}_{X_1 \rightarrow X_2 \rightarrow X_4} + \underbrace{\alpha_{21}\alpha_{32}\alpha_{43}}_{X_1 \rightarrow X_2 \rightarrow X_3 \rightarrow X_4}\right)X_1 + (\alpha_{42} + \alpha_{43}\alpha_{32})\varepsilon_2 + \alpha_{43}\varepsilon_3 + \varepsilon_4 \dots (4.16) \end{aligned}$$

(4.16) 式は、交絡因子 X_1 が、 X_4 に至る全ての経路に影響することを示している。
また、(4.15) 式の両辺に X_2 を乗じて期待値をとると、

$$\begin{aligned} E[X_4 X_2] &= \alpha_{41} \underbrace{E[X_1 X_2]}_{=r_{12}} + \alpha_{42} \underbrace{E[X_2 X_2]}_{=1} + \alpha_{43} \underbrace{E[X_3 X_2]}_{=r_{32}} + \underbrace{E[\varepsilon_4 X_2]}_0 \\ \therefore \underbrace{r_{24}}_{\text{単相関}} &= \alpha_{41}r_{12} + \alpha_{42} + \alpha_{43}r_{32} \\ &= \alpha_{42} + \alpha_{43}r_{32} + \alpha_{41}r_{12} \\ &= \underbrace{\alpha_{42}}_{\substack{X_2 \rightarrow X_4 \\ \text{直接効果}}} + \underbrace{\alpha_{32}\alpha_{43}}_{\substack{X_2 \rightarrow X_3 \rightarrow X_4 \\ \text{間接効果}}} + \underbrace{\alpha_{41}\alpha_{21}}_{\substack{X_2 \leftarrow X_1 \rightarrow X_4 \\ \text{疑似相関}}} \dots (4.17) \end{aligned}$$

(4.17) 式は、単相関が直接効果 + 間接効果 + 疑似相関の和に分解できることを示している。

2.2 複数の原因変数からの影響を受けるモデル

(1) 重回帰モデルにおける偏回帰係数

図表Ⅳ-9 は、結果変数 z が、原因変数 x と y から影響を受け、両者が相関を持つ

重回帰モデルのケースである。重回帰式は (4.18) 式で表される。

$$z = ax + by + c + \varepsilon \dots (4.18)$$

$$\therefore \varepsilon = z - ax - by - c$$

OLS 推定により、(4.19) 式に示す残差項 ε の平方和を最小化するように、パラメータ a, b, c の値を推定する。非標準化解、標準化解は (4.20) 式となる（導出は Appendix を参照⁶）。 x, y, z が標準化されているとき、 $\sqrt{S_{xx}} = \sqrt{S_{yy}} = \sqrt{S_{zz}} = 1$ となり推定値は標準化解となる。

$$L(a, b, c) = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (z_i - ax_i - by_i - c)^2 \dots (4.19)$$

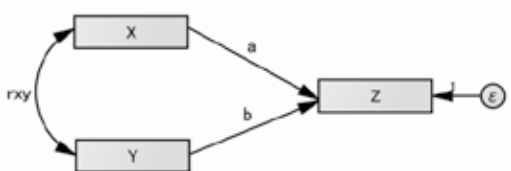
非標準化解

$$\beta = \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} \frac{r_{xz} - r_{xy}r_{yz}}{1 - r_{xy}^2} \times \frac{\sqrt{S_{zz}}}{\sqrt{S_{xx}}} \\ \frac{r_{yz} - r_{xy}r_{xz}}{1 - r_{xy}^2} \times \frac{\sqrt{S_{zz}}}{\sqrt{S_{yy}}} \\ \bar{z} - a\bar{x} - b\bar{y} \end{bmatrix}$$

標準化解

$$\beta^S = \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} \frac{r_{xz} - r_{xy}r_{yz}}{1 - r_{xy}^2} \\ \frac{r_{yz} - r_{xy}r_{xz}}{1 - r_{xy}^2} \\ \bar{z} - a\bar{x} - b\bar{y} \end{bmatrix} \dots (4.20)$$

図表Ⅳ-9 説明変数間に相関をもつ重回帰モデル



⁶ 通常では (19) 式を a, b, c で偏微分して、 a, b, c を未知数とする 3 元の正規方程式を導き、行列表示した連立方程式を解いてパラメータを推定する。ここでは教育用に高校数学で行列を計算するために、先に定数項 C のみを求めて、(1) 式に代入して消去してしまい、正規方程式を 2 次元の行列に下げた。

(4.20) 式について、具体的な数値例で確認する。図表Ⅳ-10において、説明変数 X 、 Y を魅力性、信頼性、被説明変数 Z を社会臨場感とする。

まず図 1 において、標準化されたデータを用いた単回帰分析の結果より、標準化回帰係数、すなわち、単相関係数は、 $r_{zx}=0.75$, $r_{zy}=0.78$ であるが、図 2 において重回帰分析の結果では、偏回帰係数は 0.57、0.64 となり、相関係数に一致しない。これは双方の説明変数の影響を受ける部分が除去（パーシャル・アウト）されるためである。

次に、図 3 において 2 つの説明変数間に相関を示す双方向矢印 (\leftrightarrow) を付加すると、相関係数と偏回帰係数が $r_{xy}=0.5$, $a=0.57$, $b=0.64$ となった。なお、切片 c については 0 となる。

(4.20) 式に図 1 と図 3 で求めた 3 変数間の単相関係数である、 $r_{zx}=0.75$, $r_{zy}=0.78$, $r_{xy}=0.5$ を当てはめて、2 つの偏回帰係数を求める。

$$a = \frac{r_{xz} - r_{xy}r_{yz}}{1 - r_{xy}^2} = \frac{0.75 - 0.5 \times 0.78}{1 - 0.5^2} = \frac{0.75 - 0.39}{1 - 0.25} = \frac{0.36}{0.75} = 0.48$$

$$b = \frac{r_{yz} - r_{xy}r_{xz}}{1 - r_{xy}^2} = \frac{0.78 - 0.5 \times 0.75}{1 - 0.5^2} = \frac{0.78 - 0.375}{1 - 0.25} = \frac{0.405}{0.75} = 0.54$$

この計算結果は図 3 における推定値と一致する。注意すべきは、(4.20) 式は図 3 のように重回帰モデルに双方向矢印 (\leftrightarrow) の相関を設定したパス解析モデルを前提としており、図 2 のような一般的な重回帰分析では成立していないということである。

さらに、(4.4) 式より双方向矢印 (\leftrightarrow) の相関を片側矢印 (\rightarrow) に置き換えた図 4 においても標準化偏回帰係数は、 $a=0.48$, $b=0.55$ となり、推定値はほぼ一致する。

また、図 4 において「魅力性」から「社会臨場感」、「信頼性」から「社会臨場感」への総合効果は、それぞれ、(4.13) 式、(4.14) 式より、 $0.48 + 0.5 \times 0.55 = 0.75$, $0.54 + 0.5 \times 0.48 = 0.78$ であるが、これらは 1 図における単回帰係数 $a=0.75$, $b=0.78$ に一致する。さらに (4.4) 式よりこれらは図 3 と同値モデルとなる。

統計的因果推論に生じた奇妙な現象ともやもや感（深瀬）

図表IV-10 説明変数間に相関をもつ重回帰モデル

図1 単回帰モデル

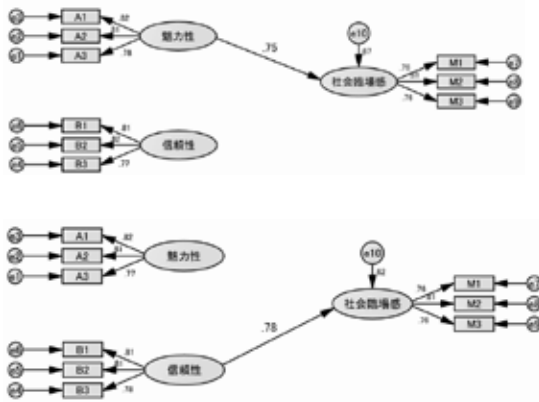


図2 重回帰モデル

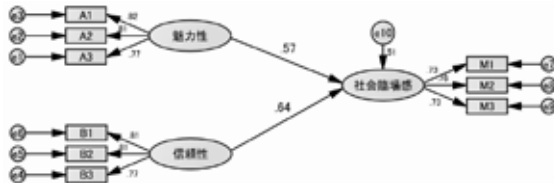


図3 重回帰モデルの説明変数間に相関パスを付加したモデル

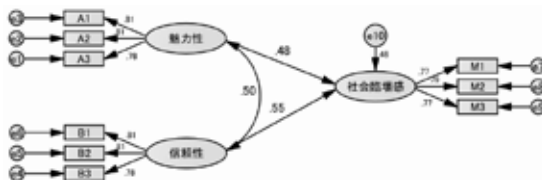
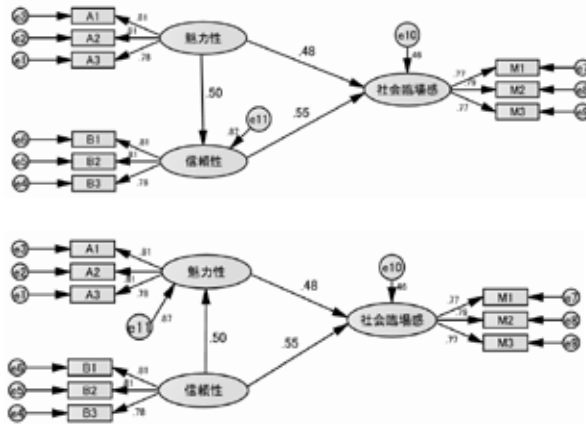


図4 説明変数間の相関を標準化パス係数に置換した連鎖モデル



(2) 重回帰モデルにおける説明変数の欠落バイアス

清水 (2017) における偏回帰係数の概念理解の事例は明瞭なため授業に使用している。図表Ⅳ-9 (前掲) の重回帰モデルおよび (4.18) 式において、 x をコーヒーを飲んだ量、 y を砂糖を入れた量、 z を体重の増加量とする。この重回帰モデルにおいて、医学的な実質科学ではコーヒーを飲んでも体重増加には影響しないとすれば、 x と y を説明変数、 z を目的変数に用いて、適切なパラメータ推定を行った場合は、 x の偏回帰係数 a は有意とならず、 y の偏回帰係数 b がプラス値で有意となるはずである。

しかし、説明変数に x のみを用い、 y を用いずに単回帰モデルで単回帰係数 a を推定した場合は、プラス値で有意となる可能性がある。すなわち、コーヒーを飲むと体重が増加するという、実質科学に反する誤った分析結果となってしまう危険性がある。

テキストには理論的な説明は記されていないので、(4.18) 式のモデルを用いて考察する。なお、説明変数 y が欠落した単回帰モデルは (4.21) 式となり、パラメータが (4.18) 式における推定値と一致しないため * を付けている。

$$z = ax + by + c + \varepsilon \dots (4.18) \text{再}$$

$$z = a^*x + c^* + \varepsilon \dots (4.21)$$

$$\therefore \varepsilon = z - a^*x - c^* \dots (4.22)$$

統計的因果推論に生じた奇妙な現象ともややもや感（深瀬）

OLS より、まず定数項 c^* を推定すると、

$$L(a, b) = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (z_i - a^* x_i - c^*)^2 \dots (4.23)$$

$$\frac{\partial L}{\partial c} = \frac{\partial}{\partial c} \sum_{i=1}^n (z_i - a^* x_i - c^*)^2 = -2 \sum_{i=1}^n (z_i - a^* x_i - c^*) = 0$$

$$\begin{aligned} \therefore nc^* &= \sum_{i=1}^n z_i - a^* \sum_{i=1}^n x_i \\ c^* &= \bar{z} - a^* \bar{x} \dots (4.24) \end{aligned}$$

これを (4.22) 式に代入すると、

$$\varepsilon_i = (z_i - \bar{z}) - a^*(x_i - \bar{x}) \dots (4.25)$$

(4.25) 式を、下記の行列を用いて行列表示すると、(4.26) 式になる。

$$e = Z - a^* X \dots (4.26)$$

ただし、
$$e = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}, \quad Z = \begin{bmatrix} z_1 - \bar{z} \\ z_2 - \bar{z} \\ \vdots \\ z_n - \bar{z} \end{bmatrix}, \quad X = \begin{bmatrix} x_1 - \bar{x} \\ x_2 - \bar{x} \\ \vdots \\ x_n - \bar{x} \end{bmatrix}$$

(4.23) 式、(4.26) 式より、残差平方和は次式となる。

$$\begin{aligned} L(a, b) &= \sum_{i=1}^n \varepsilon_i^2 = \begin{bmatrix} \varepsilon_1 & \varepsilon_2 & \cdots & \varepsilon_n \end{bmatrix} \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \\ &= {}^t e e = {}^t (Z - a^* X) (Z - a^* X) \\ &= {}^t Z Z - 2a^{*t} X Z + (a^*)^2 {}^t X X \dots (4.27) \end{aligned}$$

(4.27) 式に OLS を用いて、単回帰係数 a^* を推定すると、一階条件より

$$\begin{aligned}\therefore \frac{\partial L(a, b)}{\partial a^*} &= -2'XZ + 2a^*{}'XX = 0 \\ \therefore a^* &= ({}'XX)^{-1} {}'XZ \dots (4.28)\end{aligned}$$

(4.28) 式における Z は、(4.21) 式より導出されたものだが、 Z の観測値としては (4.18) 式のモデルを用いて正しくパラメータを推定した場合についても変わらないはずである。そこで、(4.18) 式の Z についても (4.29) 式のように行列表示して (4.28) の Z に代入すれば、(4.30) 式を得る。

$$Z = aX + bY + e \dots (4.29)$$

$$\text{ただし、} Z = \begin{bmatrix} z_1 - \bar{z} \\ z_2 - \bar{z} \\ \vdots \\ z_n - \bar{z} \end{bmatrix}, X = \begin{bmatrix} x_1 - \bar{x} \\ x_2 - \bar{x} \\ \vdots \\ x_n - \bar{x} \end{bmatrix}, Y = \begin{bmatrix} y_1 - \bar{y} \\ y_2 - \bar{y} \\ \vdots \\ y_n - \bar{y} \end{bmatrix}, e = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

※定数項の c については、 $c = \bar{z} - a\bar{x} - b\bar{y}$ だが、 X, Y, Z に含めた。

$$\begin{aligned}a^* &= ({}'XX)^{-1} {}'X \underbrace{(aX + bY + e)}_Z \\ &= a \underbrace{({}'XX)^{-1} ({}'XX)}_I + b \underbrace{({}'XX)^{-1} {}'XY}_{\frac{\sigma_{xy}}{\sigma_{xx}}} + \underbrace{({}'XX)^{-1} {}'Xe}_0 \\ &= a + b \underbrace{\left(\frac{\sigma_{xy}}{\sigma_{xx}} \right)}_{\text{bias}} = a + \underbrace{b \cdot a_{yx}}_{\text{間接効果}} \dots (4.30)\end{aligned}$$

(4.30) 式より、説明変数 y が欠落した回帰モデルで z を予測した場合、説明変数 x の単回帰係数の推定値 a^* に、 y の単回帰係数 b に x から y への単回帰係数 a_{yx} を乗じたバイアスが生じている可能性がある。すなわち、図表Ⅳ-8（前掲）における相関の分解に関する (4.17) 式より、以下の2つのケースが考えられる。

i) $x, \rightarrow z$ の関係において、 y を媒介変数とする $x, \rightarrow y \rightarrow z$ の間接効果に相当す

るバイアスが生じており、 α^* の推定値が x による直接効果とこの間接効果を合算した総合効果に相当する可能性。

- ii) $x \rightarrow z$ の関係において、 y が交絡因子となっていて、原因変数 x と結果変数 z に擬似相関による影響を与えている可能性。

（３）重回帰分析における多重共線性

テキスト等にあまり言及がみられないが、重回帰分析モデルにおいて、説明変数間に共分散ないし相関を設定した場合に、推計結果への影響を調べた。

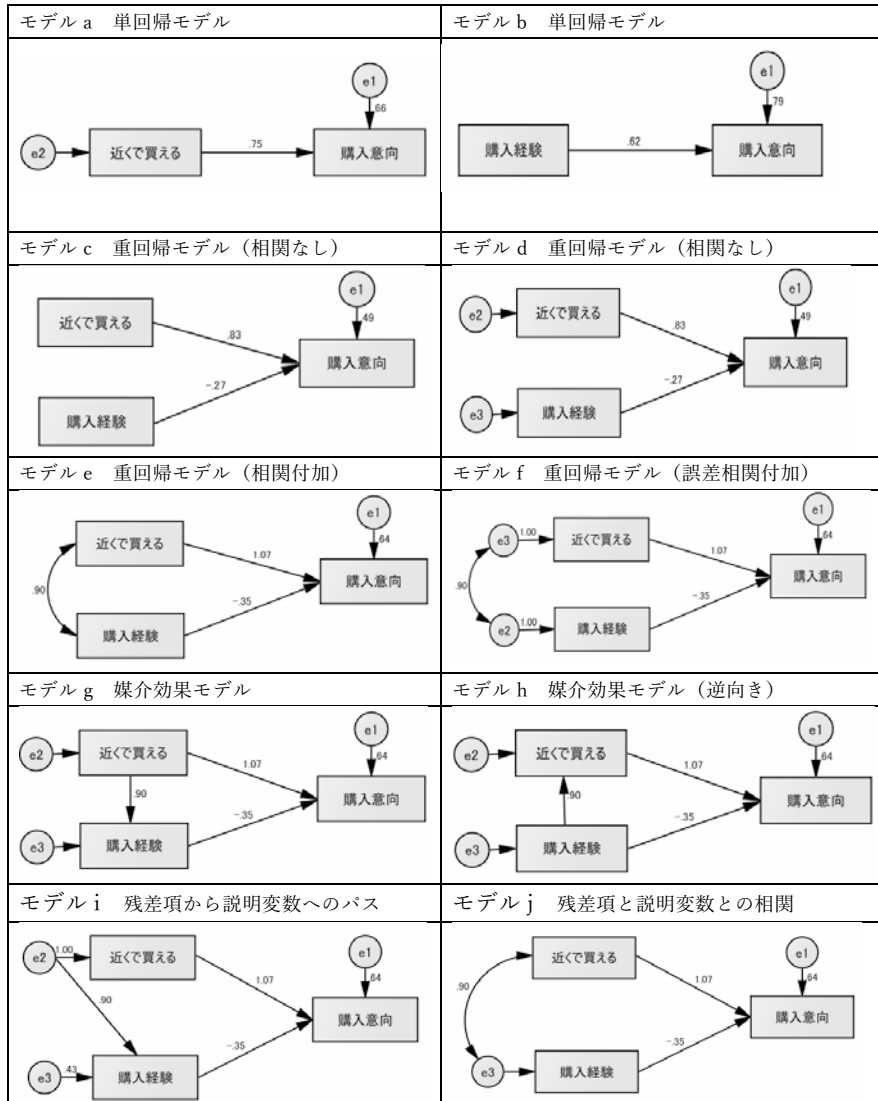
図表IV-11の事例は「購入意向」を目的変数、「近くで買える」と「購入経験」を説明変数とする重回帰モデルにおいて、2つの説明変数間に強い相関があり、多重共線性が発生している設定である。すなわち、「近くで買える」(モデルa)と「購入経験」(モデルb)の単回帰係数はともにプラス値で「購入意向」を促進させるが、重回帰モデルcにおいては「購入経験」の偏回帰係数がマイナス値に転じている。

このような場合の対処法として、インターネット上の解説サイト等では「構造方程式モデルを用いると多重共線の問題が解消される場合がある」との記述を散見する。確かに、観測変数間で多重共線性が発生している場合には、潜在変数を仮定し、これらを測定尺度に用いることで解決できる可能性は高い。

しかし、モデルeおよびモデルfのように、単に相関する観測変数間や残差変数間に、共分散ないし相関の両方向矢印をつなぐだけでは解決しない可能性がある。むしろ偏回帰係数が大きくなり多重共線性問題を増幅してしまう可能性もある。なお、(4.2)式の結果から相関の代わりに標準化パス係数を用いても推定値は同値となる。さらに、残差項を付加して残差項間、残差項と説明変数間において相関を推定しても、また標準化パス係数を推定しても、同値の推定結果となった。(モデルe～j)。

いずれにしても、多重共線性が生じた際の問題解決には、新たな潜在変数を導入して観測変数を尺度得点化する必要がある。

図表IV-11 重回帰分析における多重共線性



2.3 バックドア基準

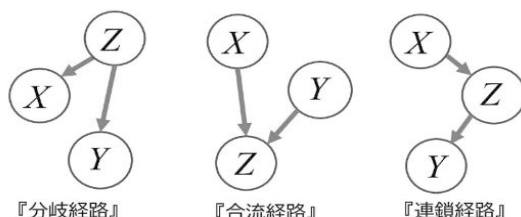
原因変数 X と結果変数 Y の間に因果関係がある場合は X と Y の間に相関が観察される。しかし、前節でみたように X と Y の間に因果関係がなくても、それぞれに関係がある要因 Z が存在する場合にも擬似相関 (spurious correlation) が観察される。このような構造を交絡 (confounding) といい、 X と Y に影響を与えている共通の要因 Z を交絡因子 (confounding factor) という (立森2016)。

重回帰モデルおよびより一般的な構造方程式モデルにおいて、パス係数のパラメータ推定値にバイアスが生じない条件については、Pearl, J. (1993b) および Pearl, J. (1995) が体系的に研究している。最終的な結論は図表IV-13に示される「バックドア基準」の定義に集約されるが、まずは基本型モデルを個別に検討する。

(1) バックドア・パスの3つの基本型モデル

バックドアとは、パス係数のパラメータ推定値にバイアスが生じない条件を $X \rightarrow Y$ に至る裏口経路への出入口の開閉に例えた表現である。バックドア・パスの基本型モデルには、図表IV-12に示される、①分岐経路 (Z は交絡因子)、②合流経路 (Z は共変量)、③連鎖経路 (Z は中間変数) の3タイプがある。バックドア基準では、それぞれ、①の「交絡因子 Z 」は固定し、②の「共変量 Z 」、③の「中間変数 Z 」については固定してはいけなとされている。

図表IV-12 バックドア・パスの基本型



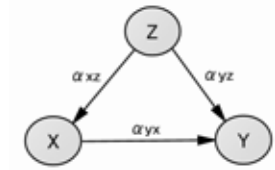
i) 分岐（フォーク）経路

$X \rightarrow Y$ の因果関係において、双方に影響を与える交絡因子 Z が分岐点に位置して、裏路に分岐経路が形成されるケースである。

変数 X, Y, Z に関する構造方程式は次式となる。

$$X = \alpha_{xz}Z + \varepsilon_x \cdots (4.31)$$

$$Y = \alpha_{yx}X + \alpha_{yz}Z + \varepsilon_y \cdots (4.32)$$



(4.32) 式の両辺に X を乗じて、期待値をとると、

$$\begin{aligned} \underbrace{E[XY]}_{\text{cov}[YX]} &= \alpha_{yx} \underbrace{E[XX]}_{\text{var}[X]} + \alpha_{yz} \underbrace{E[XZ]}_{\text{cov}[XZ]} + \underbrace{E[X\varepsilon_y]}_0 \\ \therefore \frac{\text{cov}[YX]}{\text{var}[X]} &= \alpha_{yx} + \alpha_{yz} \frac{\text{cov}[XZ]}{\text{var}[X]} \\ \text{i.e. } \alpha_{yx} &= \alpha_{yx} + \underbrace{\alpha_{yz} \alpha_{xz}}_{\text{bias}} \cdots (4.33) \end{aligned}$$

すなわち、 $X \rightarrow Y$ へのパス係数パラメータに推定バイアスが生じる可能性がある。

なお、この分析結果は、重回帰分析において説明変数が欠落した場合に生じた (4.30) 式と等価であることから、モデルから交絡因子 Z を除外したとしても解決策にはならないであろう。

しかし、(4.32) 式において、 Z を階層 (strata) 化して一定値にしてしまえば、(4.34) 式となり、交絡因子からの影響を除去して、 X と Y との関係を推定することができる。すなわち、交絡因子 Z が分岐経路を構成するケースでは、これをモデルに取り込んで固定化することによって、バックドアを閉じる必要がある。

$$Y = \alpha_{yx}X + \underbrace{\alpha_{yz}\bar{Z}}_{\text{const}} + \varepsilon_y \cdots (4.34)$$

ii) 連鎖経路

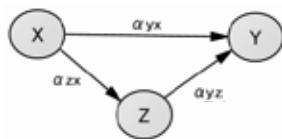
$X \rightarrow Y$ の因果関係において、経路の外に中間変数 Z が存在し、これを経由して X から Y に至る裏路の連鎖経路が形成されるケースである。

統計的因果推論に生じた奇妙な現象ともやもや感（深瀬）

構造方程式は次式となる。

$$Z = \alpha_{zx}X + \varepsilon_z \cdots (4.35)$$

$$Y = \alpha_{yx}X + \alpha_{yz}Z + \varepsilon_y \cdots (4.36)$$



(4.35) 式を (4.36) 式に代入すれば、

$$\begin{aligned} Y &= \alpha_{yx}X + \alpha_{yz}\alpha_{zx}X + \alpha_{yz}\varepsilon_z + \varepsilon_y \\ &= (\alpha_{yx} + \alpha_{yz}\alpha_{zx})X + \alpha_{yz}\varepsilon_z + \varepsilon_y \cdots (4.37) \end{aligned}$$

しかし、中間変数 Z を階層化して固定してしまうと、(4.36) 式において、

$$Y = \alpha_{yx}X + \underbrace{\alpha_{yz}\bar{Z}}_{const} + \varepsilon_y \cdots (4.38)$$

となり、重回帰モデルにおいて、中間変数 Z からの影響が遮られることになり、総合効果が、 $\alpha_{yx} + \alpha_{yz}\alpha_{zx}$ から α_{yx} に減少し、 $\alpha_{yz}\alpha_{zx}$ 分だけ過小評価されてしまう。

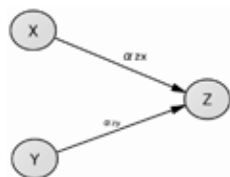
すなわち、中間変数 Z を経由するバックドア・パスが存在するケースでは、中間変数 Z を階層化により固定化せずに、バックドアを開いておく必要がある。

iii) 合流経路

2つの原因変数 X と Y の間に因果関係がなく、合流点に総合選抜を行う共変量 Z が存在して、裏路に合流経路が形成されるケースである。

構造方程式は次式となる。

$$Z = \alpha_{zx}X + \alpha_{zy}Y + \varepsilon_z \cdots (4.39)$$



X と Y を総合評価した Z を階層化して固定した場合、(4.39) 式は (4.40) 式となり、因果関係のない X と Y の間に経路が存在するかのよう擬似相関が発生する。この

ような擬似相関をセレクション・バイアス (selection bias) という。

$$\begin{aligned}\bar{Z} &= \alpha_{zx}X + \alpha_{zy}Y + \varepsilon_z \\ \therefore \alpha_{zy}Y &= \bar{Z} - \alpha_{zx}X - \varepsilon_z \\ Y &= -\frac{\alpha_{zx}}{\alpha_{zy}}X + \frac{\bar{Z} - \varepsilon_z}{\alpha_{zy}} \dots (4.40)\end{aligned}$$

すなわち、 X 、 Y の合流点に共変量 Z が位置するとき、 Z による選抜や階層化等の固定化をせず、バックドアを開いておく必要がある。

(2) バックドア基準

図表IV-13に示す Pearl, J. (1995) の原文では、バックドア基準がこれらを総括する共通概念的に定義され、表現が抽象的で難解であるため、一見すると上記の議論とはテーマが異なる印象さえ受ける。これについては、宮川 (2004) により数学的に解説されているが、さらに林・黒木 (2016) によってカジュアルな用語に置き換えて紹介されている。これらの先行研究の解説を参考にして注釈を加える。

図表IV-13 Pearl, J. (1995) におけるバックドア基準の定義

Definition 3.1 (*back-door*) A set of variables S is said to satisfy the back-door criterion relative to an ordered pair of variables (X_i, X_j) in a DAG G if

1. no node in S is a descendant of X_i , and
2. S blocks every path between X_i and X_j which contains an arrow into X_i .

Similarly, if X and Y are two disjoint subsets of nodes in G , then S is said to satisfy the back-door criterion relative to (X, Y) if it satisfies the criterion relative to any pair (x, y) such that $x \in X$ and $y \in Y$. □

出所：“Causal diagrams for empirical research,” *Biometrika*, vol. 82, p.10より抜粋。

定義 3.3.1（バックドア）① DAG G において、次の条件を満たす変数集合 S は② 順序対 (X_i, X_j) について③ バックドア基準を満たすという。

(i) S の任意の要素は X_i の④ 子孫ではない。

(ii) S は⑤ X_i に向かう矢線を含む⑥ X_i と X_j を結ぶ経路すべてを⑦ ブロックする。

同様に、 X と Y が⑧ G における⑨ 互いに排反な頂点集合であるとき、 S が⑩ 任意の $X_i \in X$ と $X_j \in Y$ に対してバックドア基準を満たすならば、 S は X と Y についてバックドア基準を満たすという

<解説>

① DAG(Direct Acyclic Graph)=有向非巡回グラフ：丸と矢印で描かれた因果構造において、 X から下流側に矢印をたどったときに Y につながる経路をいう。

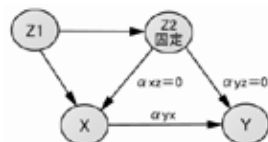
② $X_i \rightarrow X_j$ の因果関係を意味するが、非巡回的有向独立グラフでは変数間の因果関係を規定しない。

③ 「 $X \rightarrow Y$ の介入効果の推定において、バイアスが生じない条件」という意味である。

④ 「モデルに追加した説明変数は X の下流側でない」という意味であり、「 X の下流側にある中間変数、合流点を説明変数として加えてはいけない」ことを含意している。

⑤ 「外から⑥ $[X \rightarrow Y$ の因果構造] に向かう矢印において、 X に向かう矢印も含めて」、すなわち、上流側から X への入力となる（ X の裏側のバックドアから入る）矢印をさす。

⑦ 「 X と Y の上流側から X と Y の両方に影響を与える流れが全て遮断される」、すなわち、「モデルに合流点ではない交絡因子が少なくとも 1 つ（ Z_2 のみ）説明変数として追加し、階層化されることによって、交絡が消失されている状態」を意味している。



⑧ 「因果ダイアグラム（ $G = (V, E)$ ）において」、 V (vertex) は丸で示される頂点 (node)、 E (edge) は 2 つの頂点のペア、すなわち辺を示す。

⑨ グラフの頂点集合をいくつかの部分集合に分割したとき、各部分集合内の任意の 2 頂点が互いに隣接していない、つまり、どの 2 頂点も直接的な辺で結ばれていないような集合のことを指す。

⑩ 「どの 2 頂点も直接的な辺で結ばれていないようなグラフ頂点集合 (V) の部分集合 X 、 Y に属する任意の X_i, X_j について、」

V. 構造モデルにおける現象1の真相に関する考察

1.1 バックドア基準による現象1に関する考察

本節では、庄豊源(2024)による構造モデルにおいて、原因変数から結果変数への直接効果で観察されたパス係数の符号反転(現象1)について、バックドア基準の観点から、総合効果を直接効果と間接効果に分解して考察する。

(1) モデル1の検討結果

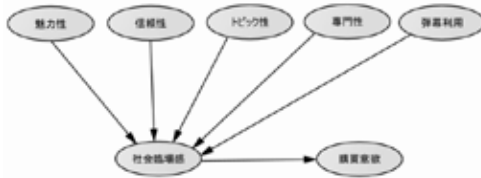
庄豊源(2024)のモデル1は図表V-1に示される。キャストのタイプ特性(原因変数)→社会臨場感(中間変数)→購買意欲(結果変数)、による連鎖経路のモデルである。ただし、簡略化のため観測変数、残差項、共分散を省略している。

中間変数の社会臨場感について階層化等による固定はしていないので、連鎖経路に関してはバックドアが開いた状態であり特に問題はない。また、連鎖経路外の分岐点に存在する可能性がある交絡因子のうち、消費者タイプについては4タイプに階層化して多母集団分析を行っており、販売対象とする商品カテゴリについても4カテゴリに分けて分析されているので、分岐経路に関してもバックドアが閉じられており、庄豊源(2024)の本文においては特に問題はない。

なお、消費者タイプおよび商品カテゴリを階層化せずに統合して分析を行っても、特に問題が発生しなかったため、これらの交絡要因については無視できる範囲であると判断して分析を行った。モデル1において、キャストの特性による購買意欲への影響力については、直接効果はなく間接効果のみである。第IV章で示した構造方程式の計算式を用いると、購買意欲への総合効果は標準化値で0.96程度と推定される。

$$\begin{aligned}\text{間接効果} &= (0.203 + 0.225 + 0.259 + 0.238 + 0.27) \times 0.85 \\ &= 1.195 \times 0.85 = 0.962\end{aligned}$$

図表V-1 モデル1 庄豊源（2024）モデル



連鎖経路：中間変数である「社会臨場感」は固定されていない。バックドアが閉じているため、バイアスは生じていない。

魅力性			0.203	***
信頼性			0.225	***
専門性	→	社会臨場感	0.259	***
弾幕利用率			0.238	***
トビック性			0.27	
社会臨場感	⇒	購買意欲	0.805	***

	社会臨場感	購買意欲	
直接効果	-	-	0
間接効果	1.195	0.805	0.9620
総合効果			0.962

GFI	AGFI	CFI	RMSEA
0.957	0.943	0.965	0.101

（2）モデル2 bの検討結果

図表V-2は、直接効果の影響力がマイナス値となる現象が観察されたモデル2bを示す。5個のキャストのタイプ特性要因が、モデル中の原因変数である「社会臨場感」と結果変数である「購買意欲」の双方に影響する交絡因子となっているため、バックドアが開いた状態である。したがって、分岐経路が構成されており、「社会臨場感」から「購買意欲」へのパスに疑似相関が生じている可能性がある。

キャストのタイプ特性要因による「購買意欲」への影響力は標準化値で、間接効果が3.34、直接効果が-2.39程度で、総合効果は0.95程度と推定される。

まず、間接効果が標準化値で3.34となり、モデル1の0.96に比べて異常に高い数値となった。これは(4.8)式の連鎖計算(1.21×2.75)において、後半の中間変数「社会臨場感」から結果変数「購買意欲」が2.75の高い数値となったためである。ところで(4.4)式によれば、標準化パス係数は単相関係数に等しいはずであるから、その最大値1を超えることは奇異にも思えるが、疑似相関が生じたために、真の標準化パス係数にプラス値の非常に強いバイアスが加算されたことによるものと考えられる。

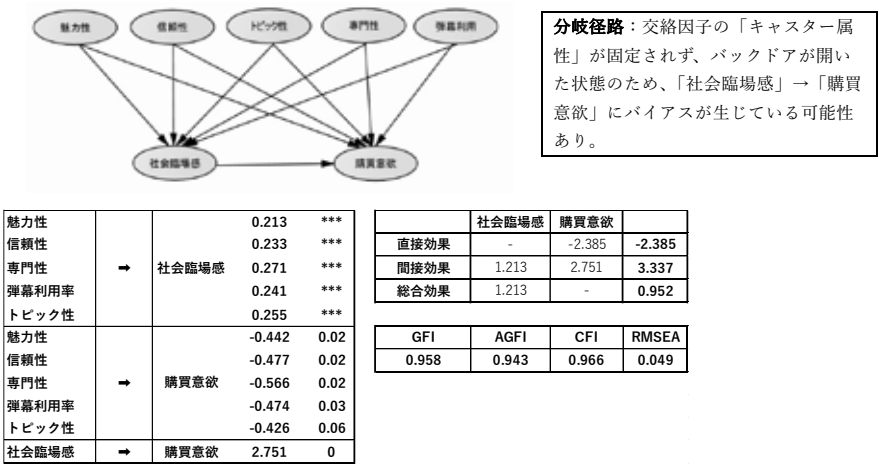
次に、直接効果がマイナス値の-2.39となった原因について考察する。

「キャストの属性」から「購買意欲」への総合効果について、「モデル1」では直接効果はなく間接効果のみだが0.96であった。「モデル2b」では、これに直接効果が付加されるが0.95でほぼ同じ値である。総合効果はいずれも変数間の単相関係数を示す値であるが、(4.17)式の単相関係数の分解式より、総合効果は、直接効果+間接効

果+擬似相関に分解することができ、逆に、これらの合計は一定の値となる。

モデル 2b ではモデル 1 に直接効果が付加されたために、中間変数と結果変数の間に強い擬似相関の影響が生じて、間接効果にはプラス値のバイアスが上乘せられて 3.34 もの高い値となったと考えられる。一方、直接効果では、間接効果に付加されたバイアスを打ち消す強さのバイアスが奪われ、それが真の標準化パス係数を上回ったために、直接効果がマイナス値に反転したことが考えられる。直接効果において観察された符号反転現象は、このような擬似相関により生じたのではないだろうか。

図表V-2 モデル2b 直接効果がマイナス値となったモデル



1.2 修正モデルの検討

直接効果において観察された符号反転現象の原因が解明された。バイアスを除去するには、分岐点において交絡因子となっている 5 個の「カスタマーの属性」を固定することによってバックドアを閉じる必要がある。しかし、庄豊源 (2024) では、カスタマーの特性要因による消費者への影響を分析目的としたため、この交絡因子が擬似相関の原因であっても、この要因を階層化により固定することは厳しい。ほかにも傾向スコア分析によるマッチング、重み付け、DID (Difference In Difference) 分析等を用いて、カスタマーの属性による影響を除去することで理論的には解決できる可能性も考えられる。しかし、交絡因子が 5 個もあり、それぞれについて直接効果を分析

すれば組み合わせが膨大になるため今後の課題とし、今回はモデル2bの構造を修正して解決を試みたい。

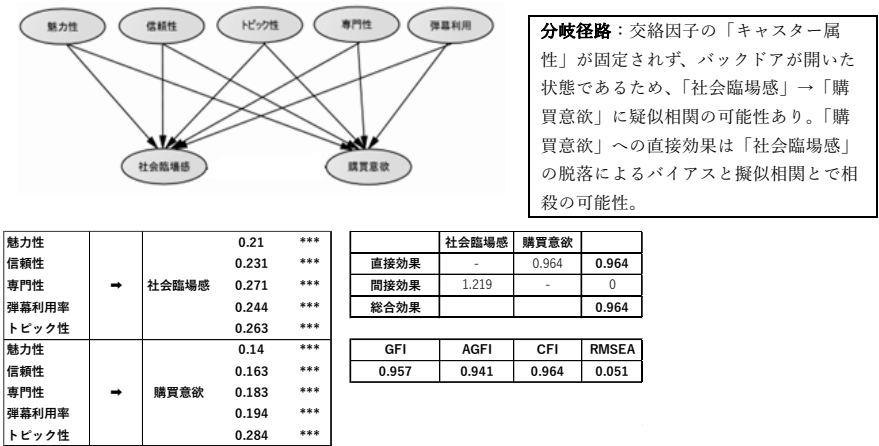
分岐経路を解消するために間接効果の経路を断ち切ることを試みるが、経路は中間変数の前後に分けられる。間接効果の後半を切断したモデル2aを図表V-3、前半を切断したモデル3を図表V-4に示す。

(1) モデル2a(間接効果の後半を切除したモデル)

間接効果後半の「社会臨場感」から「購買意欲」へのパスを絶ったモデル2aでは、(4.7)式の連鎖計算において後半の値が0となり、キャストの特性要因から購買意欲への直接効果のみのモデルとなった。総合効果はすべて直接効果に集約され、標準化値で0.96程度となった。

しかし、「社会臨場感」と「購買意欲」との間に生じている擬似相関は依然として、-0.71程度の強さで残存している。なお、モデルに擬似相関の双方向矢印の相関を付加したところ、モデル適合度指標は表に示した数値よりもわずかながら改善した。

図表V-3 モデル2a 間接効果がなく直接効果のみのモデル



(2) モデル3(間接効果の前半を切除したモデル)

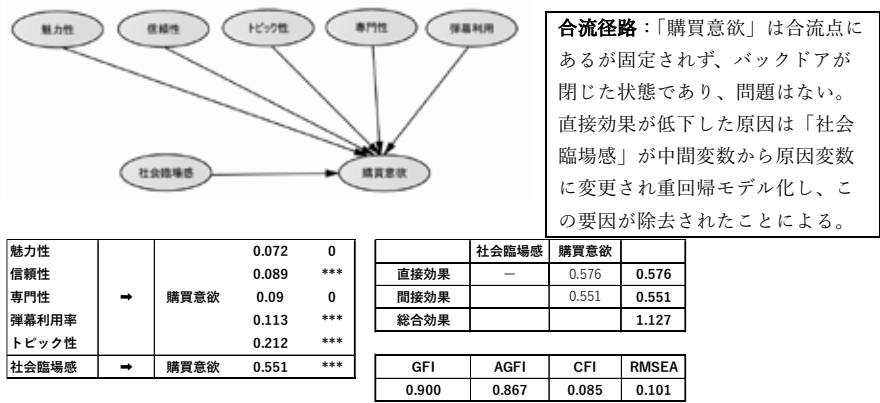
間接効果前半のキャストの特性要因から「社会臨場感」へのパスを絶ったモデル3では、間接効果はなくなり、総合効果が直接効果に集約されていると考えられるが、

数値は0.58程度であり、0.96程度であった他のモデルに比べて低い数値である。

他のモデルでは、原因変数がキャスターの特性要因のみであり、「社会臨場感」は連鎖経路における中間変数に用いられていた。一方、モデル3では重回帰モデルにおいてこれらが原因変数に用いられている点が異なる。すなわち、連鎖経路においては間接効果に「社会臨場感」による影響も含まれていたが、これが重回帰モデルにおける説明変数として独立したため、キャスターの特性要因による直接効果から「社会臨場感」による影響が除外（パーシャルアウト）されて偏回帰係数化したことが考えられる。他方の「社会臨場感」による直接効果についても偏回帰係数化されているが、原因変数が2つになったため、これらによる「購買意欲」への影響力を合計すれば1.13程度に増加したことになる。なお、「購買意欲」はこれらによる合流点に位置する共変量となる。しかし、ここでは多母集団同時分析を行わず、選抜や階層化による固定はしていないので、バックドアは開かれており、特に合流経路においてセレクション・バイアスが発生している問題も考えられない。

ただし、他のモデルよりも適合度が低下している。この理由として、本来は総合効果であるべきキャスター特性による影響が直接効果に集約され、間接効果となる連鎖経路によるパラメータ調整が不能となり、モデルが硬直化したことが影響していると考えられる。

図表V-4 モデル3 間接効果の前半を切断

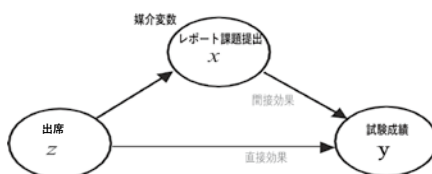


（3） Rosenbaum 流の修正モデルによる推定結果

直接効果における符号反転現象は、バックドア基準が満たされず、擬似相関バイアスの影響で生じることが判明した。そこで、図表V-3（前掲）における原因変数、結果変数に影響する交絡因子の経路のうち、原因変数に向かう経路を断ち切った結果、図表V-4（前掲）に示す6個の説明変数をもつ重回帰モデルに帰着し、Pearl が論じたバックドア基準については満たされた。しかし、図表Ⅲ-2に前掲したように統計的因果推論の研究系譜には諸説があり、それぞれの方向に広がりつつある。

岩崎（2021）はImbens、Rubin、Rosenbaum等による研究を踏まえ、重回帰分析で誤りやすい注意点を、図表V-5に示す授業データ分析の実例で紹介している。このモデルは、本研究における修正モデル3と同じ構造であり、「 z （出席）から y （試験成績）への因果のパスにおいて x （課題提出）は中間変数であり、 x （課題提出）から y （試験成績）への因果のパスでは、 z （出席）は交絡要因となる」と想定されている。このような状況におけるRosenbaum流の分析では、「出席（ z ）が試験（ y ）に与える影響を評価する場合には、モデルに課題（ x ）を取り込んではいならない、および、課題（ x ）が試験（ y ）に与える影響を評価する場合には出席（ z ）をモデルに取り入れなければならない」とし、「出席（ z ）と試験（ y ）との関係では単回帰式、課題（ x ）と試験（ y ）の関係では重回帰式における係数が評価の対象となる。重回帰式での出席（ z ）の係数は何ら意味を持たない値であり、解釈の対象とすべきではないことになる。」としている。

図表V-5 授業データの因果ダイアグラム



出所：岩崎（2020）、p276、図3

本研究においてRosenbaum流の回帰分析をするならば、図表V-4（前掲）のモデル3に修正を加えて、「社会臨場感」（ x ）による「購買意欲」（ y ）への間接効果については重回帰分析、「キャスターの属性」（ z ）による「購買意欲」（ y ）への直接効果について単回帰分析を用いる必要がある。ただし、「キャスターの属性」（ z ）は既に5

つの潜在要因による重回帰モデルが構成されているため、PLS モデル (Partial Least Squares Model) に組み替えて、これらを 1 つの潜在変数に集約した。最終的な分析モデルおよび分析結果は図表 V-6 に示される。

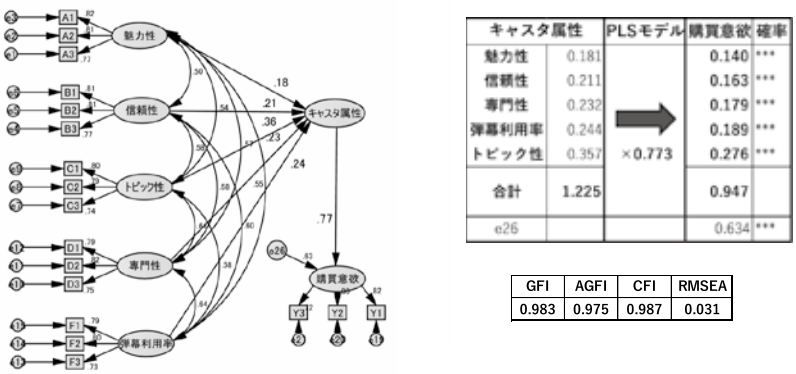
1 図において「キャスター属性」から「購買意欲」への標準化パス係数、すなわち、直接効果は 0.947 であり、他のモデルの総合効果 0.96 と近い値となった。(4.17 式) で示した単相関係数の分解式が、ほぼ成立していることから、信頼度の高い結果であると考えられる。なお潜在変数「キャスター属性」を外して直接、観測変数を説明変数とする重回帰分析を行った場合の標準化偏回帰係数は、表の右欄の数値に一致した。

また、2 図において「社会臨場感」から「購買意欲」への標準化パス係数は 0.551 であり、「キャスター属性」による直接効果よりも弱い、「購買意欲」を促進していることが確認された。ただし、「キャスター属性」から「購買意欲」への標準化パス係数 0.577 は、岩崎 (2021) によれば「何ら意味を持たない値であり、解釈の対象とすべきではない」ことになる。

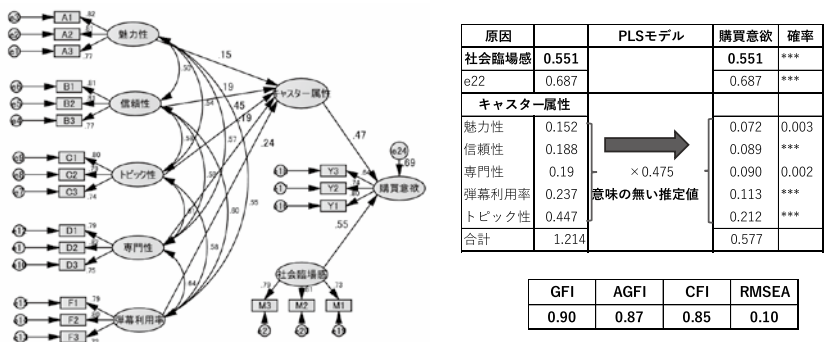
統計的因果推論に生じた奇妙な現象ともややや感（深瀬）

図表V-6 Rosenbaum 流の修正モデルおよび分析結果

1 図：単回帰モデルによる「キャスター属性」から「購買意欲」への影響分析



2 図：重回帰モデルによる「社会臨場感」から「購買意欲」への影響分析



1.3 検討結果の小括

庄豊源 (2024) によるモデル1は、キャスターのタイプ特性（原因変数）→社会臨場感（中間変数）→購買意欲（結果変数）、による連鎖経路のモデルであった。これにキャスターの属性から「購買意欲」への直接効果を付加したモデル2bでは、①キャスターのタイプ特性を交絡因子とする分岐経路、②「購買意欲」を共変量とする合流経路、の2つのバックドア・パスが形成された。

①の分岐経路については、バックドアが開いた状態であるため、交絡因子による強い擬似相関によって間接効果にプラスのバイアスが加算された一方で、直接効果には

真のパス係数を上回るマイナスのバイアスが加算されたために、パス係数がマイナス値となる現象が発生したことがわかった。

そこで、キャストのタイプ特性→「社会臨場感」→「購買意欲」の連鎖経路を前半、後半で絶ち切って、間接効果に擬似相関バイアスが加算されないようにしたところ、モデル2b、モデル3における直接効果はいずれもプラス値で有意となった。

さらに②の合流経路について、間接効果を形成する連鎖経路の前半を断ち切ったモデル3は重回帰モデルとなったが、合流経路においてセレクション・バイアスの発生はなく、「社会臨場感」から購買意欲への直接効果については、キャストのタイプ特性による影響をパーシャルアウトした偏回帰係数より求めることができ、0.551となった。

また、本研究着想の発端となったキャスト属性による購買意欲への直接効果は、単回帰モデルより0.947であることが判明し、解決に至った。

したがって、ライブコマースに夢中になるあまり購買意欲が抑制されるといったブラウジング効果が発生することは考えにくい。

VI. 測定モデルにおける現象2と現象3の真相に関する考察

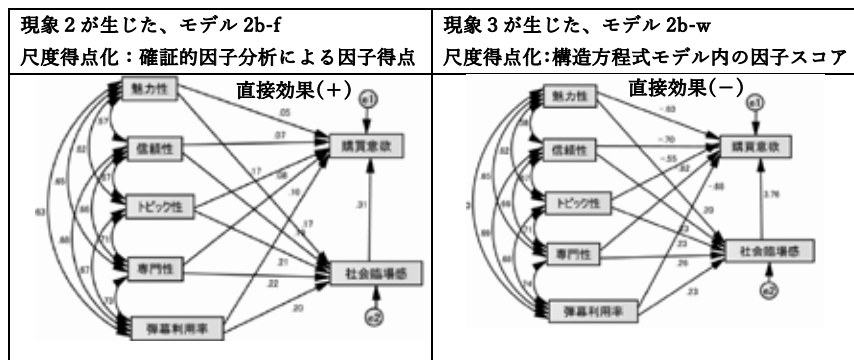
構造方程式モデルは、パス解析である「構造モデル」と、因子分析により潜在変数を予測する「測定モデル」より構成される。前述した奇妙な現象は「購買意欲」への直接効果を推定した際に観察された。第IV章冒頭にまとめたように、現象1は「構造モデル」に関する現象であり、直接効果がマイナス値となったが、第V章で解明された。

残る現象2と現象3も直接効果における符号の反転現象だが、いずれも連鎖経路モデル2bにおける「測定モデル」部分を尺度得点化したことに起因したものである。あらためて論点を図表VI-1で整理する。現象2は「測定モデル」部分を予め確証的因子分析で測定した因子得点に置換したことにより生じたマイナス値からプラス値への反転現象である。現象3は構造方程式モデル内で分析中に測定された因子スコアに置換したことにより生じたマイナス値への復帰現象である。すなわち、同一の構造方程式モデルの測定モデル部分を尺度得点化したが、因子分析と構造方程式モデル内で測定した潜在変数の測定値の違いによって分析結果が真逆となったのである。したがっ

て、因子分析で測定した因子得点は、構造方程式モデル中で異なる値に変容してデータ処理がされている可能性が浮上する。

本章では、「観測モデル」において観察されたこれらの現象について、関連する既往研究を踏まえて理論的な考察を試みる。

図表VI-1 現象2と現象3が生じたモデルと尺度得点化に用いたデータ



出所：図表IV-1（前掲）を拡大して、論点を整理した。

1. 測定モデルの尺度得点化に関する先行研究

①測定誤差による相関の希薄化問題

狩野（2002a）⁷は、重回帰分析に尺度得点を用いる場合の弊害を指摘している。問題は原データが尺度得点化される過程で、データ変換誤差が混入されるために変数の標準誤差が大きくなってしまい、変数の本質的な特性が希薄化されてしまうことである。

標準誤差が拡大すれば、相対的にパラメータ推定値や検定統計量が低下することは容易に想像できるが、構造方程式によるパラメータ推定では、生データを相関係数行列または分散共分散行列に変換してデータ処理が行われる。狩野裕（2002a）は、(6.1) 式に示すモデルを用いて、因子得点の測定値による潜在変数間の分散を、真の相関と信頼係数 α の積に分解し、 α が1よりも小さい値をとることから、潜在変数間の相関が真の相関よりも弱められることを導いている。これが希薄化（Attenuation）

⁷ 狩野裕（2002a）「共分散構造分析は、因子分析、重回帰分析にとって代わるのか？」

である。

因子得点や尺度得点は潜在変数ごとに評価されるが、それぞれの信頼性が異なれば偏回帰係数の大きさに基づいて、目的変数への影響力を比較する分析が無効となってしまう。また、偏回帰係数が小さな値であったり有意にならない場合でも、尺度の信頼性が低く因子間相関が希薄化されたために過小に評価されている可能性もある。

x, y を真の変数の得点、 X, Y を誤差を含む変数の得点、 r_{xy} を相関係数とする。

$$\begin{aligned}
 r_{XY} &= \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}} = \frac{\text{Cov}(x, y)}{\sqrt{\text{Var}(x) + \text{Var}(\varepsilon_x)}\sqrt{\text{Var}(y) + \text{Var}(\varepsilon_y)}} \\
 &= \frac{\text{Cov}(x, y)}{\sqrt{\text{Var}(x)}\sqrt{\text{Var}(y)}} \frac{\sqrt{\text{Var}(x)}\sqrt{\text{Var}(y)}}{\sqrt{\text{Var}(x) + \text{Var}(\varepsilon_x)}\sqrt{\text{Var}(y) + \text{Var}(\varepsilon_y)}} \\
 &= \frac{\text{Cov}(x, y)}{\sqrt{\text{Var}(x)}\sqrt{\text{Var}(y)}} \sqrt{\frac{\text{Var}(x)}{\text{Var}(x) + \text{Var}(\varepsilon_x)}} \sqrt{\frac{\text{Var}(y)}{\text{Var}(y) + \text{Var}(\varepsilon_y)}} \\
 &= \underbrace{\frac{\text{Cov}(x, y)}{\sqrt{\text{Var}(x)}\sqrt{\text{Var}(y)}}}_{\text{真の得点間の相関係数}} \underbrace{\sqrt{\alpha_x \alpha_y}}_{\text{信頼係数の相乗平均}} \dots (6.1)
 \end{aligned}$$

②希薄化問題の修正

米村 (2003)⁸は、希薄化 (Attenuation) とは「尺度得点化した際の構成概念は誤差項を含んでしまうため、興味の対象である構成概念間の関係 (構成概念間の相関係数やパス係数) の絶対値が小さくなる現象をいう」とし、「SEM では観測変数から誤差を分離するモデリングを行い 希薄化を修正できる」とし、2つの具体策を示している。

1つ目は「観測変数群に対して、確認的因子分析や多重指標モデルを適用する」手法で、いずれも探索的因子分析よりも測定尺度を自由に調整できる利点があり希薄化問題の軽減に有効であると思われる。しかし、次の2つ目の方法も含めて、クローンバック α 係数を1にすること、すなわち誤差を完全に除去することは困難であろう。

2つ目は「尺度得点に対して、信頼性係数 α と尺度得点の分散 σ_x^2 から誤差分散 σ_e^2

⁸ 豊田 (2003) 編『共分散構造分析 [疑問編]』における分担執筆、pp.87-89

⁹ 原文では信頼係数を ρ とし、 $\sigma_e^2 = \sigma_x^2 (1 - \rho)$ となっているが、誤植の可能性。

を算出し、固定母数として利用する」方法である。まず、誤差分散 σ_e^2 は、(6.2) 式より求められる⁹。モデル中の尺度得点の変数に誤差項を付加して、誤差分散のパラメータを手動で計算値に設定することで希薄化を修正できるとしている。

$$\alpha = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_e^2} \Leftrightarrow \sigma_e^2 = \sigma_x^2(1 - \alpha) / \alpha \dots (6.2)$$

③データ使い回し・水増し批判

小杉 (2018) は、「因子分析によって因子を特定し、因子得点を算出し、因子得点を説明変数にして別の変数に回帰分析を行う、という分析の流れをとることができますが（かつてはよく行われていました）」とした上で、「分析者にとって、得られるデータは1つしかありません。それを何度も切り貼りして、一部は因子分析に、一部は回帰分析に、あるいは何度か回帰分析を行って、という使い方をしているのは、データの水増しをしているようなものだ」と批判している。その理由については、「適合度もそれぞれに算出されて、研究全体としての統合的判断が難しいという問題があるのです」、「実質的に得られる情報は全体の共変動で、すべてのデータからなる分散共分散行列 S ただ1つの中にすべて埋め込まれているのです」（以上 p.182）などを挙げている。

すなわち、構造方程式モデルでは共分散行列を用いるが、因子分析では相関行列を用いるので平均の情報が損失すること、パス解析における連鎖的な計算過程で誤差が拡大していくことを問題視していると思われる。

2. 下位尺度得点を用いたパス解析モデルとの比較

2.1 本事例における先行研究の検証

尺度得点化したデータを回帰分析や構造方程式モデルに用いる分析手法は一般的に用いられており、テキストや図表 I-1 に前掲した *International Journal of Advertising* をはじめとする最近の学術誌にさえ今もなお掲載されていることから、小杉 (2018) の指摘は過激に感じられ、筆者にとっても耳が痛い。しかし、本研究の着想に至ったパラメータ推定値の符号反転現象が観察されており、相関の希薄化や

データの増しで、この現象を説明できるのかは不明だが、従来の分析手法に問題がある可能性は高い。

そこで、まずは先行研究の妥当性を軽く確認するため、図表Ⅵ-2に示したモデル1、モデル2a、モデル2b、モデル3の構造方程式モデル内の測定モデルの素子（四角の破線で囲った部分）を確証的因子分析（CFA）により測定した下位尺度得点（ウェイト付けをしない単純平均）に置換したパス解析モデルを作成して平均構造分析を行い、構造方程式モデルとパス解析モデルによる推定結果を比較してみた。なお、平均構造分析におけるパス係数は通常の共分散構造分析による推定値にほとんど近い値となるが、いずれも非標準化推定値で出力される。分析結果は図表Ⅵ-3に示される。

図表Ⅵ-3 非標準化パス係数の推定結果の比較

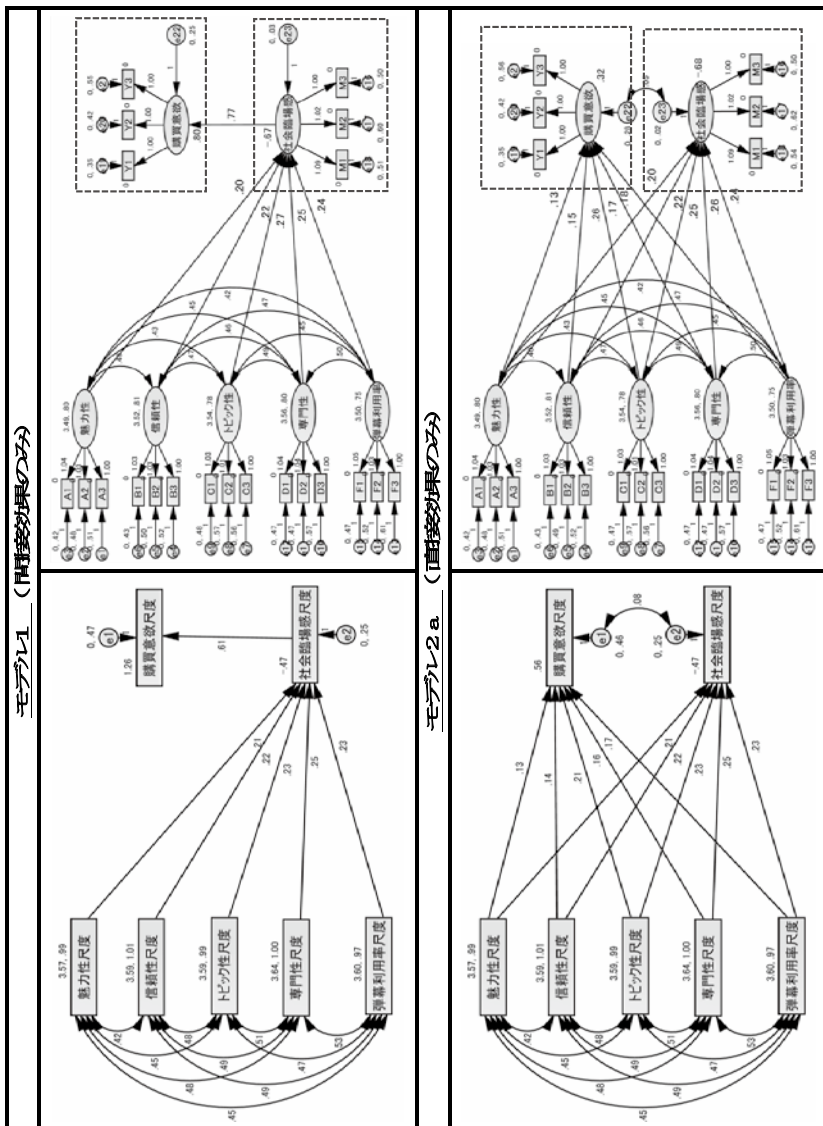
構造 方程式	From To	外生変数					内生変数 社会臨場感
		魅力性	信頼性	トピック性	専門性	弾幕利用率	
モデル1	社会臨場感	0.198	0.217	0.254	0.270	0.240	—
	購買意欲	—	—	—	—	—	0.767
モデル2a	社会臨場感	0.205	0.223	0.252	0.263	0.240	—
	購買意欲	0.132	0.150	0.261	0.167	0.182	—
モデル2b	社会臨場感	0.205	0.223	0.252	0.263	0.240	—
	購買意欲	-0.503	-0.541	-0.521	-0.650	-0.562	3.101
モデル3	社会臨場感	—	—	—	—	—	—
	購買意欲	0.060	0.072	0.173	0.075	0.098	0.429

パス 解析	From To	外生変数					内生変数 社会臨場感
		魅力性	信頼性	トピック性	専門性	弾幕利用率	
モデル1	社会臨場感	0.206	0.219	0.234	0.248	0.229	—
	購買意欲	—	—	—	—	—	0.611
モデル2a	社会臨場感	0.206	0.219	0.234	0.248	0.229	—
	購買意欲	0.128	0.143	0.214	0.158	0.166	—
モデル2b	社会臨場感	0.206	0.219	0.234	0.248	0.229	—
	購買意欲	0.066	0.077	0.143	0.083	0.096	0.303
モデル3	社会臨場感	—	—	—	—	—	—
	購買意欲	0.066	0.077	0.143	0.083	0.096	0.303

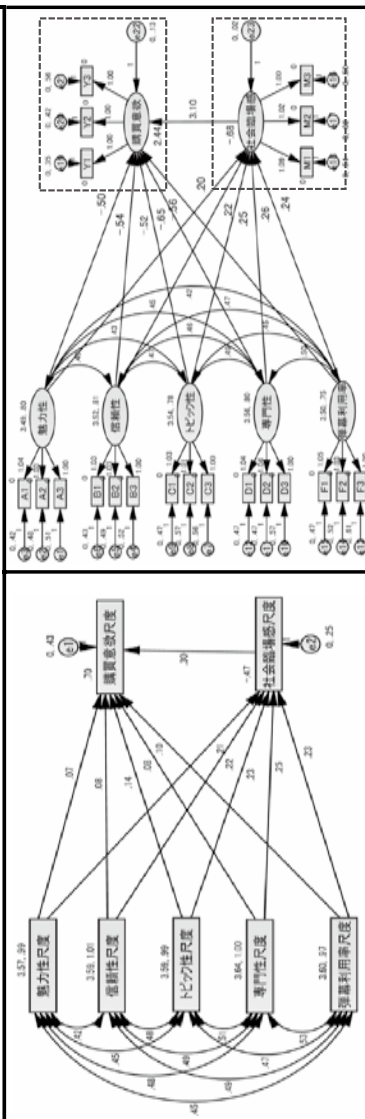
外生変数からの非標準化パス係数の推定値は、両モデル間で近い値となっている。しかし、例えばモデル1の社会臨場感に影響するキャスターの要因について両者を比較すると、信頼性、トピック性、専門性、弾幕利用では構造方程式モデル（0.217, 0.254, 0.27, 0.24）がパス解析モデル（0.219, 0.234, 0.248, 0.229）を上回るが、魅力性については、構造方程式モデル（0.19）よりもパス解析モデル（0.206）が大きく、優劣の逆転がみられる。このような傾向は、モデル2a、モデル2bについても同様である。特にモデル2bにおいては、構造方程式モデルでは購買意欲へのパス係数が全てマイナス値となるが、パス解析モデルでは全てプラス値となり真逆の分析結果となった。

さらに、内生変数からのパス係数では推定値の乖離が大きくなる傾向がみられた。この原因として、パス解析モデルではパス係数の推定値が比較的に安定しているのに対し、構造方程式モデルではモデルの形状によって変動する傾向が影響している可能

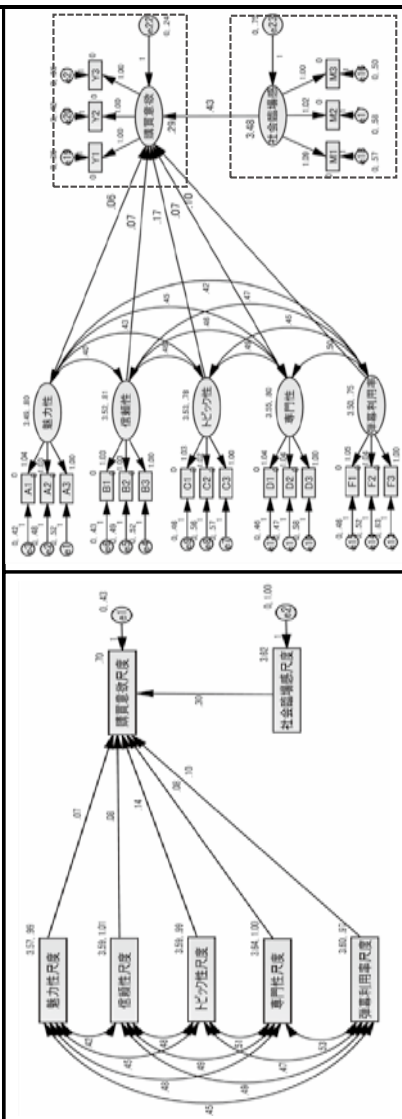
図表VI-2 比較対象する構方程式モデル（平均共分散構造分析）



モデル2b (間接効果+直接効果)



モデル3 (直接効果+社会総効果)



注：片方向の矢印脇の数値は非標準化パス係数、両方向の矢印の脇の数値は共分散を示す。

性がある。これらの結果は、通常の構造方程式モデルによる標準化パス係数で比較しても同じであった。

このように、「目的変数への影響力を比較する分析が無効」とする狩野（2002a）の指摘は本事例にも該当し、これを支持する結果となった。

2.2 因子平均と因子分散の比較

さらに平均構造分析の結果を深掘りし、因子平均と因子分散の値からパス解析モデルと構造方程式モデルを比較する。平均構造分析では、外生変数については因子平均が出力されるが、社会臨場感と購買意欲などの内生変数については平均値が出力されずに切片として出力される。豊田（2007）は、内生変数の因子平均は、切片の値に原因変数からの直接効果と間接効果を合計して求めるとしている。

また、分散についても外生変数については因子分散が出力されるが、内生変数については出力されない。足立（2006）によれば、直接には測定できないため残差項 e のパス係数を1に固定してその分散値から受ける影響を独自の因子分散とし、これに原因変数から影響を受ける直接効果、間接効果を合計した値を内生変数の因子分散とする。

図表VI-5において、まず、いずれのパス解析モデルにおいても、魅力性尺度（因子平均3.57、分散0.99）をはじめとする弾幕利用尺度までの外生変数の平均値については3.5前後の値であり、構造方程式モデルとの推定値は僅差であったが、分散についてはパス解析モデルの方が大きい値となる傾向がみられた。

また、内生変数における切片については、モデル2bを除いてパス解析モデルの方が大きい値となる傾向がみられた。特にモデル2aにおいては、パス解析モデルの購買意識尺度は0.559であるのに対して、構造方程式モデルでは0.274であるように、多くの原因変数からの影響を受ける内生変数ほど、切片の値が小さくなる傾向がみられた。

前に掲げた先行研究では、パス解析による問題の発生原因が測定尺度の α 係数などの信頼性に起因している。しかし、本事例では α 係数は全て0.8以上の高い値だが、内生変数の切片の符号が一致しないこと、多くの変数との関連を持つほど乖離が拡大することから、測定誤差以外にも原因がある可能性が考えられる。

2.3 平均共分散構造分析による考察

さらに因子分散についてもモデルを比較し、因子平均の分析結果と併合して考察する。ただし、Amos19で出力される因子分散の推定値は外生変数のみである。そこで、内生変数の因子分散は、足立（2006）に倣い原因変数の分散値を非標準化パス係数の2乗でウェイト付けして合計し、誤差変数から受ける「独自の因子分散」に合算した。

なお後先になるが、基準とした確証的因子分析（CFA）における因子平均および因子分散については、以下のように算出した。

まず、確証的因子分析（CFA）により、因子負荷量が単純構造となるように、魅力性、信頼性、トピック性、専門性、弾幕利用率、社会臨場感、購買意欲の7因子を抽出した。なお、先行研究において、測定尺度の信頼性に起因する問題が議論されるが、図表VI-4に示されるように、本分析では全ての α 係数が0.8を上回り良好だった。

次に、これらを潜在変数とする構造方程式モデルによる独立モデルを作成してから、社会臨場感、購買意欲を除く5因子間に共分散を設定して、平均構造分析を行った。

7因子は全て外生変数であり、因子平均および因子分散の推定値を直接に求められる。

図表VI-4 確証的因子分析（CFA）による測定尺度の信頼度

	魅力性	信頼性	トピック性	専門性	弾幕利用率	社会臨場感	購買意欲
α 係数	.842	.840	.819	.832	.816	.818	.827
ω 係数	.761	.724	.724	.833	.801	.821	.828

平均構造分析に基づく因子平均および因子分散の算出結果は図表VI-5（前掲）に示される。外生変数の平均値はいずれのモデルについても、CFAの値とほぼ近い値となった。ただし、構造方程式モデルに比べてパス解析モデルの方が、若干誤差が大きかった。内生変数の切片の値についてはモデルによって異なり、マイナス値が多数見られた。

しかし、切片に直接効果および間接効果の影響を加えた平均値については、いずれのモデルもCFAの因子平均値に近い値となり、推定値のバラツキは極めて小さかった。

つまり、内生変数の平均値 M は一定の値となり、切片を m_0 、間接効果を m_l 、直接効果を m_d とすれば、次の関係式が成り立っているようである。

$$M = m_0 + m_l + m_d \cdots (6.3)$$

図表VI-5 平均共分散構造分析による潜在因子の平均と分散

表1. 構造方程式モデルにおける因子平均の推定結果

平均	観測変数の平均値					社会臨場感の平均値			購買意欲の平均値			
	魅力性	信頼性	トビック性	専門性	彈幕利用率	切片	直接効果	平均値	切片	間接効果	直接効果	平均値
モデル1	3.489	3.518	3.536	3.555	3.503	-0.668	4.152	3.484	0.797	2.672	0	3.469
モデル2a	3.489	3.518	3.536	3.555	3.503	-0.706	4.191	3.485	0.274	0.000	3.188	3.462
モデル2b	3.489	3.518	3.536	3.555	3.503	-0.682	4.167	3.485	2.441	10.805	-9.780	3.467
モデル3	3.488	3.517	3.534	3.554	3.501	3.484	0	3.484	0.291	1.495	1.684	3.470
CFA	3.488	3.518	3.535	3.554	3.501	3.477	0	3.477	3.459	0	0	3.459

表2. パス解析モデルにおける因子平均の推定結果

平均	観測変数の平均値					社会臨場感の平均値			購買意欲の平均値			
	魅力性	信頼性	トビック性	専門性	彈幕利用率	切片	直接効果	平均値	切片	間接効果	直接効果	平均値
モデル1	3.57	3.593	3.592	3.643	3.6	-0.473	4.091	3.618	1.265	2.210	0	3.475
モデル2a	3.57	3.593	3.592	3.643	3.6	-0.473	4.091	3.618	0.559	0.000	2.913	3.472
モデル2b	3.57	3.593	3.592	3.643	3.6	-0.473	4.091	3.618	0.702	1.096	1.674	3.472
モデル3	3.57	3.593	3.592	3.643	3.6	3.617	0	3.617	0.702	1.096	1.674	3.472
CFA	3.488	3.518	3.535	3.554	3.501	3.477	0	3.477	3.459	0	0	3.459

表3. 構造方程式モデルにおける因子分散の推定結果

分散	観測変数の分散					社会臨場感の分散			購買意欲の分散			
	魅力性	信頼性	トビック性	専門性	彈幕利用率	分散	直接効果	総分散	分散	間接効果	直接効果	総分散
モデル1	0.8	0.81	0.782	0.797	0.754	0.034	0.221	0.255	0.252	0.221	0.000	0.473
モデル2a	0.799	0.809	0.779	0.796	0.752	0.008	0.225	0.233	0.267	0.000	0.137	0.404
モデル2b	0.8	0.81	0.782	0.797	0.754	0.015	0.222	0.237	0.135	2.280	1.227	3.642
モデル3	0.785	0.794	0.732	0.781	0.723	0.754	0	0.754	0.244	0.139	0.040	0.423
CFA	0.8	0.81	0.782	0.797	0.754	0.75	0	0.750	0.695	0	0	0.695

表4. パス解析モデルにおける因子分散の推定結果

分散	観測変数の分散					社会臨場感の分散			購買意欲の分散			
	魅力性	信頼性	トビック性	専門性	彈幕利用率	分散	直接効果	総分散	分散	間接効果	直接効果	総分散
モデル1	0.995	1.006	0.986	1.003	0.973	0.468	0.257	0.725	0.249	0.257	0.000	0.506
モデル2a	0.995	1.006	0.986	1.003	0.973	0.249	0.257	0.506	0.458	0.000	0.134	0.592
モデル2b	0.995	1.006	0.986	1.003	0.973	0.249	0.257	0.506	0.435	0.046	0.257	0.739
モデル3	0.995	1.006	0.986	1.003	0.973	1	0	1.000	0.435	0.092	0.046	0.573
CFA	0.8	0.81	0.782	0.797	0.754	0.75	0	0.750	0.695	0	0	0.695

誤差項で測った独自の因子分散について両モデルを比較すると、全体的に構造方程式モデルに比べてパス解析モデルの方が大きい値となった。このため、CFA による基準値との乖離も広がっている。ただし、外生変数については双方ともにモデル間における推定値のバラツキは極めて小さくほぼ一定の値となっている。

しかし、内生変数については、モデルによって独自の分散値が異なり、CFA の値と異なる。全体的に構造方程式モデルに比べてパス解析モデルの方が大きい値となったため、CFA による基準値との乖離も広がった。

また、独自の分散に間接効果、直接効果を加えた総分散についても、モデルにより値が異なり CFA の値に一致しない。全体的に構造方程式モデルに比べてパス解析モデルの方が大きい値となった。ただし、唯一、奇妙な現象が観察されたモデル 2b については、構造方程式モデル（表 3）において、購買意欲の総分散（3.642）が表 4 に示すパス解析モデル（0.739）を大きく上回った。この原因は、社会的臨場感を媒介する間接効果による分散（2.28）の影響が異常に強いことによると考えられる。

2.4 数理モデルによる考察

モデル 2b において、なぜ、間接効果による分散が異常値になったのかを簡単な数理モデルで考察する。厳密な計算ではないが、モーメント（積率）の次元より分散値が平均値の 2 乗に比例すると仮定して、CFA における潜在変数の分散との乖離を概算する。外生変数、間接効果、直接効果の分散をそれぞれ、 $v_0 = km_0^2$, $v_I = km_I^2$, $v_D = km_D^2$, ($k > 0$) とすれば、総分散 V は次式となり、平均の内訳項目 m_0 , m_I , m_D のそれぞれについての増加関数である。

$$V = v_0 + v_I + v_D = k(m_0^2 + m_I^2 + m_D^2) \cdots (6.4)$$

$$\nabla V = \begin{bmatrix} \frac{\partial V}{\partial m_0} \\ \frac{\partial V}{\partial m_I} \\ \frac{\partial V}{\partial m_D} \end{bmatrix} = 2k \begin{bmatrix} m_0 \\ m_I \\ m_D \end{bmatrix} \cdots (6.5)$$

そこで、内生変数の平均値 M の配分についてラグランジュ乗数法を用いて総分散 V の最小化問題を解くと、

切片 m_0 、間接効果 m_I 、直接効果 m_D に対して、 $m_0^* = m_I^* = m_D^* = \frac{M}{3}$ となるよう

に等配分される場合に、総分散 V は最小化されることがわかる。すなわち、配分に偏りが生じて、特定の項目が大きくなると総分散 V が拡大することになる。

$$\begin{cases} \text{Min } V = k(m_0^2 + m_I^2 + m_D^2) \dots (6.6) \\ \text{s.t. } M = m_0 + m_I + m_D \end{cases}$$

$$L = k(m_0^2 + m_I^2 + m_D^2) + \lambda(M - m_0 - m_I - m_D) \dots (6.7)$$

最小化の一階条件より、 M の最適配分、 $m_0^* = m_I^* = m_D^*$ が次式のように定まる。

$$\begin{cases} \frac{\partial L}{\partial m_0} = 2km_0 - \lambda = 0 \\ \frac{\partial L}{\partial m_I} = 2km_I - \lambda = 0 \\ \frac{\partial L}{\partial m_D} = 2km_D - \lambda = 0 \\ \frac{\partial L}{\partial \lambda} = M - m_0 - m_I - m_D = 0 \end{cases} \Leftrightarrow \begin{cases} m_0^* = \frac{\lambda}{2k} = \frac{M}{3} \\ m_I^* = \frac{\lambda}{2k} = \frac{M}{3} \\ m_D^* = \frac{\lambda}{2k} = \frac{M}{3} \end{cases} \dots (6.8)$$

$$\therefore M - \frac{\lambda}{2k} - \frac{\lambda}{2k} - \frac{\lambda}{2k} = M - \frac{3\lambda}{2k} = 0 \Leftrightarrow \frac{\lambda}{2k} = \frac{M}{3}$$

この結果は、潜在因子間の因果関係进行分析する場合は、構造方程式モデルを構築する方が潜在因子の平均が間接効果や直接効果に偏らずに分散されやすいため、総分散値が減少するという重要な可能性を意味する。

図表VI-5において、中間変数の「社会臨場感」、結果変数の「購買意欲」についても、確かにCFAの分散に比べて、構造方程式モデルの分散が小さい値となっている。

因子平均を柔軟に配分できるのは、より多くのパラメータをもつ構造方程式モデルの方がパス解析モデルよりも調整機能が高いためである。この数理モデルでは、分散値が平均値の2乗に比例すると仮定し、内生変数の独自の因子分散をその誤差分散値で測定した。したがって、因子分散についても調整機能が高いことから、独自の因子分散を縮小できていた構造方程式モデルでは、よりデータに適合した柔軟なパラメータ推定が可能になると考えられる。

表3で奇妙な現象が生じたモデル2bにおいて、「購買意欲」の総分散値が3.642と突出して高い値となった。これは交絡因子の影響を受けて「社会臨場感」から「購買

意欲」へのパス係数にバイアスが生じて、間接効果に平均値を超える10.8が配分された一方、直接効果には-9.78のマイナス値が配分され、極端な偏在が生じたためでありモデルの不備を適切に警告した。一方、パス解析は異常への対応ができなかった。

3. モデル構造中における潜在変数の変容仮説

3.1 因子得点を用いたパス解析モデルとの比較

実証分析では潜在変数の測定値に下位尺度得点が広く使用され、狩野(2002)は測定尺度の信頼係数に起因する希薄化を指摘した。本稿ではそれ以外のモデル構造に起因する柔軟性に着目する。先行研究の米村(2003)を踏まえ、分析に因子得点を使用することで測定誤差の軽減を図った上で、構造方程式モデルとパス解析モデルを比較する。

①構造方程式モデルにおける因子スコアの測定方法

豊田(2007)より、潜在因子 i ($i = 1 \sim 7$)について回答者 j ($j = 1, 2, \dots, 3188$)が有する構成概念スコア f_{ij} を、アンケート回答における第 i 因子の測定尺度の観測変数 Z_{ijk} を平均周りに中心化し、因子得点ウエイト行列 W における潜在因子を決定するための尺度(観測変数) k のウエイト w_{ik} をかけて、次式で求めた。

$$f_{ij} = \sum_{k=1}^3 w_{ik} (Z_{ijk} - \bar{Z}_{i \cdot k}) = \sum_{k=1}^3 w_{ik} \left(Z_{ijk} - \frac{1}{3188} \sum_{j=1}^{3188} Z_{ijk} \right) \dots (6.9)$$

②パス解析モデルにおける因子得点の測定方法

確証的因子分析(CFA)を用いると因子負荷量を単純構造化できるので、構造方程式モデルと同様に、他の潜在変数の測定尺度の影響を受けずに潜在変数を測定できる。さらに因子得点には因子負荷量が重み付けされるため、単純平均で求めた下位尺度得点よりも構造方程式モデルの測定値に近づくはずである¹⁰。

¹⁰ 因子分析における共通因子と独自因子とは形式的な区別に過ぎず、独自因子についても固有値分解を行っているが、因子得点を用いても独自因子の情報は反映されないため測定誤差が無くなるわけではない。

因子得点を求めるには、まず以下のようにして因子負荷量行列を求める必要がある。詳細については Appendix に回すが、要約すれば、生データより相関係数行列を作り、最も素朴な主因子法では、固有値分解（スペクトル分解）によって、共通因子負荷量行列の積と独自因子の平方に分解して求める。さらに因子得点は、回答データの基準化行列、相関行列の逆行列、因子負荷量行列の積として求めることができる¹¹。

まず因子に用いる変数（標準化）間の相関係数行列を行列積の和に分解する。

$$\text{因子分析の基礎方程式： } \mathbf{R} = \mathbf{W}^t \mathbf{W} + \mathbf{E}^2 \dots (6.10)$$

$$\begin{bmatrix} 1 & r_{12} & \dots & r_{1m} \\ r_{21} & 1 & \dots & r_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ r_{m1} & r_{2m} & \dots & 1 \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1m} \\ w_{21} & w_{22} & \dots & w_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ w_{m1} & w_{m2} & \dots & w_{mm} \end{bmatrix} \begin{bmatrix} w_{11} & w_{21} & \dots & w_{m1} \\ w_{12} & w_{22} & \dots & w_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ w_{12} & w_{2m} & \dots & w_{mm} \end{bmatrix} + \begin{bmatrix} e_{11}^2 & e_{12}^2 & \dots & e_{1m}^2 \\ e_{21}^2 & e_{22}^2 & \dots & e_{2m}^2 \\ \vdots & \vdots & \ddots & \vdots \\ e_{m1}^2 & e_{m2}^2 & \dots & e_{mm}^2 \end{bmatrix}$$

相関係数行列 因子負荷量行列（共通因子） 共通因子の転置行列 因子負荷量行列（独自因子）

$$\therefore \mathbf{R} - \mathbf{E}^2 = \mathbf{W}^t \mathbf{W} \quad E \cong 0 \text{ のとき、 } \mathbf{R} \cong \mathbf{W}^t \mathbf{W} \dots (6.11)$$

$$\begin{bmatrix} 1 & r_{12} & \dots & r_{1m} \\ r_{21} & 1 & \dots & r_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ r_{m1} & r_{2m} & \dots & 1 \end{bmatrix} - \begin{bmatrix} e_{11}^2 & e_{12}^2 & \dots & e_{1m}^2 \\ e_{21}^2 & e_{22}^2 & \dots & e_{2m}^2 \\ \vdots & \vdots & \ddots & \vdots \\ e_{m1}^2 & e_{m2}^2 & \dots & e_{mm}^2 \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1m} \\ w_{21} & w_{22} & \dots & w_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ w_{m1} & w_{m2} & \dots & w_{mm} \end{bmatrix} \begin{bmatrix} w_{11} & w_{21} & \dots & w_{m1} \\ w_{12} & w_{22} & \dots & w_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ w_{12} & w_{2m} & \dots & w_{mm} \end{bmatrix}$$

\mathbf{R} \mathbf{E}^2 \mathbf{W} \mathbf{W}^t

さらに因子得点 \mathbf{F} については、回答データの基準化行列 \mathbf{Z} 、相関行列の逆行列 \mathbf{R}^{-1} 、因子負荷量 \mathbf{W} の積として得ることができる。

¹¹ 確証的因子分析における因子得点の計算法には、回帰、Bartlett 法、Anderson-Rubin 法などがある。構造方程式モデルの「因子得点」は筆者が使用した Amos19 では出力されないため豊田（2007）に従った。

$$\mathbf{Z} = \mathbf{F}'\mathbf{W}\dots (6.12)$$

$$\begin{aligned} \therefore \mathbf{F} &= \mathbf{Z}(\mathbf{W}')^{-1} = \mathbf{Z}(\mathbf{W}')^{-1} \mathbf{I} = \mathbf{Z}(\mathbf{W}')^{-1} \underbrace{(\mathbf{W}^{-1}\mathbf{W})}_{\mathbf{I}} \\ &= \mathbf{Z} \underbrace{\mathbf{W}^{-1}\mathbf{W}^{-1}}_{\text{combine}} \mathbf{W} = \mathbf{Z} \underbrace{(\mathbf{W}'\mathbf{W})^{-1}}_{\mathbf{R}} \mathbf{W} = \mathbf{Z}\mathbf{R}^{-1}\mathbf{W}\dots (6.13) \end{aligned}$$

$$\begin{bmatrix} f_{11} & f_{12} & \dots & f_{1m} \\ f_{21} & f_{22} & \dots & f_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ f_{n1} & f_{n2} & \dots & f_{nm} \end{bmatrix} = \begin{bmatrix} z_{11} & z_{12} & \dots & z_{1m} \\ z_{21} & z_{22} & \dots & z_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ z_{n1} & z_{n2} & \dots & z_{nm} \end{bmatrix} \begin{bmatrix} 1 & r_{12} & \dots & r_{1m} \\ r_{21} & 1 & \dots & r_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ r_{m1} & r_{2m} & \dots & 1 \end{bmatrix}^{-1} \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1m} \\ w_{21} & w_{22} & \dots & w_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ w_{m1} & w_{m2} & \dots & w_{mm} \end{bmatrix}$$

因子得点行列 基準化した回答行列 相関係数行列の逆行列 因子負荷量行列

3.2 散布図による考察結果

図表VI-2に示したモデル1、モデル2a、モデル2b、モデル3の構造方程式モデル内の測定モデルにおける7つの潜在変数（魅力性・信頼性・トピックス性・専門性、社会臨場感、購買意欲）の因子得点を、確証的因子分析（CFA）による因子得点と比較するため、CFAも含めアンケート回答より3,188人分の構成概念スコアを測定した。

しかし、因子得点も構成概念スコアも全回答者の潜在因子の平均は0、分散は1に標準化されたデータであるため、モデル間の測定値を比較したり、F検定やt検定、分散分析等を用いて、両者の平均値や分散の有意差を検証することは困難である。

そこで、まず視覚化による考察がしやすいように、横軸に確証的因子分析（CFA）で測定した因子得点、縦軸に構造方程式モデル内における構成概念スコアをとり、図表VI-6に示す散布図を作成するとともに、両者の相関係数も求めた。6つの潜在変数¹²について縦方向に両者の乖離状況を比較すると、原因変数（魅力性・信頼性・トピックス性・専門性）では直線状に近く相関係数も0.99以上の高い値となった。しかし、内生変数について、結果変数（購買意欲）、特に中間変数（社会臨場感）では、構造方程式モデルの構成概念スコアとCFA測定値の因子得点との乖離が大きく広がり、相関係数が0.91台にまで低下している。すなわち、多くの変数との因果関係を持ち、影響を及ぼしあう潜在変数ほど乖離が大きくなる傾向がみられる。

¹² 「弾幕利用」については紙幅の制約のため割愛するが、傾向は他の原因変数と同様である。

また、変数の種類別にモデルにおける両得点の乖離を横方向に比較すると、原因変数については、モデル3が直線状となり相関係数が1で最も乖離が小さく、中間変数ではモデル1の相関係数が0.93台で0.91台の他のモデルよりも若干乖離が小さい。すなわち、原因変数と結果変数との因果関係について、間接効果または直接効果のみの経路しかなく因果関係が少ない単純なモデルほど乖離が小さくなる傾向がみられる。

以上の分析結果からは、モデル2bにおける測定モデルの尺度得点化と、直接効果のパラメータ推定結果が真逆となった現象2および現象3との関係を特定できない。しかし、構造方程式モデルのパラメータ推定において、モデル内で潜在変数間に因果関係が構築されることによって構成概念スコアが変容するため、モデルの外で尺度得点化された因子得点とは異なるデータが処理されている可能性が考えられる。

3.3 因子得点と構成概念スコアの乖離に関する仮説

因子分析は、①観測変数とその背後に仮定される潜在変数との因果構造の解明と、②観測変数の実現値の原因となる潜在変数の測定に用いられる。②の用途において、潜在変数の因子得点行列 F は、測定尺度に採用した観測変数のデータをもとに、前に導出した関係式、 $F = ZR^{-1}W \cdots (6.13)$ により推定され、標準化された観測変数行列 Z と因子負荷量行列 W の影響を受ける。

図表VI-2（前掲）において、各構造方程式モデルを構成する測定モデルのユニットに着目すると、まず、原因変数（魅力性・信頼性・トピックス性・専門性・弾幕利用率）となる潜在変数の測定尺度は、それぞれ3個の観測変数を従えている。さらに、モデル1では「社会臨場感」へ、モデル3では「購買意欲」へ、モデル2aとモデル2bではその両方へのパスが伸びている。これらの測定モデル・ユニットにおいて、測定尺度である3つの観測変数と、因果関係をもつ潜在変数とが区別されずに、第4個目、第5個目の測定尺度として扱われている可能性も考えられる。因子分析の測定尺度にはない潜在変数が付加されることで因子得点の測定値に影響を受けていることが考えられる。「社会臨場感」については「購買意欲」の測定モデル・ユニットが付加されるため、因子スコアに強く影響を受けていることが考えられる。原因変数（5個のキャスター特性）については、付加される潜在変数を5個で共有しているため、因子得点への影響が弱かったものと考えられる。

図表VI-6（前掲）の散布図における分散値は1であり、平均共分散構造分析で求めた値とは異なるが、分散が縮小すれば平均値付近に分布が集中することになるだろう。あらためて散布図をみると、45度線の中央部分が横方向に広がっていることがわかる。すなわち、各モデルにおける潜在因子の分散が縮小したために、CFAの分散が相対的に広がったようにみえているのではないだろうか。ただし、モデル2bにおける「購買意欲」については、分散値が大きいことから相対的に縦方向に広がったと考えられる。

内生変数の分散値の変化要因は、(6.3)式における因子平均 M の配分であった。(6.12)式において、アンケート回答の観測変数行列 Z は所与なので、 M の配分に影響している可能性があるのは因子負荷量行列 W と相関係数行列の逆行列 R^{-1} に絞られる。そこで、各モデルにおける因子負荷量行列 W のデータを図表VI-7に集約した。

図表VI-6 構造方程式モデルと確証的因子分析との因子得点の乖離

	モデル 1 間接効果を仮定	モデル 2a 直接効果を仮定	モデル 2b 間接効果＋直接効果	モデル 3 潜在変数重回帰
魅力性	$r = .997^{**}$ 	$r = .996^{**}$ 	$r = .997^{**}$ 	$r = 1.000^{**}$
信頼性	$r = .996^{**}$ 	$r = .995^{**}$ 	$r = .996^{**}$ 	$r = 1.000^{**}$
トピックス性	$r = .994^{**}$ 	$r = .992^{**}$ 	$r = .994^{**}$ 	$r = .999^{**}$
専門性	$r = .995^{**}$ 	$r = .994^{**}$ 	$r = .995^{**}$ 	$r = 1.000^{**}$
社会臨場感	$r = .934^{**}$ 	$r = .918^{**}$ 	$r = .915^{**}$ 	$r = .988^{**}$
購買意欲	$r = .970^{**}$ 	$r = .972^{**}$ 	$r = .968^{**}$ 	$r = .967^{**}$

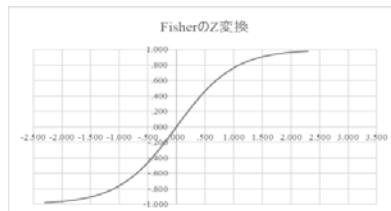
注：確証的因子分析（CFA）も含め、Amos19により原因変数間に相関をもたせて因子得点を測定した。紙幅の制約で、弾幕利用率を省略したが、他の原因変数と類似した散佈図となった。

まず、各モデルの「購買意欲」から測定尺度 Y1、Y2、Y3 への因子負荷量について確証的因子分析 (CFA) の測定値を基準として比較すると、いずれも Y1 への因子負荷量 (基準0.77) は増加し、Y2 と Y3 への因子負荷量 (基準0.84、0.74) は減少している。すなわち、構造方程式モデル内で、CFA における因子負荷量行列 \mathbf{W} が変化している可能性がある。この結果を踏まえると次式、 $\mathbf{R} \equiv \mathbf{W}'\mathbf{W} \cdots (6.4)$ より、相関係数行列 \mathbf{R} 、さらにその逆行列である \mathbf{R}^{-1} も変化していることが考えられる。

そこで、CFA における因子得点間、構造方程式モデルの構成概念スコア間の相関係数行列をそれぞれ R_0 、 R_i として、両者の相関係数行列に関する差の検定を行った。まず、各相関係数行列に対し (6.14) 式に示す Fisher の Z 変換を施すと、相関係数 r の分布が正規分布に近似される。この変換により、図表 VI-8 のように縦軸に示される相関係数 r の定義域 $[-1 \leq r \leq 1]$ が横軸に示される Z の値域 $(-\infty < Z < +\infty)$ に変換される。

$$Z = \frac{1}{2} \ln \frac{1+r}{1-r} \cdots (6.14)$$

図表 VI-8 相関係数 r に対する Fisher の Z 変換



注：縦軸 = 相関係数 r 、横軸 = Fisher の Z

一般的な (対応関係のない) 2つの相関係数 r_i と r_0 間において、(6.15) 式で定義される両者の差の検定統計量 T は、2次元正規分布に従うことが知られている。この統計検定量 T について、相関係数の差の検定を行った結果を図表 VI-9 に示す。

$$T_i = \frac{Z_i - Z_0}{\sqrt{\frac{1}{n_i - 3} + \frac{1}{n_0 - 3}}} = \frac{\frac{1}{2} \ln \frac{1+r_i}{1-r_i} - \frac{1}{2} \ln \frac{1+r_0}{1-r_0}}{\sqrt{\frac{1}{n_i - 3} + \frac{1}{n_0 - 3}}} \cdots (6.15)$$

統計的因果推論に生じた奇妙な現象ともやもや感（深瀬）

図表Ⅵ-7 各潜在変数から測定尺度への因子負荷量W（標準化パス係数）

	潜在変数	観測変数	CFA	モデル1	モデル2a	モデル2b	モデル3
原因変数（外生変数）	魅力性	A1	0.815	0.816	0.815	0.815	0.816
		A2	0.813	0.807	0.806	0.808	0.811
		A3	0.772	0.777	0.777	0.777	0.773
		購買意欲	—	—	0.140	-0.442	0.072
		社会臨場感	—	0.203	0.210	0.213	—
	信頼性	B1	0.807	0.809	0.808	0.808	0.808
		B2	0.815	0.810	0.809	0.810	0.814
		B3	0.774	0.777	0.777	0.777	0.773
		購買意欲	—	—	0.163	-0.477	0.089
		社会臨場感	—	0.225	0.231	0.233	—
	トピック性	C1	0.8	0.795	0.793	0.795	0.803
		C2	0.788	0.786	0.785	0.787	0.789
		C3	0.743	0.750	0.749	0.750	0.740
		購買意欲	—	—	0.284	-0.426	0.212
		社会臨場感	—	0.270	0.263	0.255	—
	専門性	D1	0.794	0.793	0.792	0.793	0.794
		D2	0.824	0.819	0.817	0.819	0.823
		D3	0.754	0.760	0.760	0.760	0.754
		購買意欲	—	—	0.183	-0.566	0.090
		社会臨場感	—	0.259	0.271	0.271	—
	弹幕利用率	F1	0.793	0.790	0.788	0.790	0.793
		F2	0.805	0.799	0.797	0.799	0.805
		F3	0.724	0.733	0.734	0.734	0.724
		購買意欲	—	—	0.194	-0.474	0.113
		社会臨場感	—	0.238	0.244	0.241	—
中間変数	社会臨場感	M1	0.743	0.780	0.779	0.771	0.728
		M2	0.84	0.779	0.770	0.767	0.806
		M3	0.742	0.784	0.773	0.783	0.788
		購買意欲	—	0.805	—	2.751	0.551
結果変数	購買意欲	Y1	0.769	0.834	0.828	0.839	0.796
		Y2	0.844	0.796	0.799	0.793	0.740
		Y3	0.739	0.710	0.716	0.706	0.644
残差	e22	購買意欲	—	—	0.615	0.446	—
	e23	社会臨場感	—	—	-0.124	-0.163	—

3.4 潜在変数のモデル構造適応仮説

仮説検証1では、直接効果において符号反転現象が生じたモデル2bとCFAにおける潜在因子間の相関行列を比較した。まず、双方について無相関検定(帰無仮説を無相関とする)を行ったところ、全ての相関係数が1%水準で有意であることが確認された。相関係数は全体的にモデル2bの方が高く、特に中間変数の「社会臨場感」については全ての潜在変数と0.8以上の高い相関がみられたが、CFAの相関行列にはこのような高い相関はみられなかった。

次に双方の相関行列の乖離度合いについてカイ2乗検定で検証した結果、顕著な差(p 値0.00)が認められた。さらに相関係数行列間の差の検定を用いて、双方に乖離が見られる箇所を探索し、「社会臨場感」と全ての潜在変数間の相関、「購買意欲」と全ての潜在変数間の相関に差があることが特定された。すなわち、潜在変数間の相関構造が変容していることから、その逆行列である R^{-1} も変化し、(6.6)式より因子得点行列Fに影響していることが考えられる。(潜在変数のモデル構造適応仮説)。

仮説検証2では、直接効果がプラス値となったモデル2a(間接効果なし)と直接効果がマイナス値に反転したモデル2b(間接効果あり)との差を検証した。モデル2aの相関係数はモデル2bよりも僅差ながら全体的に高めで、特に「社会臨場感」と「購買意欲」間における相関係数($2a = 0.89$ 、 $2b = 0.85$)には顕著な差がみられた。両モデルの相関行列の乖離について、カイ2乗検定の結果から有意な差が認められた。さらに相関係数行列間の差の検定結果より、具体的に差がある箇所として唯一「社会臨場感」と「購買意欲」間の相関係数が特定され、1%水準で有意な差が認められた。この有意差は両モデルの違いを決定づける間接効果の有無に関連すると考えられる。

ではなぜ、モデル2bではモデル2aよりも相関係数が低く、影響が弱いにも関わらず、直接効果の符号反転という奇妙な現象が生じたのか、以下のように考察した。モデル2bでは「社会臨場感」から「購買意欲」への因果関係において、双方に5つの原因変数が影響しており、図表IV-9(前掲)に示した分岐型経路のバックドア基準を満たさないため、これらが交絡因子となり強い疑似相関が生じていると考えられる。ここでIV章で導出した(4.17)式より、単相関は直接効果+間接効果+疑似相関に分解でき、間接効果を仮定するモデル2bでは両変数間の間接効果に疑似相関によるバイアスの一部が吸収される。数式上の単相関の和は一定だが、相関係数行列において

統計的因果推論に生じた奇妙な現象ともややや感（深瀬）

間接効果は別の変数間の相関であるため、両変数間の相関からは控除されることから、モデル 2b の相関が低下するのではないかと。

$$\therefore \underbrace{r_{24}}_{\text{単相関}} = \alpha_{41}r_{12} + \alpha_{42} + \alpha_{43}r_{23} = \underbrace{r_{42}}_{\substack{X_2 \rightarrow X_4 \\ \text{直接効果}}} + \underbrace{r_{43}r_{32}}_{\substack{X_2 \rightarrow X_3 \rightarrow X_4 \\ \text{間接効果}}} + \underbrace{r_{41}r_{21}}_{\substack{X_2 \leftarrow X_1 \rightarrow X_4 \\ \text{疑似相関}}} \dots (4.17)$$

図表 VI-9 モデル間における相関係数行列の差の検定結果

仮説検証 1：確証的因子分析（CFA）vs.モデル 2b（間接効果あり）

χ^2 値 = 2112.31
df = 21
 ρ 値 = .000

確証的因子分析（n = 3188） 相関係数行列

	魅力性	信頼性	トビック性	専門性	弾幕利用	社会臨場	購買意欲
魅力性	1.000						
信頼性	.571 **	1.000					
トビック性	.622 **	.668 **	1.000				
専門性	.650 **	.659 **	.705 **	1.000			
弾幕利用率	.629 **	.683 **	.672 **	.729 **	1.000		
社会臨場感	.686 **	.716 **	.737 **	.757 **	.745 **	1.000	
購買意欲	.526 **	.557 **	.600 **	.588 **	.583 **	.656 **	1.000

モデル2b-f(n = 3188) 相関係数行列

	魅力性	信頼性	トビック性	専門性	弾幕利用	社会臨場	購買意欲
魅力性	1.000						
信頼性	.576 **	1.000					
トビック性	.626 **	.672 **	1.000				
専門性	.654 **	.665 **	.710 **	1.000			
弾幕利用率	.633 **	.689 **	.679 **	.735 **	1.000		
社会臨場感	.800 **	.836 **	.864 **	.884 **	.870 **	1.000	
購買意欲	.662 **	.699 **	.762 **	.737 **	.734 **	.888 **	1.000

差の統計量（正規分布を仮定したZ値）

	魅力性	信頼性	トビック性	専門性	弾幕利用	社会臨場	購買意欲
魅力性	---						
信頼性	0.297	---					
トビック性	0.227	0.351	---				
専門性	0.290	0.408	0.407	---			
弾幕利用率	0.218	0.426	0.466	0.542	---		
社会臨場感	10.311 **	12.269 **	14.588 **	16.083 **	14.750 **	---	
購買意欲	8.402 **	9.465 **	12.276 **	10.706 **	10.742 **	24.982 **	---

仮説検証 2：モデル 2a（間接効果なし）vs.モデル 2b（間接効果あり）

 χ^2 値 = 41.10

df = 21

 p 値 = .005

モデル2a-f(n = 3188) 相関係数行列

	魅力性	信頼性	トビック性	専門性	弾幕利用	社会臨場	購買意欲
魅力性	1.000						
信頼性	.578 **	1.000					
トビック性	.627 **	.673 **	1.000				
専門性	.655 **	.665 **	.712 **	1.000			
弾幕利用率	.634 **	.688 **	.680 **	.736 **	1.000		
社会臨場感	.803 **	.839 **	.869 **	.889 **	.874 **	1.000	
購買意欲	.669 **	.706 **	.773 **	.747 **	.743 **	.852 **	1.000

モデル2b-f (n = 3188) 相関係数行列

	魅力性	信頼性	トビック性	専門性	弾幕利用	社会臨場	購買意欲
魅力性	1.000						
信頼性	.576 **	1.000					
トビック性	.626 **	.672 **	1.000				
専門性	.654 **	.665 **	.710 **	1.000			
弾幕利用率	.633 **	.689 **	.679 **	.735 **	1.000		
社会臨場感	.800 **	.836 **	.864 **	.884 **	.870 **	1.000	
購買意欲	.662 **	.699 **	.762 **	.737 **	.734 **	.888 **	1.000

差の統計量（正規分布を仮定したZ値）

	魅力性	信頼性	トビック性	専門性	弾幕利用	社会臨場	購買意欲
魅力性	---						
信頼性	0.077	---					
トビック性	0.087	0.038	---				
専門性	0.096	0.056	0.131	---			
弾幕利用率	0.102	0.038	0.128	0.096	---		
社会臨場感	0.382	0.405	0.767	0.906	0.746	---	
購買意欲	0.512	0.544	1.009	0.889	0.865	5.969 **	---

** $p < .01$, * $p < .05$, + $p < .10$

4. 奇妙な現象2および現象3の真相に迫る

4.1 潜在変数のモデル構造適応仮説による検証

仮説検証1の結果より、構造方程式モデルでは、2つの潜在変数である「社会臨場感」と「購買意欲」との相関が、確証的因子分析における因子得点間よりも有意に高められていたことが確認された。このことが、購買意欲にどのように影響するのか、重回帰分析を行って検証した。

結果は図表VI-10に示され、因子分析による因子得点を使用した場合、標準化偏回帰係数は全てプラス値で有意となり、説明変数間の相関を示す指標であるVIF (Value Inflation Factor) は全て4未満であり、多重共線性問題は発生していない。

一方、構造方程式モデルにおける因子スコアを使用した場合は、キャスター特性に関する標準化偏回帰係数が全てマイナス値で有意となり、これらのVIF値が6.3～10近くに高まっていた。さらに「社会臨場感」に至ってはVIF値が100を超える異常に高い値となり、極めて強い多重共線性問題が発生していることが確認された。これにより、符号の反転が生じたことが考えられる。

例えば、「魅力性」の回帰係数の符号について、(4.20)式に仮説検証1の相関係数を代入すれば、両者の符号が反転することがわかる（ただし、重回帰モデルにおいて双方向矢印(↔)を付していないため、図表VI-10の出力結果には一致しない）。

$$a = \frac{r_{xz} - r_{xy}r_{yz}}{1 - r_{xy}^2} = \begin{cases} \frac{0.526 - 0.656 \times 0.686}{1 - 0.656^2} > 0 \dots (\text{因子分析}) \\ \frac{0.662 - 0.800 \times 0.888}{1 - 0.800^2} < 0 \dots (\text{モデル2b}) \end{cases}$$

また、モデル適合度を決定係数 R^2 値で比較すると、因子分析による因子得点を使用した場合は0.477だが、構造方程式モデルにおける因子スコアを使用した場合は0.98であり、モデル適合度に格段の違いがある。

これらの結果から、構造方程式モデルでは潜在変数の測定値が、モデルの構造に適応して変容した結果（潜在変数のモデル構造適応仮説）、極めて強い多重共線性問題が発生し、標準化パス係数の推定値がマイナス値となる奇妙な現象3が観察されたことが考えられる。一方、因子分析による因子得点を使用した場合は、データは硬直的

でモデルの構造に十分に適応できていないことが幸いして、多重共線性問題の発生に至らず、標準化パス係数の推定値がプラス値に反転する現象²が観察されたということが、一連の奇妙な現象の真相ではないだろうか。

図表Ⅵ-10 使用データの違いによる重回帰分析結果の比較

①使用データ：確証的因子分析による因子得点

標準化係数		目的変数 = 購買意欲（購買意欲）		
変数名	購買意欲	95%下限	95%上限	VIF
魅力性	.055 **	0.018	0.092	2.136
信頼性	.074 **	0.034	0.113	2.442
トピック性	.167 **	0.126	0.208	2.684
専門性	.083 **	0.039	0.127	3.035
弾幕利用率	.096 **	0.054	0.139	2.882
社会臨場感	.307 **	0.259	0.356	3.705
R^2	.477 **			

②使用データ：モデル2bの構造方程式モデルにおける因子スコア

標準化係数		目的変数 = 購買意欲（購買意欲）		
変数名	購買意欲	95%下限	95%上限	VIF
魅力性	-.940 **	-0.953	-0.928	6.356
信頼性	-1.027 **	-1.041	-1.014	7.595
トピック性	-1.022 **	-1.036	-1.007	9.211
専門性	-1.200 **	-1.216	-1.185	9.990
弾幕利用率	-1.035 **	-1.049	-1.020	8.469
社会臨場感	5.342 **	5.291	5.392	106.681
R^2	.980 **			

** P<.01, * p<.05、+p<.10

4.2 潜在変数のモデル構造適応の詳細

最後に、図表Ⅵ-11より、原データにおける観測変数間の相関が、構造方程式モデル2bにおいて、どのように変化したのか、「購買意欲」（Y 群）を中心に比較すると、M 群と Y 群における差は、A～F 群に比べて大きく、特に M3 と Y1 との相関が原データの0.7と比べて大きく低下している。とりわけ、キャストの特性→「社会臨場

感」→「購買意欲」の間接効果において、後半の経路が寸断されたモデル 2 a では0.50、前半の経路が寸断されたモデル 3 では0.35に半減している。

また、「購買意欲」（Y 群）内の測定尺度間の相関についても、モデル 3 における Y 2 と Y 3 との相関では原データの0.62から0.48に低下している。

すなわち、潜在変数のモデル構造適応仮説の前段階として、多くのパラメータを擁する構造方程式モデルでは、モデルの形状に適用して観測変数間の相関 R が調整されたがって、その逆行列である R^{-1} も変化していることになる。その結果として、分析に使用される潜在変数の因子得点に変容するというメカニズムが解明された。

図表 VI-11 観測変数間の相関係数行列

潜在変数	観測変数	相関係数行列											
		標本の共分散			モデル 1			モデル 2a			モデル 2b		
		購買意欲			購買意欲			購買意欲			購買意欲		
		Y1	Y2	Y3	Y1	Y2	Y3	Y1	Y2	Y3	Y1	Y2	Y3
魅力生	A1	0.42	0.37	0.33	0.41	0.39	0.35	0.40	0.38	0.34	0.40	0.37	0.33
	A2	0.40	0.31	0.27	0.40	0.38	0.34	0.39	0.38	0.34	0.39	0.37	0.33
	A3	0.41	0.37	0.33	0.39	0.37	0.33	0.38	0.36	0.33	0.38	0.36	0.32
信頼性	B1	0.45	0.39	0.34	0.42	0.40	0.36	0.42	0.40	0.36	0.41	0.39	0.35
	B2	0.42	0.36	0.33	0.42	0.40	0.36	0.42	0.40	0.36	0.42	0.39	0.35
	B3	0.41	0.37	0.32	0.40	0.39	0.35	0.40	0.38	0.34	0.40	0.38	0.34
トビック	C1	0.46	0.41	0.39	0.43	0.41	0.37	0.45	0.43	0.39	0.45	0.42	0.38
	C2	0.43	0.39	0.38	0.43	0.41	0.36	0.44	0.43	0.38	0.44	0.42	0.37
	C3	0.43	0.39	0.34	0.41	0.39	0.35	0.42	0.41	0.36	0.42	0.40	0.35
専門性	D1	0.44	0.40	0.35	0.44	0.42	0.37	0.43	0.42	0.37	0.43	0.41	0.36
	D2	0.45	0.40	0.35	0.45	0.43	0.39	0.44	0.43	0.38	0.44	0.42	0.37
	D3	0.43	0.40	0.35	0.42	0.40	0.36	0.41	0.40	0.36	0.41	0.39	0.35
弾幕利用	F1	0.43	0.39	0.37	0.43	0.41	0.36	0.42	0.41	0.37	0.42	0.40	0.36
	F2	0.43	0.38	0.35	0.43	0.41	0.37	0.43	0.41	0.37	0.43	0.40	0.36
	F3	0.41	0.39	0.35	0.40	0.38	0.34	0.40	0.38	0.34	0.39	0.37	0.33
社会臨場感	M1	0.48	0.47	0.44	0.52	0.50	0.45	0.50	0.49	0.44	0.54	0.51	0.46
	M2	0.55	0.49	0.38	0.52	0.50	0.45	0.50	0.48	0.43	0.54	0.51	0.45
	M3	0.70	0.52	0.40	0.53	0.50	0.45	0.50	0.48	0.43	0.55	0.52	0.46
購買意欲	Y1	1.00	0.65	0.57	1.00	0.66	0.59	1.00	0.66	0.59	1.00	0.67	0.59
	Y2		1.00	0.62		1.00	0.57		1.00	0.57		1.00	0.56
	Y3			1.00			1.00			1.00			1.00

VII 結びに代えて

本研究では、学生の構造方程式モデル分析において生じた、①モデル構造および②尺度得点化に起因する奇妙な現象の解明を目指すとともに、③統計的因果推論の誤用を防止するための安全点検も行った。

①モデル構造に関する現象については、間接効果のモデルに直接効果のパスを追加したことにより生じたものであり、ライブコマースにおけるキャスターの特性が参加者の購買意欲を抑制するという、実質科学に反する分析結果であった。社会臨場感の醸成によりライブコマースに夢中になるあまり購買意欲が抑制されるブラウジング効果の可能性も視野にいて検証を行った。このメカニズムに関する結論として、直接効果が付加されたことにより分岐経路が構成されたために、原因変数が交絡因子となり、変数間の強力なバイアスによって生じた現象であることが解明された。すなわち、間接効果が過剰に増強された一方で、直接効果が過大に減衰してマイナス値に転じたことによるものであった。

この現象は、J.Peael (1995) などにより30年近く前から指摘されてきた擬似相関のテーマの延長線上におかれた、推定バイアスの極めて強いケースであった。学生を指導する職にありながら即時に気付かなかった点是不注意極まりなく誹りも免れない。「構造モデル」において、バックドアを閉じる手法については、傾向スコアマッチング法、操作変数法、差分の差分等々の手法が開発されており、活用していきたい。

しかし、今もなお媒介効果の分析等は、実証分析のテキスト、学会報告、ジャーナル誌等でも頻繁に扱われており、気付かないまま不適切な分析結果が報告されている危険性も否定できない。ただし、発生するバイアスが相対的に小さい場合はそのまま使用することも許容されるべきであろう。最大の問題は推定結果の適否を判断しにくいことである。

このテーマについては、2024年11月に愛媛大学で開催された「統計的因果推論の最先端」研究集会において、黒木学氏（横国大）・吉田悠夏氏（黒木研究室）による感応度に関連する応用研究が報告されたので、今後の課題として早急にキャッチアップしていきたい。

②尺度得点化において発生した現象は、構造方程式モデルにおいて直接効果のパス係数がマイナス値となった現象を単純化して解明するために、尺度得点化してパス解

析モデルに作り直したところパス係数がプラス値に転じ、構造方程式モデルとパス解析モデルによる分析結果が真逆になったことである。

尺度得点化したデータの二次使用については、狩野裕（2002）をはじめ小杉（2018）など既に20年以上も前から批判されているが、その根拠は「データの希薄化」や「データの使い回し」など、潜在変数の測定尺度の信頼性、因子分析における平均情報の損失、連鎖計算プロセスにおける誤差の累積など、データ劣化に起因するものであった。

これらの先行研究に対して、本研究では、下位尺度得点、因子得点を用いた平均構造分析、数理モデルによる考察から、「潜在変数のモデル構造適応仮説」を導き、「データの硬直性と柔軟性」から尺度得点化により生じた現象を解明した。すなわち、多数の観測変数と推定パラメータを擁する構造方程式モデルでは自由度が高く、パラメータを細かく調整して分散を配分できるので、モデル構造に応じて柔軟に潜在変数を変容させることができる。この結果、「購買意欲」を結果変数とする重回帰分析モデルにおいて、原因変数であるキャストの特性と中間変数である「社会臨場感」との間に極めて強い多重共線性問題が発生していたために、パス係数の推定値がマイナス値となったと考えられる。

一方、尺度得点化したパス解析モデルでは、パラメータが少なく自由度も低いため潜在変数が「硬直化」し、解像度が低下することから交絡因子により生じたバイアスに対して十分なパラメータ調整ができなくなる可能性がある。このことが幸いして、多重共線性問題が発生するまでには至らなかったのである。このことは、構造方程式モデルにおける測定モデル素子を尺度得点に置換することで、重回帰分析においても適合度指標の数値が悪化した分析結果とも整合しており、「潜在変数のモデル構造適応仮説」を支持する分析結果となった。

以上が本研究によって得られた知見である。

③の安全点検について、本稿では統計的因果推論の誤用を防止するために、ソフトウェアによる基本的な分析結果について、簡単なトイ・モデルを用いて計算過程を確認して安全確認に努めた。高度な統計分析ができるソフトウェアを手軽に入手できるようになり、文化系の学生も基礎的な理論を学ばずに使用する機会も広がった。また、プログラミングでは、理論的な検討がなされないままフローチャートによる連鎖的

なデータ処理が行われている。特にビッグデータを扱う場合には、「測定モデル」や「次元削減」(Dimensionality reduction)により尺度得点化されたデータを2次、3次、多段階に利用して連鎖処理を行う大規模なシステムが構築されることもあるだろう。

しかし、本稿のように分析結果が実質科学とは真逆になるケースも確実に実在し、理論を無視してデータサイエンスによる分析結果を過信することは非常に危険である。ただし、有用な分析手法を封印してしまうのではなく、ふぐ料理のように毒を除去して美味を賞味する道を探索するべきである。

そのためには、連鎖処理において生じるバイアスの程度を推定して、処理の妥当性を判断する指標が欲しいが、未だ記されているテキストが見当たらない。今後の課題として、重回帰分析における多重共線性問題を判定する VIF(Value Inflation Factor) 値のように、文科系学生にも簡単に扱える指標を開発したい。

Appendix 1 重回帰分析の偏回帰係数の導出

【重回帰モデル】

$$z_i = ax_i - by_i - c + \varepsilon_i \cdots (A.1)$$

ただし、 z : 結果変数、 x, y : 原因変数、 a, b, c : パラメータ、 ε : 残差項とする。

(A.1) 式より残差平方和を、パラメータ a, b, c を変数とする次式で示す。

$$L(a, b, c) = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (z_i - ax_i - by_i - c)^2 \cdots (A.2)$$

(A.2) 式を最小化するように、パラメータ a, b, c を定める。最小化の一階条件より

$$\frac{\partial L}{\partial c} = \frac{\partial}{\partial c} \sum_{i=1}^n (z_i - ax_i - by_i - c)^2 = -2 \sum_{i=1}^n (z_i - ax_i - by_i - c) = 0$$

$$\sum_{i=1}^n z_i - a \sum_{i=1}^n x_i - b \sum_{i=1}^n y_i - nc = 0$$

$$\therefore c = \frac{\sum_{i=1}^n z_i}{n} - a \frac{\sum_{i=1}^n x_i}{n} - b \frac{\sum_{i=1}^n y_i}{n} = \bar{z} - a\bar{x} - b\bar{y} \cdots (A.3)$$

(A.2) 式の C に、(A.3) 式を代入して C を消去すると、

$$L(a, b) = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n \left((z_i - \bar{z}) - a(x_i - \bar{x}) - b(y_i - \bar{y}) \right)^2 \dots (A.4)$$

ここで、重回帰式を残差について解いた (A.4) 式の括弧内の式を、下記の 2 次元行列を用いて行列表示すると、(A.5) 式になる。

$$\mathbf{e} = \mathbf{Z} - \mathbf{X}\boldsymbol{\beta} \dots (A.5)$$

$$\text{ただし、} \mathbf{Z} = \begin{bmatrix} z_1 - \bar{z} \\ z_2 - \bar{z} \\ \vdots \\ z_n - \bar{z} \end{bmatrix}, \mathbf{X} = \begin{bmatrix} x_1 - \bar{x} & y_1 - \bar{y} \\ x_2 - \bar{x} & y_2 - \bar{y} \\ \vdots & \vdots \\ x_n - \bar{x} & y_n - \bar{y} \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} a \\ b \end{bmatrix}, \mathbf{e} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

よって、残差平方和に関する (A.4) 式は次式となり、(A.5) 式を代入すれば (A.6) 式を得る。

$$\begin{aligned} L(a, b) &= \sum_{i=1}^n \varepsilon_i^2 = \begin{bmatrix} \varepsilon_1 & \varepsilon_2 & \dots & \varepsilon_n \end{bmatrix} \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} = {}^t \mathbf{e} \mathbf{e} \\ &= {}^t (\mathbf{Z} - \mathbf{X}\boldsymbol{\beta}) (\mathbf{Z} - \mathbf{X}\boldsymbol{\beta}) = ({}^t \mathbf{Z} - {}^t \boldsymbol{\beta}^t \mathbf{X}) (\mathbf{Z} - \mathbf{X}\boldsymbol{\beta}) \\ &= {}^t \mathbf{Z} \mathbf{Z} - {}^t \mathbf{Z} \mathbf{X} \boldsymbol{\beta} - \underbrace{{}^t \boldsymbol{\beta}^t \mathbf{X} \mathbf{Z}}_{= {}^t \mathbf{Z} \mathbf{X} \boldsymbol{\beta} \dots *} + {}^t \boldsymbol{\beta}^t \mathbf{X} \mathbf{X} \boldsymbol{\beta} \dots (A.6) \end{aligned}$$

注* (A.6) 式の第 3 項における ${}^t \boldsymbol{\beta}^t \mathbf{X} \mathbf{Z} = {}^t \mathbf{Z} \mathbf{X} \boldsymbol{\beta}$ となる計算過程について、
次式の第 3 項は 2 次形式（横ベクトル・行列・縦ベクトルの積）になっている。

$${}^t \boldsymbol{\beta}^t \mathbf{X} \mathbf{Z} = {}^t (\mathbf{X}\boldsymbol{\beta}) \mathbf{Z} = ({}^t \mathbf{Z} \mathbf{X} \boldsymbol{\beta}) = {}^t \mathbf{Z} \mathbf{X} \boldsymbol{\beta}$$

一般の行列ではその転置行列と一致しないが、この場合は答えがスカラーの次元になるため、転置を施しても値が変わらない。

$$\therefore L(a, b) = {}^t \mathbf{Z} \mathbf{Z} - 2 {}^t \mathbf{Z} \mathbf{X} \boldsymbol{\beta} + {}^t \boldsymbol{\beta}^t \mathbf{X} \mathbf{X} \boldsymbol{\beta} \dots (A.7)$$

(A.2) 式の次元を下げた、残差平方和 (A.7) 式の最小化問題について、 β で偏微分して、得られた正規方程式を解くと、

$$\begin{aligned}
 \frac{\partial L(a,b)}{\partial \beta} &= \left[\frac{\partial L(a,b)}{\partial a} \right. \\
 &\quad \left. \frac{\partial L(a,b)}{\partial b} \right] = 0 \\
 \frac{\partial L(a,b)}{\partial \beta} &= \frac{\partial}{\partial \beta} ({}^t\mathbf{Z}\mathbf{Z} - 2{}^t\mathbf{Z}\mathbf{X}\beta + {}^t\beta{}^t\mathbf{X}\mathbf{X}\beta) \\
 &= -2{}^t\mathbf{Z}\mathbf{X} + {}^t\mathbf{X}\mathbf{X}\beta + \underbrace{{}^t\beta{}^t\mathbf{X}\mathbf{X}}_{{}^t({}^t\mathbf{X}\mathbf{X}\beta)={}^t\mathbf{X}\mathbf{X}\beta} \\
 &= -2{}^t\mathbf{Z}\mathbf{X} + 2{}^t\mathbf{X}\mathbf{X}\beta = 0 \\
 \therefore {}^t\mathbf{X}\mathbf{X}\beta &= {}^t\mathbf{Z}\mathbf{X} \\
 \therefore \beta &= ({}^t\mathbf{X}\mathbf{X})^{-1} {}^t\mathbf{Z}\mathbf{X} \dots (A.8)
 \end{aligned}$$

(A.8) 式において、先にパーツの項を計算する。

$$\begin{aligned}
 {}^t\mathbf{X}\mathbf{X} &= \begin{bmatrix} x_1 - \bar{x} & x_2 - \bar{x} & \cdots & x_n - \bar{x} \\ y_1 - \bar{y} & y_2 - \bar{y} & \cdots & y_n - \bar{y} \end{bmatrix} \begin{bmatrix} x_1 - \bar{x} & y_1 - \bar{y} \\ x_2 - \bar{x} & y_2 - \bar{y} \\ \vdots & \vdots \\ x_n - \bar{x} & y_n - \bar{y} \end{bmatrix} \\
 &= \begin{bmatrix} \sum_{i=1}^n (x_i - \bar{x})^2 & \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \\ \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) & \sum_{i=1}^n (y_i - \bar{y})^2 \end{bmatrix} = \begin{bmatrix} S_{xx} & S_{xy} \\ S_{xy} & S_{yy} \end{bmatrix} \\
 \therefore ({}^t\mathbf{X}\mathbf{X})^{-1} &= \begin{bmatrix} S_{xx} & S_{xy} \\ S_{xy} & S_{yy} \end{bmatrix}^{-1} = \frac{1}{S_{xx}S_{yy} - S_{xy}^2} \begin{bmatrix} S_{yy} & -S_{xy} \\ -S_{xy} & S_{xx} \end{bmatrix} \\
 {}^t\mathbf{X}\mathbf{Z} &= \begin{bmatrix} x_1 - \bar{x} & x_2 - \bar{x} & \cdots & x_n - \bar{x} \\ y_1 - \bar{y} & y_2 - \bar{y} & \cdots & y_n - \bar{y} \end{bmatrix} \begin{bmatrix} z_1 - \bar{z} \\ z_2 - \bar{z} \\ \vdots \\ z_n - \bar{z} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n (x_i - \bar{x})(z_i - \bar{z}) \\ \sum_{i=1}^n (y_i - \bar{y})(z_i - \bar{z}) \end{bmatrix} = \begin{bmatrix} S_{xz} \\ S_{yz} \end{bmatrix}
 \end{aligned}$$

統計的因果推論に生じた奇妙な現象ともややや感（深瀬）

よって β は (A.8) 式より、

$$\begin{aligned}
 \beta &= \begin{bmatrix} a \\ b \end{bmatrix} = (\text{'XX})^{-1} \text{'XZ} = \underbrace{\frac{1}{S_{xx}S_{yy} - S_{xy}^2} \begin{bmatrix} S_{yy} & -S_{xy} \\ -S_{xy} & S_{xx} \end{bmatrix}}_{(\text{'XX})^{-1}} \underbrace{\begin{bmatrix} S_{xz} \\ S_{yz} \end{bmatrix}}_{\text{'XZ}} \\
 &= \frac{1}{S_{xx}S_{yy} - S_{xy}^2} \begin{bmatrix} S_{yy}S_{xz} - S_{xy}S_{yz} \\ -S_{xy}S_{xz} + S_{xx}S_{yz} \end{bmatrix} = \begin{bmatrix} \frac{S_{yy}S_{xz} - S_{xy}S_{yz}}{S_{xx}S_{yy} - S_{xy}^2} \\ \frac{-S_{xy}S_{xz} + S_{xx}S_{yz}}{S_{xx}S_{yy} - S_{xy}^2} \end{bmatrix} \\
 &= \begin{bmatrix} \frac{\frac{S_{yy}S_{xz}}{S_{xx}S_{yy}} - \frac{S_{xy}S_{yz}}{S_{xx}S_{yy}}}{1 - \frac{S_{xy}^2}{S_{xx}S_{yy}}} \\ \frac{\frac{-S_{xy}S_{xz}}{S_{xx}S_{yy}} + \frac{S_{xx}S_{yz}}{S_{xx}S_{yy}}}{1 - \frac{S_{xy}^2}{S_{xx}S_{yy}}} \end{bmatrix} = \begin{bmatrix} \frac{\frac{S_{xz}}{S_{xx}} - \frac{S_{xy}S_{yz}}{S_{xx}S_{yy}}}{1 - r_{xy}^2} \\ \frac{\frac{-S_{xy}S_{xz}}{S_{xx}S_{yy}} + \frac{S_{yz}}{S_{xx}S_{yy}}}{1 - r_{xy}^2} \end{bmatrix} = \begin{bmatrix} \frac{\left(\frac{S_{xz}}{\sqrt{S_{xx}}\sqrt{S_{yy}}} - \left(\frac{S_{xy}}{\sqrt{S_{xx}}\sqrt{S_{yy}}} \right) \left(\frac{S_{yz}}{\sqrt{S_{xx}}\sqrt{S_{yy}}} \right) \right) \frac{\sqrt{S_{xz}}}{\sqrt{S_{xx}}}}{1 - r_{xy}^2} \\ \frac{\left(\frac{-S_{xy}S_{xz}}{\sqrt{S_{xx}}\sqrt{S_{yy}}\sqrt{S_{xx}}\sqrt{S_{yy}}} + \frac{S_{yz}}{\sqrt{S_{yy}}\sqrt{S_{xx}}} \right) \frac{\sqrt{S_{xz}}}{\sqrt{S_{yy}}}}{1 - r_{xy}^2} \end{bmatrix} = \begin{bmatrix} \frac{r_{xz} - r_{xy}r_{yz}}{1 - r_{xy}^2} \times \frac{\sqrt{S_{xz}}}{\sqrt{S_{xx}}} \\ \frac{r_{yz} - r_{xy}r_{xz}}{1 - r_{xy}^2} \times \frac{\sqrt{S_{xz}}}{\sqrt{S_{yy}}} \end{bmatrix}
 \end{aligned}$$

変数 Z 、 X 、 Y が標準化されている場合は、標準化偏回帰係数は次式となり、また、切片項は (A.3) 式において、 $\bar{z} = \bar{x} = \bar{y} = 0$ より $c = 0$ となる。

$$\beta = \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} \frac{r_{xz} - r_{xy}r_{yz}}{1 - r_{xy}^2} \times \frac{\sqrt{S_{xz}}}{\sqrt{S_{xx}}} \\ \frac{r_{yz} - r_{xy}r_{xz}}{1 - r_{xy}^2} \times \frac{\sqrt{S_{xz}}}{\sqrt{S_{yy}}} \\ \bar{z} - a\bar{x} - b\bar{y} \end{bmatrix} = \begin{bmatrix} \frac{r_{xz} - r_{xy}r_{yz}}{1 - r_{xy}^2} \\ \frac{r_{yz} - r_{xy}r_{xz}}{1 - r_{xy}^2} \\ 0 \end{bmatrix} \dots (A.9)$$

下記に示す一般の重回帰モデルについて、

$$z = a_1x_1 + a_2x_2 + \dots + a_nx_n + c \dots (A.10)$$

標準化偏回帰係数と切片は、(A.11) 式となる。

$$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} S_{11} & S_{12} & \cdots & S_{1n} \\ S_{21} & S_{22} & \cdots & S_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ S_{n1} & S_{n2} & \cdots & S_{nn} \end{pmatrix} \begin{pmatrix} S_{1y} \\ S_{2y} \\ \vdots \\ S_{ny} \end{pmatrix} \cdots (A.11)$$

$$c = \bar{z} - a_1 \bar{x}_1 - a_2 \bar{x}_2 - \cdots - a_n \bar{x}_n$$

ただし、

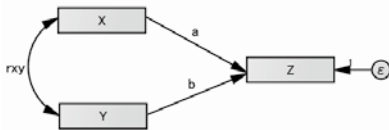
$$S_{ii} = \sum_{k=1}^n (x_{ik} - \bar{x}_i)^2, \quad S_{ij} = \sum_{k=1}^n (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j), \quad S_{iy} = \sum_{k=1}^n (x_{ik} - \bar{x}_i)(y_k - \bar{y})$$

Appendix2 偏回帰係数と偏相関係数の関係

図 A.1のような (A.12) 式で示される重回帰モデルにおいて、説明変数間に相関をもつ場合について、標準化偏回帰係数における影響を考察する。

$$z = ax + by + c \cdots (A.12)$$

図 A.1 説明変数間に相関をもつ重回帰モデル



x の影響を除いた z と、 x の影響を除いた y との偏相関係数 $r_{zy \cdot x}$ は次式となる。

$$r_{zy \cdot x} = \frac{r_{zy} - r_{xy} r_{zx}}{\sqrt{1 - r_{xy}^2} \sqrt{1 - r_{zx}^2}} \cdots (A.13)$$

以下では、(A.13) 式を導出する。

(A.12) 式は、必ず重心を通過することから、

統計的因果推論に生じた奇妙な現象ともやや感（深瀬）

$$\bar{z} = a\bar{x} + b\bar{y} + c \dots (A.14)$$

$$(A.12) - (A.14) \Rightarrow z = a(x - \bar{x}) + b(y - \bar{y}) + \bar{z}$$

$$z = \frac{S_{xz}}{S_{xx}}(x - \bar{x}) + \frac{S_{yz}}{S_{yy}}(y - \bar{y}) + \bar{z} \dots (A.15)$$

(A.15) 式において、 x を一定の値に固定したとき、 x の影響を除いた z の残差を $e_{z \cdot x}$ 、 x の影響を除いた y の残差を $e_{y \cdot x}$ とし、それぞれについて期待値をとると、

$$\begin{cases} e_{z \cdot x} = z - \bar{z} - \frac{S_{zx}}{S_{xx}}(x - \bar{x}) \\ e_{y \cdot x} = y - \bar{y} - \frac{S_{yx}}{S_{xx}}(x - \bar{x}) \end{cases} \dots (A.16) \Rightarrow \begin{cases} E[e_{z \cdot x}] = E\left[\underbrace{z - \bar{z}}_{=0} - \frac{S_{zx}}{S_{xx}}\left(\underbrace{x - \bar{x}}_{=0}\right)\right] = 0 \\ E[e_{y \cdot x}] = E\left[\underbrace{y - \bar{y}}_{=0} - \frac{S_{yx}}{S_{xx}}\left(\underbrace{x - \bar{x}}_{=0}\right)\right] = 0 \end{cases} \dots (A.17)$$

偏相関係数は、次式で定義される。

$$r_{zy \cdot x} = \frac{S_{e_{z \cdot x} e_{y \cdot x}}}{\sqrt{S_{e_{z \cdot x}}} \sqrt{S_{e_{y \cdot x}}}} = \frac{Cov[e_{z \cdot x}, e_{y \cdot x}]}{\sqrt{E[e_{z \cdot x}^2]} \sqrt{E[e_{y \cdot x}^2]}} \dots (A.18)$$

$$\text{ただし、} \begin{cases} Cov[e_{z \cdot x}, e_{y \cdot x}] = E[(e_{z \cdot x} - E[e_{z \cdot x}])(e_{y \cdot x} - E[e_{y \cdot x}])] = E[e_{z \cdot x} e_{y \cdot x}] \\ S_{e_{z \cdot x}} = Var[e_{z \cdot x}] = E[(e_{z \cdot x} - E[e_{z \cdot x}])^2] = E[e_{z \cdot x}^2] \\ S_{e_{y \cdot x}} = Var[e_{y \cdot x}] = E[(e_{y \cdot x} - E[e_{y \cdot x}])^2] = E[e_{y \cdot x}^2] \end{cases}$$

以下では、(A.18) 式のパーツを計算する。

分子について、

$$S_{e_{z \cdot x} e_{y \cdot x}} = \frac{1}{n} \sum_{i=1}^n \left(e_{zi \cdot x} - \underbrace{\bar{e}_{zi \cdot x}}_0 \right) \left(e_{yi \cdot x} - \underbrace{\bar{e}_{yi \cdot x}}_0 \right) = \frac{1}{n} \sum_{i=1}^n e_{zi \cdot x} e_{yi \cdot x} = E[e_{z \cdot x} e_{y \cdot x}] \dots (A.19)$$

(A.19) 式に (A.16) 式を代入する

$$\begin{aligned}
 S_{e_{z \cdot x}, e_{y \cdot x}} &= E[e_{z \cdot x}, e_{y \cdot x}] = E\left[(z - \bar{z}) - \frac{S_{zx}}{S_{xx}}(x - \bar{x}), (y_i - \bar{y}) - \frac{S_{yx}}{S_{xx}}(x - \bar{x})\right] \\
 &= \frac{1}{n} \sum_{i=1}^n \left[(z_i - \bar{z}) - \frac{S_{zx}}{S_{xx}}(x_i - \bar{x}) \right] \left[(y_i - \bar{y}) - \frac{S_{yx}}{S_{xx}}(x - \bar{x}) \right] \\
 &= E\left[(z - \bar{z})(y_i - \bar{y}) - \frac{S_{zx}}{S_{xx}}(x - \bar{x})(y_i - \bar{y}) - \frac{S_{yx}}{S_{xx}}(x - \bar{x})(z - \bar{z}) + \frac{S_{zx}}{S_{xx}} \frac{S_{yx}}{S_{xx}}(x - \bar{x})(x - \bar{x})\right] \\
 &= E\left[(z - \bar{z})(y_i - \bar{y})\right] - \frac{S_{zx}}{S_{xx}} E[(x - \bar{x})(y_i - \bar{y})] - \frac{S_{yx}}{S_{xx}} E[(x - \bar{x})(z - \bar{z})] + \frac{S_{zx}}{S_{xx}} \frac{S_{yx}}{S_{xx}} E[(x - \bar{x})^2] \\
 &= S_{zy} - \frac{S_{zx} S_{xy}}{S_{xx}} - \frac{S_{zx} S_{yx}}{S_{xx}} + \frac{S_{zx} S_{yx}}{S_{xx}} = S_{zy} - \frac{S_{zx} S_{xy}}{S_{xx}} \\
 &= \left(\frac{S_{zy}}{\sqrt{S_{zz}} \sqrt{S_{yy}}} - \frac{S_{zx}}{\sqrt{S_{zz}} \sqrt{S_{xx}}} \frac{S_{xy}}{\sqrt{S_{xx}} \sqrt{S_{yy}}} \right) \sqrt{S_{zz} S_{yy}} \\
 &= (r_{zy} - r_{zx} r_{xy}) \sqrt{S_{zz} S_{yy}} \\
 \therefore E[e_{z \cdot x}, e_{y \cdot x}] &= (r_{zy} - r_{zx} r_{xy}) \sqrt{S_{zz} S_{yy}} \quad (A.20)
 \end{aligned}$$

(A.13) 式の分母について、

$$\begin{aligned}
 S_{e_{z \cdot x}} &= E[e_{z \cdot x}^2] = E\left[(z - \bar{z})^2 - 2 \frac{S_{zx}}{S_{xx}}(z - \bar{z})(x - \bar{x}) + \left(\frac{S_{zx}}{S_{xx}}\right)^2 (x - \bar{x})^2\right] \\
 &= E[(z - \bar{z})^2] - 2 \frac{S_{zx}}{S_{xx}} E[(z - \bar{z})(x - \bar{x})] + \left(\frac{S_{zx}}{S_{xx}}\right)^2 E[(x - \bar{x})^2] \\
 &= S_{zz} - 2 \frac{S_{zx}^2}{S_{xx}} + \left(\frac{S_{zx}}{S_{xx}}\right)^2 S_{xx} = S_{zz} - \frac{S_{zx}^2}{S_{xx}} = S_{zz} \left(1 - \frac{S_{zx}^2}{S_{zz} S_{xx}}\right) \\
 \therefore S_{e_{z \cdot x}} &= E[e_{z \cdot x}^2] = S_{zz} (1 - r_{zx}^2) \quad (A.21)
 \end{aligned}$$

$$\begin{aligned}
 S_{e_{y \cdot x}} &= E[e_{y \cdot x}^2] = S_{yy} - 2 \frac{S_{yx}^2}{S_{xx}} + \left(\frac{S_{yx}}{S_{xx}}\right)^2 \\
 &= S_{yy} \left(1 - \frac{S_{xy}^2}{S_{xx} S_{yy}}\right) = S_{yy} (1 - r_{xy}^2) \quad (A.22)
 \end{aligned}$$

(A.18) 式に (A.20) 式、(A.21) 式、(A.22) 式を代入すると、(A.13) 式が導出される。

$$\begin{aligned}
 r_{zy \cdot x} &= \frac{S_{e_{z \cdot x} e_{y \cdot x}}}{\sqrt{S_{e_{z \cdot x}}} \sqrt{S_{e_{y \cdot x}}}} = \frac{Cov[e_{z \cdot x}, e_{y \cdot x}]}{\sqrt{E[e_{z \cdot x}^2]} \sqrt{E[e_{y \cdot x}^2]}} \\
 &= \frac{(r_{zy} - r_{zx} r_{xy}) \sqrt{S_{zz} S_{yy}}}{\sqrt{S_{zz} (1 - r_{zx}^2)} \sqrt{S_{yy} (1 - r_{xy}^2)}} = \frac{r_{zy} - r_{zx} r_{xy}}{\sqrt{1 - r_{zx}^2} \sqrt{1 - r_{xy}^2}} \frac{\sqrt{S_{zz} S_{yy}}}{\sqrt{S_{zz}} \sqrt{S_{yy}}} \\
 &= \frac{r_{zy} - r_{zx} r_{xy}}{\sqrt{1 - r_{zx}^2} \sqrt{1 - r_{xy}^2}} \dots (A.13)
 \end{aligned}$$

(A.13) 式の偏相関係数と、(A.9) 式で導出した標準化偏回帰係数と比較すると、

$$\begin{aligned}
 b &= \frac{r_{yz} - r_{xy} r_{xz}}{1 - r_{xy}^2} = \underbrace{\frac{r_{zy} - r_{zx} r_{xy}}{\sqrt{1 - r_{zx}^2} \sqrt{1 - r_{xy}^2}}}_{r_{zy \cdot x}} \times \frac{\sqrt{1 - r_{zx}^2}}{\sqrt{1 - r_{xy}^2}} = r_{zy \cdot x} \times \frac{\sqrt{1 - r_{zx}^2}}{\sqrt{1 - r_{xy}^2}} \dots (A.23) \\
 \Leftrightarrow r_{zy \cdot x} &= b \times \frac{\sqrt{1 - r_{xy}^2}}{\sqrt{1 - r_{zx}^2}}
 \end{aligned}$$

(A.9) 式より、 x が目的変数 z と説明変数 y に等しく影響するとき、すなわち、 $r_{zx} = r_{yx}$ のとき、標準化偏相関係数 b は偏相関係数 $r_{zy \cdot x}$ に一致することがわかる。

Appendix3 因子分析における因子得点の計測

1. 因子得点の測定の手順

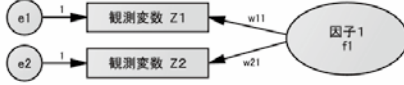
因子分析における因子得点は以下の手順で計測される。

- (1) 基準化した回答データ行列 \mathbf{Z} を相関係数行列 \mathbf{R} に変換する。
- (2) 相関係数行列 \mathbf{R} を固有値（スペクトル）分解して因子負荷量ベクトル \mathbf{W} を求める。
- (3) 因子負荷量行列、相関係数行列の逆行列 \mathbf{R}^{-1} 、基準化した回答データ行列 \mathbf{Z} より、因子得点行列 \mathbf{F} を求める。

2. 因子負荷量行列 \mathbf{W} の導出

(1) 1 因子モデルの因子負荷量ベクトル \mathbf{W} 導出

図 A.2 2 観測変数の 1 因子モデル



標準化された観測変数行列 \mathbf{Z} に関する構造方程式モデルは次式で与えられる。

$$\begin{cases} Z_1 = w_{11}f_1 + e_1 \\ Z_2 = w_{21}f_1 + e_2 \end{cases}$$

$$\mathbf{Z} = \begin{bmatrix} \mathbf{Z}_1 & \mathbf{Z}_2 \end{bmatrix} = \begin{bmatrix} z_{11} & z_{12} \\ z_{21} & z_{22} \\ \vdots & \vdots \\ z_{n1} & z_{n2} \end{bmatrix}, \quad \mathbf{f}_1 = \begin{bmatrix} f_{11} \\ f_{21} \\ \vdots \\ f_{n1} \end{bmatrix}, \quad \mathbf{E} = \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 \end{bmatrix} = \begin{bmatrix} e_{11} & e_{11} \\ e_{21} & e_{22} \\ \vdots & \vdots \\ e_{n1} & e_{n2} \end{bmatrix}$$

i : オブザベーション ($i=1, 2, \dots, n$)

観測変数行列 \mathbf{Z} の分散共分散行列を \mathbf{R} とする。

$$\mathbf{R} = \begin{bmatrix} \text{Var}(Z_1) & \text{Cov}(Z_1, Z_2) \\ \text{Cov}(Z_2, Z_1) & \text{Var}(Z_2) \end{bmatrix}$$

分散共分散行列 \mathbf{R} の各成分である分散と共分散を、構造方程式モデルより求めて代入する。

$$\begin{aligned} \text{Var}(Z_1) &= \text{Var}(w_{11}f_1 + e_1) = w_{11}^2 \underbrace{\text{Var}(f_1)}_{=1} + 2 \underbrace{\text{Cov}(f_1, e_1)}_{=0} + \text{Var}(e_1) = w_{11}^2 + \text{Var}(e_1) \\ \text{Var}(Z_2) &= \text{Var}(w_{21}f_1 + e_2) = w_{21}^2 \text{Var}(f_1) + 2 \text{Cov}(f_1, e_2) + \text{Var}(e_2) = w_{21}^2 + \text{Var}(e_2) \\ \text{Cov}(Z_1, Z_2) &= \text{Cov}(Z_2, Z_1) = \text{Cov}(w_{11}f_1 + e_1, w_{21}f_1 + e_2) \\ &= w_{11}w_{21} \underbrace{\text{Cov}(f_1, f_1)}_{=1} + w_{11} \underbrace{\text{Cov}(f_1, e_2)}_{=0} + w_{21} \underbrace{\text{Cov}(e_1, f_1)}_{=0} + w_{11}w_{21} \underbrace{\text{Cov}(e_1, e_2)}_{=0} = w_{11}w_{21} \end{aligned}$$

得られた分散共分散行列 \mathbf{R} から残差行列 \mathbf{E} を除いた $\mathbf{R}-\mathbf{E}$ は、因子負荷量ベクトル \mathbf{W} の積に分解できる。

統計的因果推論に生じた奇妙な現象ともややも感（深瀬）

$$\begin{aligned}\therefore R &= \begin{bmatrix} \text{Var}(Z_1) & \text{Cov}(Z_1, Z_2) \\ \text{Cov}(Z_2, Z_1) & \text{Var}(Z_2) \end{bmatrix} = \begin{bmatrix} w_{11}^2 + \text{Var}(e_1) & w_{11}w_{21} \\ w_{11}w_{21} & w_{21}^2 + \text{Var}(e_2) \end{bmatrix} \\ &= \begin{bmatrix} w_{11}^2 & w_{11}w_{21} \\ w_{11}w_{21} & w_{21}^2 \end{bmatrix} + \begin{bmatrix} \text{Var}(e_1) & 0 \\ 0 & \text{Var}(e_2) \end{bmatrix} = \underbrace{\begin{bmatrix} w_{11} \\ w_{21} \end{bmatrix}}_{\mathbf{W}} \underbrace{\begin{bmatrix} w_{11} & w_{21} \end{bmatrix}}_{\mathbf{W}^t} + \underbrace{\begin{bmatrix} \text{Var}(e_1) & 0 \\ 0 & \text{Var}(e_2) \end{bmatrix}}_{\mathbf{E}}\end{aligned}$$

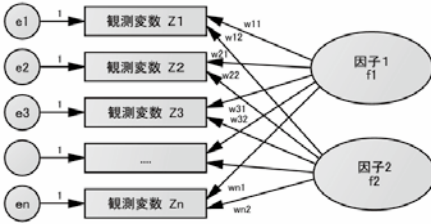
$\mathbf{R} = \mathbf{W}\mathbf{W}^t + \mathbf{E}$ （因子分析の基本方程式）

(2) 2 因子モデルの因子負荷量ベクトル \mathbf{W} 導出

3 つの観測変数 \mathbf{Z} に対して、因子 \mathbf{F} が 2 つある場合、

図 A.3 3 観測変数の 2 因子モデル

$$\mathbf{Z} = \mathbf{F}\mathbf{W}^t + \mathbf{e}$$



z_{ij} : 観測変数、 f_{ik} : 潜在変数、 w_{jk} : 因子負荷量、 e_{ij} : 残差項、 i : オブザベーション

($i = 1, 2, \dots, n$)、 j : 観測変数の識別 ($j = 1, 2, 3$)、 k : 潜在変数の識別 ($k = 1, 2$)

$$\mathbf{Z} = \begin{bmatrix} z_{11} & z_{12} & z_{13} \\ z_{21} & z_{22} & z_{23} \\ \vdots & \vdots & \vdots \\ z_{n1} & z_{n2} & z_{n3} \end{bmatrix} \quad \mathbf{F} = \begin{bmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \\ \vdots & \vdots \\ f_{n1} & f_{n2} \end{bmatrix} \quad \mathbf{W} = \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \\ \vdots & \vdots \\ w_{31} & w_{32} \end{bmatrix} \quad \mathbf{e} = \begin{bmatrix} e_{11} & e_{12} & e_{13} \\ e_{21} & e_{22} & e_{23} \\ \vdots & \vdots & \vdots \\ e_{n1} & e_{n2} & e_{n3} \end{bmatrix}$$

$$\begin{bmatrix} z_{11} & z_{12} & z_{13} \\ z_{21} & z_{22} & z_{23} \\ \vdots & \vdots & \vdots \\ z_{n1} & z_{n2} & z_{n3} \end{bmatrix} = \begin{bmatrix} f_{11} & f_{12} \\ f_{12} & f_{22} \\ \vdots & \vdots \\ f_{1n} & f_{2n} \end{bmatrix} \begin{bmatrix} w_{11} & w_{21} & w_{31} \\ w_{12} & w_{22} & w_{32} \end{bmatrix} + \begin{bmatrix} e_{11} & e_{12} & e_{13} \\ e_{21} & e_{22} & e_{23} \\ \vdots & \vdots & \vdots \\ e_{n1} & e_{n2} & e_{n3} \end{bmatrix}$$

因子分析の基礎方程式：この関係式は任意の数の観測変数と因子に対して成り立つ。

$$\mathbf{R} = \mathbf{W}_1 \mathbf{W}_1^t + \mathbf{W}_2 \mathbf{W}_2^t + \mathbf{E}$$

$$\therefore \mathbf{E} = \mathbf{R} - \mathbf{W}_1 \mathbf{W}_1^t - \mathbf{W}_2 \mathbf{W}_2^t$$

ただし、 $\mathbf{E} = \begin{bmatrix} \text{Var}[e_1] & 0 & 0 \\ \text{Cov}(e_1, e_2) & \text{Var}[e_2] & 0 \\ 0 & 0 & \text{Var}[e_3] \end{bmatrix}$

$$\begin{aligned} \mathbf{R} &= \begin{bmatrix} 1 & \text{Cov}(Z_1, Z_2) & \text{Cov}(Z_1, Z_3) \\ \text{Cov}(Z_2, Z_1) & 1 & \text{Cov}(Z_2, Z_3) \\ \text{Cov}(Z_3, Z_1) & \text{Cov}(Z_3, Z_2) & 1 \end{bmatrix} \\ &= \mathbf{W} \mathbf{W}^t + \mathbf{E} = \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \\ w_{31} & w_{32} \end{bmatrix} \begin{bmatrix} w_{11} & w_{21} & w_{31} \\ w_{12} & w_{22} & w_{32} \end{bmatrix} + \mathbf{E} \\ &= \begin{bmatrix} w_{11}^2 + w_{12}^2 & w_{11}w_{21} + w_{12}w_{22} & w_{11}w_{31} + w_{12}w_{32} \\ w_{21}w_{11} + w_{22}w_{12} & w_{21}^2 + w_{22}^2 & w_{21}w_{31} + w_{22}w_{32} \\ w_{31}w_{11} + w_{32}w_{12} & w_{31}w_{21} + w_{32}w_{22} & w_{31}^2 + w_{32}^2 \end{bmatrix} + \mathbf{E} \\ &= \begin{bmatrix} w_{11}^2 & w_{11}w_{21} & w_{11}w_{31} \\ w_{21}w_{11} & w_{21}^2 & w_{21}w_{31} \\ w_{31}w_{11} & w_{31}w_{21} & w_{31}^2 \end{bmatrix} + \begin{bmatrix} w_{12}^2 & w_{12}w_{22} & w_{12}w_{32} \\ w_{22}w_{12} & w_{22}^2 & w_{22}w_{32} \\ w_{32}w_{12} & w_{32}w_{22} & w_{32}^2 \end{bmatrix} + \mathbf{E} \\ &= \underbrace{\begin{bmatrix} w_{11} \\ w_{21} \\ w_{31} \end{bmatrix}}_{\mathbf{W}_1} \underbrace{\begin{bmatrix} w_{11} & w_{21} & w_{31} \end{bmatrix}}_{\mathbf{W}_1^t} + \underbrace{\begin{bmatrix} w_{12} \\ w_{22} \\ w_{32} \end{bmatrix}}_{\mathbf{W}_2} \underbrace{\begin{bmatrix} w_{12} & w_{22} & w_{32} \end{bmatrix}}_{\mathbf{W}_2^t} + \mathbf{E} \end{aligned}$$

※「因子分析の基礎方程式」によって、理論的には、因子負荷量行列に分解できるが、具体的な因子負荷量の値を特定できない。

3. 相関係数行列 \mathbf{R} のスペクトル分解（固有値分解）

「因子分析の基礎方程式」に基づく具体的な因子負荷量ベクトル \mathbf{W} を、固有値、固有ベクトルを求めて解析的（代数的）に特定する方法である。

（1）主因子法による考察

因子分析における共通因子の抽出方法には、最尤法、最小2乗法、反復主因子法などがあるが、モデルの解析的な理解のため、最も単純な主因子法について考察する。

例：3 観測変数の2 因子モデル

（手順1）行列のスペクトル分解（固有値分解）

行列を $[\text{固有値 } \lambda] \times [\text{固有ベクトル } \bar{\mathbf{x}}] \times [\text{その転置ベクトル } \bar{\mathbf{x}}']$ の積和に分解できる。

$$\mathbf{R} = \lambda_1 \bar{\mathbf{x}}_1 \bar{\mathbf{x}}_1' + \lambda_2 \bar{\mathbf{x}}_2 \bar{\mathbf{x}}_2' + \lambda_3 \bar{\mathbf{x}}_3 \bar{\mathbf{x}}_3' \quad (\text{因子分析の基礎方程式})$$

分散共分散行列（相関行列） \mathbf{R} をスペクトル分解すると、固有値 λ の平方根 \times 固有ベクトル $\bar{\mathbf{x}}$ によって、近似的に因子負荷量ベクトル \mathbf{w} が求められる。

（手順1）観測変数についての相関係数行列 \mathbf{R} の固有値と固有ベクトルを求める。

$$\mathbf{R}\mathbf{x} = \lambda\mathbf{x}$$

$$\mathbf{R} = \begin{bmatrix} 1 & \text{Cov}[Z_1, Z_2] & \text{Cov}[Z_1, Z_3] \\ \text{Cov}[Z_2, Z_1] & 1 & \text{Cov}[Z_2, Z_3] \\ \text{Cov}[Z_3, Z_1] & \text{Cov}[Z_3, Z_2] & 1 \end{bmatrix} \quad \bar{\mathbf{x}} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

$$(\mathbf{R} - \lambda \mathbf{I}) \bar{\mathbf{x}} = \mathbf{0}$$

固有方程式が、非自明解を持つためには、係数行列の行列式が0でなければならない。

$$|\mathbf{R} - \lambda \mathbf{I}| = 0$$

$$\begin{vmatrix} 1-\lambda & \text{Cov}[Z_1, Z_2] & \text{Cov}[Z_1, Z_3] \\ \text{Cov}[Z_2, Z_1] & 1-\lambda & \text{Cov}[Z_2, Z_3] \\ \text{Cov}[Z_3, Z_1] & \text{Cov}[Z_3, Z_2] & 1-\lambda \end{vmatrix} = 0$$

$$(1-\lambda)^3 + \text{Cov}[Z_2, Z_1]\text{Cov}[Z_3, Z_2]\text{Cov}[Z_1, Z_3] + \text{Cov}[Z_3, Z_1]\text{Cov}[Z_2, Z_3]\text{Cov}[Z_1, Z_2] - \text{Cov}[Z_1, Z_3]^2(1-\lambda) - \text{Cov}[Z_1, Z_2]^2(1-\lambda) - \text{Cov}[Z_2, Z_3]^2(1-\lambda) = 0$$

$$(1-\lambda)^3 - [\text{Cov}[Z_1, Z_2]^2 + \text{Cov}[Z_2, Z_3]^2 + \text{Cov}[Z_1, Z_3]^2](1-\lambda)$$

$$+ 2\text{Cov}[Z_2, Z_1]\text{Cov}[Z_3, Z_2]\text{Cov}[Z_1, Z_3] = 0$$

$$\lambda = \lambda_1 > \lambda_2 > \lambda_3$$

固有値 λ を冒頭の式に代入すれば、それぞれの固有値

$\lambda_1 > \lambda_2 > \lambda_3$ に対して、固有ベクトル $\vec{\mathbf{x}}_1$ $\vec{\mathbf{x}}_2$ $\vec{\mathbf{x}}_3$ が求まる。

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{21} & x_{31} \\ x_{12} & x_{22} & x_{32} \\ x_{13} & x_{23} & x_{33} \end{bmatrix} = [\vec{\mathbf{x}}_1 \quad \vec{\mathbf{x}}_2 \quad \vec{\mathbf{x}}_3] \quad \mathbf{X}^t = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix} = \begin{bmatrix} \vec{\mathbf{x}}_1 \\ \vec{\mathbf{x}}_2 \\ \vec{\mathbf{x}}_3 \end{bmatrix}$$

その際に、解が不定にならないよう $\vec{\mathbf{x}}$ の長さが1になるように制約をかける。

固有ベクトル $\vec{\mathbf{x}}$ は、互いに直交し、一次独立の関係がある。

$$\begin{cases} \mathbf{R}\vec{\mathbf{x}}_1 = \lambda_1\vec{\mathbf{x}}_1 \\ \mathbf{R}\vec{\mathbf{x}}_2 = \lambda_2\vec{\mathbf{x}}_2 \\ \mathbf{R}\vec{\mathbf{x}}_3 = \lambda_3\vec{\mathbf{x}}_3 \end{cases} \Rightarrow \begin{cases} \|\vec{\mathbf{x}}_1\| = \|\vec{\mathbf{x}}_2\| = \|\vec{\mathbf{x}}_3\| = 1 \\ \vec{\mathbf{x}}_1 \perp \vec{\mathbf{x}}_2 \perp \vec{\mathbf{x}}_3 \end{cases}$$

(手順2) 固有ベクトルの正規直交性を利用して、2次形式の積を作り対角化する

$$\mathbf{R}\bar{\mathbf{x}}_i = \lambda_i \bar{\mathbf{x}}_i$$

$$\mathbf{X}'\mathbf{R}\mathbf{X} = \begin{bmatrix} \bar{\mathbf{x}}_1 \\ \bar{\mathbf{x}}_2 \\ \bar{\mathbf{x}}_3 \end{bmatrix} \mathbf{R} \begin{bmatrix} \bar{\mathbf{x}}_1 & \bar{\mathbf{x}}_2 & \bar{\mathbf{x}}_3 \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{x}}_1 \\ \bar{\mathbf{x}}_2 \\ \bar{\mathbf{x}}_3 \end{bmatrix} \underbrace{\begin{bmatrix} \lambda_1 \bar{\mathbf{x}}_1 & \lambda_2 \bar{\mathbf{x}}_2 & \lambda_3 \bar{\mathbf{x}}_3 \end{bmatrix}}_{R[\bar{\mathbf{x}}_1 \ \bar{\mathbf{x}}_2 \ \bar{\mathbf{x}}_3]}$$

$$= \begin{bmatrix} \lambda_1 \bar{\mathbf{x}}_1' \bar{\mathbf{x}}_1 & \lambda_2 \bar{\mathbf{x}}_1' \bar{\mathbf{x}}_2 & \lambda_3 \bar{\mathbf{x}}_1' \bar{\mathbf{x}}_3 \\ \lambda_1 \bar{\mathbf{x}}_2' \bar{\mathbf{x}}_1 & \lambda_2 \bar{\mathbf{x}}_2' \bar{\mathbf{x}}_2 & \lambda_3 \bar{\mathbf{x}}_2' \bar{\mathbf{x}}_3 \\ \lambda_1 \bar{\mathbf{x}}_3' \bar{\mathbf{x}}_1 & \lambda_2 \bar{\mathbf{x}}_3' \bar{\mathbf{x}}_2 & \lambda_3 \bar{\mathbf{x}}_3' \bar{\mathbf{x}}_3 \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} = \mathbf{\Lambda}$$

$$\therefore \mathbf{R} = \mathbf{X}(\mathbf{X}'\mathbf{R}\mathbf{X})\mathbf{X}' = \mathbf{X} \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} \mathbf{X}' = \begin{bmatrix} \bar{\mathbf{x}}_1 & \bar{\mathbf{x}}_2 & \bar{\mathbf{x}}_3 \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} \begin{bmatrix} \bar{\mathbf{x}}_1' \\ \bar{\mathbf{x}}_2' \\ \bar{\mathbf{x}}_3' \end{bmatrix}$$

$$= \begin{bmatrix} \bar{\mathbf{x}}_1 & \bar{\mathbf{x}}_2 & \bar{\mathbf{x}}_3 \end{bmatrix} \begin{bmatrix} \lambda_1 \bar{\mathbf{x}}_1' \\ \lambda_2 \bar{\mathbf{x}}_2' \\ \lambda_3 \bar{\mathbf{x}}_3' \end{bmatrix} = \lambda_1 \bar{\mathbf{x}}_1 \bar{\mathbf{x}}_1' + \lambda_2 \bar{\mathbf{x}}_2 \bar{\mathbf{x}}_2' + \lambda_3 \bar{\mathbf{x}}_3 \bar{\mathbf{x}}_3'$$

$$\begin{array}{l} \left. \begin{array}{l} \bar{\mathbf{x}}_i = \begin{bmatrix} x_{i1} \\ x_{i2} \\ x_{i3} \end{bmatrix} \quad \bar{\mathbf{x}}_i' = \begin{bmatrix} x_{i1} & x_{i2} & x_{i3} \end{bmatrix} \\ \text{注}^* \quad \bar{\mathbf{x}}_i' \bar{\mathbf{x}}_i = \begin{bmatrix} x_{i1} & x_{i2} & x_{i3} \end{bmatrix} \begin{bmatrix} x_{i1} \\ x_{i2} \\ x_{i3} \end{bmatrix} = \sum_{j=1}^3 x_{ij}^2 = \|\bar{\mathbf{x}}_i\|^2 = 1 \\ \bar{\mathbf{x}}_i' \bar{\mathbf{x}}_i = \begin{bmatrix} x_{i1} \\ x_{i2} \\ x_{i3} \end{bmatrix} \begin{bmatrix} x_{i1} & x_{i2} & x_{i3} \end{bmatrix} = \begin{bmatrix} x_{i1}x_{i1} & x_{i1}x_{i2} & x_{i1}x_{i3} \\ x_{i2}x_{i1} & x_{i2}x_{i2} & x_{i2}x_{i3} \\ x_{i3}x_{i1} & x_{i3}x_{i2} & x_{i3}x_{i3} \end{bmatrix} \end{array} \right\} \end{array}$$

(手順3) スペクトル分解による因子抽出

$$\begin{aligned} \begin{bmatrix} 1 & \text{Cov}[Z_1, Z_2] & \text{Cov}[Z_1, Z_3] \\ \text{Cov}[Z_2, Z_1] & 1 & \text{Cov}[Z_2, Z_3] \\ \text{Cov}[Z_3, Z_1] & \text{Cov}[Z_3, Z_2] & 1 \end{bmatrix} &= \lambda_1 \underbrace{\begin{bmatrix} x_{11} & x_{12} & x_{13} \end{bmatrix}}_{\bar{\mathbf{x}}_1'} + \lambda_2 \underbrace{\begin{bmatrix} x_{21} \\ x_{22} \\ x_{23} \end{bmatrix}}_{\bar{\mathbf{x}}_2'} + \lambda_3 \underbrace{\begin{bmatrix} x_{31} \\ x_{32} \\ x_{33} \end{bmatrix}}_{\bar{\mathbf{x}}_3'} \\ &= \begin{bmatrix} \sqrt{\lambda_1} \begin{bmatrix} x_{11} \\ x_{12} \\ x_{13} \end{bmatrix} & \sqrt{\lambda_2} \begin{bmatrix} x_{21} \\ x_{22} \\ x_{23} \end{bmatrix} & \sqrt{\lambda_3} \begin{bmatrix} x_{31} \\ x_{32} \\ x_{33} \end{bmatrix} \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} \begin{bmatrix} x_{11} & x_{12} & x_{13} \end{bmatrix} \\ \sqrt{\lambda_2} \begin{bmatrix} x_{21} & x_{22} & x_{23} \end{bmatrix} \\ \sqrt{\lambda_3} \begin{bmatrix} x_{31} & x_{32} & x_{33} \end{bmatrix} \end{bmatrix} = \begin{bmatrix} \sqrt{\lambda_1} x_{11} & \sqrt{\lambda_2} x_{21} & \sqrt{\lambda_3} x_{31} \\ \sqrt{\lambda_1} x_{12} & \sqrt{\lambda_2} x_{22} & \sqrt{\lambda_3} x_{32} \\ \sqrt{\lambda_1} x_{13} & \sqrt{\lambda_2} x_{23} & \sqrt{\lambda_3} x_{33} \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} x_{11} & \sqrt{\lambda_1} x_{12} & \sqrt{\lambda_1} x_{13} \\ \sqrt{\lambda_2} x_{21} & \sqrt{\lambda_2} x_{22} & \sqrt{\lambda_2} x_{23} \\ \sqrt{\lambda_3} x_{31} & \sqrt{\lambda_3} x_{32} & \sqrt{\lambda_3} x_{33} \end{bmatrix} \end{aligned}$$

観測変数の分散共分散行列が、行列とその転置行列との自乗積に分解できたので、これを因子負荷量行列 \mathbf{W} とみなすことができる。

$$\begin{bmatrix} \sqrt{\lambda_1}x_{11} & \sqrt{\lambda_2}x_{21} & \sqrt{\lambda_3}x_{31} \\ \sqrt{\lambda_1}x_{12} & \sqrt{\lambda_2}x_{22} & \sqrt{\lambda_3}x_{32} \\ \sqrt{\lambda_1}x_{13} & \sqrt{\lambda_2}x_{23} & \sqrt{\lambda_3}x_{33} \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1}x_{11} & \sqrt{\lambda_1}x_{12} & \sqrt{\lambda_1}x_{13} \\ \sqrt{\lambda_2}x_{21} & \sqrt{\lambda_2}x_{22} & \sqrt{\lambda_2}x_{23} \\ \sqrt{\lambda_3}x_{31} & \sqrt{\lambda_3}x_{32} & \sqrt{\lambda_3}x_{33} \end{bmatrix} = \underbrace{\begin{bmatrix} w_{11} & w_{21} & w_{31} \\ w_{12} & w_{22} & w_{32} \\ w_{13} & w_{23} & w_{33} \end{bmatrix}}_{\mathbf{w}} \underbrace{\begin{bmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \\ w_{31} & w_{32} & w_{33} \end{bmatrix}}_{\mathbf{w}^t}$$

(補足) スペクトル分解を用いない因子抽出

\mathbf{X} が正規直交行列である場合は、自乗積が単位行列になる。

$$\mathbf{X}\mathbf{X}^t = \begin{bmatrix} \bar{\mathbf{x}}_1 & \bar{\mathbf{x}}_2 & \bar{\mathbf{x}}_3 \end{bmatrix} \begin{bmatrix} \bar{\mathbf{x}}_1 \\ \bar{\mathbf{x}}_2 \\ \bar{\mathbf{x}}_3 \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{x}}_1\bar{\mathbf{x}}_1 & \bar{\mathbf{x}}_1\bar{\mathbf{x}}_2 & \bar{\mathbf{x}}_1\bar{\mathbf{x}}_3 \\ \bar{\mathbf{x}}_2\bar{\mathbf{x}}_1 & \bar{\mathbf{x}}_2\bar{\mathbf{x}}_2 & \bar{\mathbf{x}}_2\bar{\mathbf{x}}_3 \\ \bar{\mathbf{x}}_3\bar{\mathbf{x}}_1 & \bar{\mathbf{x}}_3\bar{\mathbf{x}}_2 & \bar{\mathbf{x}}_3\bar{\mathbf{x}}_3 \end{bmatrix} = \begin{bmatrix} \|\mathbf{x}_1\|^2 & 0 & 0 \\ 0 & \|\mathbf{x}_2\|^2 & 0 \\ 0 & 0 & \|\mathbf{x}_3\|^2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \mathbf{I}$$

よって、

$$\mathbf{R}\bar{\mathbf{x}}_i = \lambda_i \bar{\mathbf{x}}_i$$

$$\mathbf{R} \begin{bmatrix} \bar{\mathbf{x}}_1 & \bar{\mathbf{x}}_2 & \bar{\mathbf{x}}_3 \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{x}}_1 & \bar{\mathbf{x}}_2 & \bar{\mathbf{x}}_3 \end{bmatrix} \mathbf{\Lambda}$$

$$\mathbf{R}\mathbf{X} = \mathbf{X}\mathbf{\Lambda}$$

$$\therefore \mathbf{R} = \underbrace{\mathbf{R}\mathbf{X}\mathbf{X}^t}_{\mathbf{I}} = \mathbf{X}\mathbf{\Lambda}\mathbf{X}^t$$

すなわち、

$$\begin{aligned} & \begin{bmatrix} 1 & \text{Cov}[Z_1, Z_2] & \text{Cov}[Z_1, Z_3] \\ \text{Cov}[Z_2, Z_1] & 1 & \text{Cov}[Z_2, Z_3] \\ \text{Cov}[Z_3, Z_1] & \text{Cov}[Z_3, Z_2] & 1 \end{bmatrix} = \begin{bmatrix} x_{11} & x_{21} & x_{31} \\ x_{12} & x_{22} & x_{32} \\ x_{13} & x_{23} & x_{33} \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix} \\ & = \underbrace{\begin{bmatrix} x_{11} & x_{21} & x_{31} \\ x_{12} & x_{22} & x_{32} \\ x_{13} & x_{23} & x_{33} \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} & 0 & 0 \\ 0 & \sqrt{\lambda_2} & 0 \\ 0 & 0 & \sqrt{\lambda_3} \end{bmatrix}}_{\mathbf{X}} \begin{bmatrix} \sqrt{\lambda_1} & 0 & 0 \\ 0 & \sqrt{\lambda_2} & 0 \\ 0 & 0 & \sqrt{\lambda_3} \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix} \\ & = \begin{bmatrix} \sqrt{\lambda_1}x_{11} & \sqrt{\lambda_2}x_{21} & \sqrt{\lambda_3}x_{31} \\ \sqrt{\lambda_1}x_{12} & \sqrt{\lambda_2}x_{22} & \sqrt{\lambda_3}x_{32} \\ \sqrt{\lambda_1}x_{13} & \sqrt{\lambda_2}x_{23} & \sqrt{\lambda_3}x_{33} \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1}x_{11} & \sqrt{\lambda_1}x_{12} & \sqrt{\lambda_1}x_{13} \\ \sqrt{\lambda_2}x_{21} & \sqrt{\lambda_2}x_{22} & \sqrt{\lambda_2}x_{23} \\ \sqrt{\lambda_3}x_{31} & \sqrt{\lambda_3}x_{32} & \sqrt{\lambda_3}x_{33} \end{bmatrix} = \underbrace{\begin{bmatrix} w_{11} & w_{21} & w_{31} \\ w_{12} & w_{22} & w_{32} \\ w_{13} & w_{23} & w_{33} \end{bmatrix}}_{\mathbf{w}} \underbrace{\begin{bmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \\ w_{31} & w_{32} & w_{33} \end{bmatrix}}_{\mathbf{w}^t} \end{aligned}$$

参考文献

- 足立浩平 (2005)、『多変量データ解析法』、ナカニシヤ出版
- 飯塚久哲 (2003)、「測定誤差の問題（希薄化について）」豊田秀樹編『共分散構造分析 [疑問編]』、朝倉書店
- 笠井樹希・科野智咲 (2019)、「非日常的店舗の具現化に向けた構成要素の探索、2018年度ゼミ卒業論集
- 金本拓 (2024)、『因果推論』、オーム社
- 狩野裕 (2002)、「構造方程式モデリングは、因子分析、分散分析、パス解析のすべてにとつて代わるのか」、『行動計量学』第29巻 2号
- 小杉考司 (2018)、『言葉と数式で理解する多変量解析入門』、北大路書房
- 塩谷茂明 (2009)「航行船舶の潮流・海上風の影響」『日本航海学会誌』172巻 ,pp.7-16
- 清水裕司・荘島宏二郎 (2017)、『社会心理学のための統計学』、誠心書房
- 庄豊源 (2024)、「製品普及におけるライブコマースを利用したキャストの起用」、ゼミ生修士論文 (学長賞)
- 高橋将宣 (2022)、『統計的因果推論の理論と実装』、共立出版
- 立森久照 (2016)、「因果推論とははじめ」、『岩波データサイエンス 3 因果推論－実世界のデータから世界を読む』、岩波書店
- 豊田秀樹 (2007)、『共分散構造分析 [Amos 編]』、東京図書
- 倪映傑 (2020)、「日本のインバウンド観光市場におけるコロナ禍からの復興－中国人留学生を登用する地域観光マーケティング」、ゼミ修士論文 (学長賞)
- ニ・テッケツ (2020)、「コロナ禍から日本のインバウンドをV字回復させる魔法：留学生とインフルエンサー起用によるライブコマース誘致策」株式会社 NTT データ数理システム、2020年度 VMStudio & TMStudio 学生研究奨励賞 佳作受賞論文
- 林岳彦 (2024)、『初めての統計的因果推論』、岩波書店
- 林岳彦・黒木学 (2016)、「相関と因果と丸と矢印のはなし」、『岩波データサイエンス 3 因果推論－実世界のデータから世界を読む』、岩波書店
- 星野崇宏 (2009)、『調査観察データの統計科学』、岩波書店
- 松永晃治 & 松浦萌 (2018)、「日本の保険営業実績における影響要因分析」、2017年度ゼミ卒業研究
- 宮川雅美 (2004)、『統計的因果推論』、朝倉書店
- 季丹 (2016)「网络临场感对传播效果的影响研究——基于信任关系的中介效应分析」现代传播 (中国传媒大学学报)
- 米村大介 (2003)、「潜在変数の数とパス係数の関係」、『共分散構造分析 [疑問編]』、朝倉書店、pp.86
- Alexander P. Schouten,Loes Janssen (2020), Maegan Verspaget, “Celebrity vs. Influencer endorsements in advertising: the role of identification, credibility,and Product-Endorser fit”, International Journal of Advertising
- Green,J.A. (1992), “Testing whether correlation matrices are different from each other Developmental”, *Psychology*, vol. 28, 215-224

- Lawrence A. Crosby, Kenneth R. Evans,& Deborah Cowles (1990), Relationship Quality in Service Selling: An Interpersonal Influence Perspective, "Journal of Marketing"
- Judea Pearl (1995). "Causal diagrams for empirical research". *Biometrika*, 82, 669-710.
- Judea Pearl (2009), "Causality: Models, Reasoning and Inference", Cambridge University Press, 2nd edition, 2009.
- Wright, S. (1934). The method of path coefficients. *Annals of Mathematical Statistics*, 5, 161-215
- Judea Pearl and Dana Mackenzie (2003), *The New Science of Cause and Effect*
- Jin Tian and Judea Pearl (2000), Probabilities of causation: Bounds and identification. *Annals of Mathematics and Artificial Intelligence*, 28 (1-4): 287-313,
- Tyler J VanderWeele and Peng Ding. (2017) Sensitivity analysis in observational research: introducing the e-value. *Annals of Internal Medicine*, 167 (4) : 268-274, 2017.
- John Short, Ederyn Williams, B. Christie (1976), *Psychology, Computer Science, Sociology The social psychology of telecommunications*, Wiley,

